

EdData II

Manual para la evaluación inicial de la lectura en niños de educación primaria



18 de noviembre, 2009

**EdData II Asistencia Técnica y Gerencia, Orden de Trabajo Número 3
Número de Contrato EHC-E-03-04-00004-00
Objetivo Estratégico 3**

Este manual se produjo para la Agencia de Los Estados Unidos para el Desarrollo Internacional. El original en inglés se produjo para la Oficina de Desarrollo Humano del Banco Mundial. Fue elaborado por RTI International, con la adaptación realizada al español por el consultor Juan E. Jiménez.

Manual para la evaluación inicial de la lectura en niños de educación primaria

Preparado para
Oficina de Desarrollo Económico, Agricultura y Comercio (EGAT/ED)
Agencia de los Estados Unidos para el Desarrollo Internacional

Sometido por

RTI International
3040 Cornwallis Road
Post Office Box 12194
Research Triangle Park, NC 27709-2194

RTI International es el nombre comercial de Research Triangle Institute.

Adaptación realizada al español por

Juan E. Jiménez
Catedrático de Psicología Evolutiva y de la Educación, Departamento de Psicología
Evolutiva y de la Educación, Facultad de Psicología
Universidad La Laguna, Campus de Guajara 38200
La Laguna, Tenerife, Islas Canarias
922-317545
ejimenez@ull.es

Las perspectivas de los autores que se expresan en este informe no reflejan necesariamente las opiniones ni de la Agencia de los Estados Unidos para el Desarrollo Internacional ni del Gobierno de los Estados Unidos.

Índice

Gráficas	iii
Tablas.....	iii
Abreviaturas	iv
Glosario de términos.....	vi
Agradecimientos	viii
I. Introducción	1
¿Por qué concentrarse en la lectura en los grados iniciales?.....	1
Midiendo el aprendizaje: el instrumento y el manual.....	2
Aplicaciones mundiales.....	3
Manual y contenido.....	5
II. Objetivo y usos de EGRA.....	7
Usando EGRA para identificar las necesidades del sistema	8
Usos adicionales de EGRA (con modificaciones).....	10
Identificación.....	10
Evaluación de las intervenciones.....	11
Cómo no debe usarse EGRA.....	12
EGRA no es una herramienta de rendición de cuentas de alto impacto.....	12
EGRA no resulta idóneo para las comparaciones entre lenguas	12
III. Marco conceptual e investigaciones de base	14
Evaluación inicial del aprendizaje de la lectura	14
Conciencia general del lenguaje escrito.....	16
Conocimiento alfabético y aprendizaje de las correspondencias grafema- fonema	17
Conciencia fonológica.....	19
Fluidez.....	21
Competencia lingüística (vocabulario y comprensión oral).....	21
El desarrollo de la lectura como una sucesión de etapas.....	22
Causas que explican el desarrollo de la lectura en lengua española.....	24
La transparencia de la ortografía y el aprendizaje de la lectura.....	25
Estudios comparativos	27
IV. Taller de adaptación e investigación de EGRA.....	30
Taller de adaptación e investigación	30
Nota sobre el diseño del instrumento y la estrategia de codificación.....	32
Nota sobre la ética de la investigación y la <i>Institutional Review Board</i> (IRB, Junta Revisora Institucional).....	33
Revisión de los componentes del instrumento	33
1. Conocimiento del nombre de las letras	35
2. Conocimiento de los sonidos de las letras	38
3. Conciencia fonológica.....	40
4. Conocimiento de palabras simples	42
5. Descodificación de palabras sin sentido (pseudopalabras).....	44
6. Lectura y comprensión de un párrafo (texto vinculado).....	45

7. Comprensión oral	47
8. Dictado	48
Otros posibles componentes del instrumento y razones de su exclusión	49
Traducción	50
V. Formadores de EGRA y trabajo de campo	52
La experiencia piloto del instrumento	52
Calculando la confiabilidad entre evaluadores	55
Llegada al centro educativo	56
Seleccionando educandos y llevando a cabo la evaluación	56
Enseñando lecciones para el trabajo de campo	57
VI. Analizando los datos de EGRA	58
Limpiando e ingresando datos	58
Análisis de datos: Usando Excel para analizar los datos	60
Tamaño de la muestra	62
VII. Usando EGRA: implicaciones para el diálogo de políticas	64
Usando resultados para informar el diálogo de políticas	64
Influyendo en la toma de decisiones de políticos y funcionarios	64
Cambiando la enseñanza de la lectura	67
Usando los datos para informar de los resultados a los centros educativos	67
Referencias	69
Anexo A. Normas sobre fluidez de la lectura oral en inglés para Estados Unidos	81
Anexo B. Consideraciones acerca del tamaño de la muestra en la <i>Early Grade Reading Assessment</i>	82
Anexo C. Evaluando la calidad técnica del instrumento EGRA	109
Anexo D. Carta abierta del subdirector general, África del Sur, a los directores de centros educativos	112
Anexo E. Agenda de taller de recuperación en lectura en Kenia, en base a los resultados de EGRA	115
Anexo F. Ejemplos de planes de lecciones para la recuperación de lectura, en base a los resultados de EGRA	118

Gráficas

Gráfica 1. Sedes de pruebas de EGRA a nivel mundial	4
Gráfica 2. El ciclo continuo de la mejora en el aprendizaje de los educandos.....	8
Gráfica 3. Puntajes de los educandos de palabras por minuto, 1º y 2º grado	10
Gráfica 4. El modelo dual del desarrollo ortográfico	22

Tablas

Tabla 1. Consistencia de las correspondencias grafema-fonema (CGF) y fonema-grafema (CFG) en las vocales, en inglés y francés (palabras monosílabas).....	26
Tabla 2. Ejemplo de agenda: Taller de adaptación e investigación de EGRA	32
Tabla 3. Revisión de los componentes del instrumento	35
Tabla 4. Letras en la lengua española: Frecuencia de uso	36
Tabla 5. Ejemplo de agenda: formación de examinadores y trabajo de campo piloto	53
Tabla 6. Estimados de educandos y centros educativos muestreados, y número de examinadores.....	54
Tabla 7. Ejemplos de entradas en el libro de claves	59
Tabla 8. Un ejemplo de referencias entre países con fines comparativos	66

Abreviaturas

AKF	Aga Khan Foundation [Fundación Aga Khan]
ANOVA	análisis of variance [análisis de varianza]
CCV	consonante-consonante-vocal
CF	conciencia fonológica
CFG	correspondencia fonema-grafema
CGF	correspondencia grafema-fonema
CIASES	Centro de Investigación y Acción Educativa Social [ONG nicaragüense]
CLE	[prueba de] Conocimientos sobre el Lenguaje Escrito
CTOPP	Comprehensive Test of Phonological Processing [Prueba Global de Procesamiento Fonológico]
CV	consonante-vocal
CVC	consonante-vocal-consonante
DEFF	design effect [efecto de diseño]
DFID	Department for International Development [Bretaña]
DIBELS	Dynamic Indicators of Basic Early Literacy
EGR	Early Grade Reading [proyecto, Kenia]
EGRA	Early Grade Reading Assessment [Prueba Diagnóstica de Lectura Inicial]
EMACK	Education for Marginalized Children in Kenya [proyecto; Educación de Niños Marginados en Kenia]
EPT	Educación para Todos [Education for All (EFA) en inglés; campaña de las Naciones Unidas]
FSP	frecuencia silábica posicional
IC	intervalo de confianza
ICC	intraclass correlation coefficient [coeficiente de correlación intraclases]
IRB	institutional review board [junta revisora institucional]
LAMP	Literacy Assessment and Monitoring Programme [UNESCO]
LCD	liquid-crystal display
LCPM	Letras leídas correctamente por minuto
LQAS	Lot Quality Assurance Sampling [muestreo de aceptación de lotes]
OCDE	Organización para la Cooperación y el Desarrollo Económico
ODM	Objetivo de Desarrollo del Milenio
ONG	organización no gubernamental
PASEC	Programme d'Analyse des Systems Educatifs de la Confemen
PCPM	palabras leídas correctamente por minuto
PISA	Programme for International Student Assessment [programa de la Organización para la Cooperación y el Desarrollo Económico]

PPVT	Peabody Picture Vocabulary Test
SAQMEC	Southern Africa Consortium for the Measurement of Educational Quality
SD	standard deviation [desviación estándar]
SE	standard error [error estándar]
TIMSS	Trends in International Mathematics and Science Study
TL	tiempo de latencia
TOPA	Test of Phonological Awareness
TR	tiempo de reacción
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
USAID	United States Agency for International Development [Agencia de Los Estados Unidos para el Desarrollo Internacional]

Glosario de términos

Automaticidad/fluidez. Es el nivel de dominio mediante el cual el individuo no es consciente o no tiene que realizar un esfuerzo mental para convertir los grafemas en sonidos al identificar una palabra escrita; en otras palabras, el lector realiza la descodificación rápidamente para concentrarse en la comprensión de lo leído.

Conciencia fonológica. Es la habilidad de manipular conscientemente las unidades más pequeñas y que no tienen significado del lenguaje oral. Se suelen distinguir distintos niveles de conciencia fonológica, dependiendo de la unidad lingüística con que estemos tratando. Decimos que hay una conciencia silábica cuando intentamos identificar alguna sílaba, o el número de sílabas de una palabra que escuchamos a nivel oral. Se entiende por conciencia intrasilábica el intento de manipular partes de la sílaba (i.e., el principio y la rima), por ejemplo identificar si dos palabras riman o no. Y por conciencia fonémica entendemos el intento de identificar algún fonema o número de fonemas en una palabra que escuchamos a nivel oral.

Conocimiento alfabético. Se refiere a un procedimiento de lectura que utiliza el principio alfabético y que consiste en que las palabras son leídas ensamblando los sonidos (fonemas) que las grafías (grafemas) representan.

Conocimiento metafonológico (véase conciencia fonológica).

Derivación (o palabra derivada). Son aquellas palabras que se forman con palabras primitivas o simples, a las que se añaden prefijos o sufijos (v.g., prefijos- *des-peinada*; sufijos- *pan-adero*) (véase morfema).

Descodificación (véase conocimiento alfabético)

Dígrafo. Es un grupo de dos letras que representan un solo sonido. En español, los dígrafos que corresponden a sonidos no representados por una letra individual son ch, ll, qu, gu (delante de e ó i) y rr. En este caso algunos dígrafos se consideran grafemas (véase grafema).

Efecto suelo. Se trata de un término estadístico que indica la presencia de un rango no muy grande de variabilidad, esto es, puntuaciones muy bajas. Si para valorar los conocimientos en una materia tomamos un examen y todos obtienen 0 se dice que hay un efecto suelo; en cambio, si todos obtienen 10 se dice que existe un efecto techo.

Fonemas. Son sonidos del habla que permiten distinguir las palabras en una lengua. Así, los sonidos /p/ y /b/ son fonemas del español porque existen palabras como /pala/ y /bala/, que tienen distinto significado y cuya pronunciación sólo difiere en relación a estos dos sonidos.

Forma flexiva. Cuando las palabras experimentan variaciones en sus morfemas finales. Así, por ejemplo, los cambios en la forma de una palabra según el género ('niño'/'niña'), el número ('un lápiz'/'dos lápices ') o el tiempo ('hoy llueve'/'ayer llovía').

Grafema. Es la unidad más elemental de un sistema de escritura. Los grafemas representan fonemas (ver fonema). Varios grafemas pueden representar un sólo fonema (es el caso de la ch en español) (véase dígrafo). Un grafema puede estar compuesto por una o más letras, o por una letra con marca diacrítica (v.g., «é»/«e» en francés).

Grupo consonántico. Un grupo de dos consonantes consecutivas en posición inicial en una sílaba (v.g., 'fr' en 'fresa').

Logográfico. Es un procedimiento de lectura basado en el reconocimiento de las palabras escritas a partir de claves visuales (v.g., color, tamaño, etc.) y de ciertos indicios gráficos locales (v.g., las primeras letras de las palabras).

Medida de fluidez. Es una medida de la competencia lectora, entendida como la habilidad para leer con exactitud y rapidez (véase automaticidad/fluidez).

Método fonológico. Es un método de enseñanza de la lectura que enseña el conocimiento de las correspondencias grafema-fonema.

Morfema. Es la unidad lingüística más pequeña con significado. Es diferente de una palabra, ya que las palabras están compuestas por uno o varios morfemas (v.g., un morfema = mar; tres morfemas = contra-dic-ción). Los morfemas que pueden formar una palabra por sí solos se llaman morfemas libres (v.g., mar), y los que tienen que agruparse con otros morfemas para formar palabras son los morfemas ligados o dependientes (v.g., niñ+o = niño; niñ+a = niña; niñ+ita = niñita, etc.).

Ortografía. El arte de escribir las palabras según las normas ortográficas.

Ortografía superficial o transparente. Dícese de aquellos sistemas de escritura con una ortografía alfabética y en donde se respeta en mayor o menor grado la coherencia de las correspondencias entre grafemas y fonemas, es decir que a cada fonema le corresponde un grafema único y específico. Entre los sistemas transparentes se encuentran el español, el alemán y el italiano.

Ortografía profunda u opaca. Dícese de aquellos sistemas de escritura con ortografía alfabética que presentan una mayor irregularidad en las correspondencias fonológicas, debido a que también representan unidades morfológicas. Ejemplos de ello serían el inglés y el francés.

Principio. La parte inicial de la sílaba que precede al núcleo vocálico y que puede estar formada por una o dos consonantes (v.g., g- o gr en español).

Ortográfico. Procedimiento de lectura que utiliza el principio ortográfico, a saber, que las palabras escritas codifican no sólo las unidades de base de la lengua oral (los fonemas), sino también otras marcas como las morfológicas (ver morfema).

Rima. Es una parte de la sílaba que está compuesta por un núcleo vocálico y por la coda, que es la consonante final (y que siempre es opcional). Por ejemplo, la palabra *sol* es una palabra monosílaba, es decir que coincide con la sílaba y está formada por el principio –s y por la rima –ol. A su vez, la rima está compuesta por la vocal –o y la coda –l.

Agradecimientos

Este manual es el producto de una colaboración en curso entre una gran comunidad de académicos, profesionales, funcionarios gubernamentales y profesionales del desarrollo educativo, para promover la causa de una temprana evaluación y adquisición de la lectura entre los niños de la escuela primaria en países de bajos ingresos.

Aunque no es posible agradecer a todos los que contribuyeron al desarrollo y proliferación del Early Grade Reading Assessment (EGRA; la Prueba Diagnóstica de Lectura Inicial), vaya un reconocimiento especial a Helen Abadzi, Marilyn Jager Adams, Rebecca Adams, Rukmini Banerji, Danielle Bechenec, Penelope Bender, Sandra Bertoli, Joanne Capper, Vanessa Castro, Colette Chabbott, Madhav Chavan, Marguerite Clarke, Penny Collins, Luis Crouch, Marcia Davidson, Joe DeStefano, Maria Diarra, Zakeya El-Nahas, Deborah Fredo, Ward Heneveld, Robin Horn, Sandra Hollingsworth, Juan E. Jiménez, Matthew Jukes, Cheryl Kim, Medina Korda, José Ramón Laguna, Nathalie Lahire, Sylvia Linan-Thompson, Corrine McComb, Emily Miksic, Amy Mulcahy-Dunn, Lily Mulatu, Lynn Murphy, Robert Prouty, Alastair Rodd, Mike Royer, Momar Sambe, Ernesto Schiefelbein, Dana Schmidt, Philip Seymour, Linda Siegel, Jennifer Spratt, Liliane Sprenger-Charolles, Helen Stannard, Jim Stevens, Carmen Strigel, Ian Smythe, Sana Tibi, Gerald Tindal, Palesa Tyobeka, Dan Wagner y Jim Wile.

Marcia Davidson, Sandra Hollingsworth, Sylvia Linan-Thompson y Liliane Sprenger-Charolles hicieron extensos comentarios en la evaluación por pares de este manual, así como sugerencias para el desarrollo de instrumentos.

El desarrollo del EGRA no habría sido posible sin el respaldo de las organizaciones no gubernamentales y los Equipos de Evaluación de EGRA de los ministerios de educación en Afganistán, Bangladesh, Egipto, Gambia, Guyana, Haití, Honduras, Jamaica, Kenia, Liberia, Malí, Nicaragua, Níger, Perú, Senegal y África del Sur. Vaya nuestro profundo agradecimiento a los docentes, alumnos y sus familias por su participación y continua fe en los beneficios de la educación. Por ello, estamos en deuda con ellos y buscamos mejorar los resultados de lectura de todos los niños alrededor del mundo.

Amber Gove es la responsable de la autoría primaria del manual, con la contribución de Juan E. Jiménez, Luis Crouch, Amy Mulcahy-Dunn y Marguerite Clarke. Las opiniones expresadas en este documento son las de los autores y no necesariamente reflejan el parecer de la Agencia de los Estados Unidos para el Desarrollo Internacional o el Banco Mundial. Por favor dirija sus preguntas o comentarios a Amber Gove en agove@rti.org.

I. Introducción

¿Por qué concentrarse en la lectura en los grados iniciales?

Los países en todo el mundo han venido impulsando la escolarización obligatoria en la escuela primaria a tasas históricamente sin precedentes. Buscando honrar los compromisos de la campaña Educación para Todos (EPT) y los Objetivos de Desarrollo del Milenio (ODM) de las Naciones Unidas, los países de ingresos medios, con el respaldo internacional, están escolarizando a los niños en la escuela primaria casi a tasas similares de los países de altos ingresos. ¿Pero los alumnos están aprendiendo?

Las evidencias, cuando contamos con ellas, indican que el aprendizaje del educando promedio en la mayoría de los países de bajos ingresos es sumamente bajo. Una evaluación reciente de los préstamos del Banco Mundial para educación muestra que las mejoras en el aprendizaje están retrasándose significativamente con respecto a las mejoras en el acceso a la educación (Banco Mundial: Independent Evaluation Group, 2006). Los resultados de aquellos pocos países de bajos ingresos que participan en evaluaciones internacionales como PISA o TIMSS (y efectuando inferencias a partir de los resultados procedentes de evaluaciones regionales tales como PASEC y SACMEQ),¹ indican que el niño mediano en un país de bajos ingresos tiene un desempeño correspondiente aproximadamente al del 3º percentil de la distribución de un país de altos ingresos (esto es, peor que el 97 por ciento de los educandos evaluados en este último país).² A partir de dichos resultados podemos decir qué cosa *no* sabían los educandos de un país de bajos ingresos, pero no podemos establecer qué cosa sí sabían (frecuentemente porque tuvieron un puntaje tan malo que la prueba no podía precisar su ubicación en el continuo del conocimiento). Es más, la mayoría de las evaluaciones nacionales e internacionales están basadas en pruebas de papel y lápiz tomadas a alumnos de 4º grado o más (esto es, asumen que pueden leer y escribir). A partir de estos resultados no siempre es posible saber si los alumnos tienen un puntaje bajo porque carecen de los conocimientos evaluados por la prueba, o porque no cuentan con las habilidades elementales de lectura y de comprensión.

La capacidad de leer y entender un texto simple es una de las habilidades más fundamentales que un niño puede aprender. Sin un alfabetismo elemental hay pocas posibilidades de que el niño pueda escapar al ciclo intergeneracional de la pobreza. Y sin embargo, en muchos países los alumnos escolarizados en la escuela incluso hasta por 6 años son incapaces de leer y entender un texto simple. Evidencias recientes indican que para aprender a leer bien es de esencial importancia aprenderlo tanto *tempranamente* como a un *ritmo* suficiente (con comprensión). La fase de alfabetización se hace más difícil al crecer los educandos; los niños

¹ Programme for International Student Assessment (PISA; un programa de la Organización para la Cooperación y el Desarrollo Económico [OCDE]); Trends in International Mathematics and Science Study (TIMSS); Programme d'Analyse des Systems Educatifs de la Confemen (PASEC); Southern Africa Consortium for the Measurement of Educational Quality (SACMEQ).

² Véase, por ejemplo, el desagregado en percentiles del cuadro D.1 en Mullins, Martin, González y Chrostowski (2004). Podemos derivar conclusiones similares a las del informe OCDE sobre el PISA (OCDE, 2004), por ejemplo en el cuadro 2.1.c. Por lo general sólo los países de ingreso medio participan en estas evaluaciones internacionales. Examinando los pocos países pobres que sí participan en ellas, y vinculando a éstos y a los de ingreso medio que toman parte en evaluaciones regionales como PASEC y SACMEQ, podemos extrapolar que en los países pobres el niño medio debe estar alcanzando alrededor del 3º percentil de la distribución del país desarrollado (Crouch y Winkler, 2007). Por ejemplo, el desempeño medio en el 8º grado en Ghana en TIMSS 2003 fue de 274, pero el desempeño promedio del 5º percentil en los países de altos ingresos fue de 376. En unos cuantos de los países de ingreso medio más avanzados, como Brasil o Túnez, el desempeño medio puede hallarse por encima del 5º percentil de los países de altos ingresos.

que no lo aprenden en los primeros grados tienen mayores probabilidades de repetir y de eventualmente abandonar la escuela. Los esfuerzos globales por ampliar el acceso a la educación podrían verse reducidos si los padres, enfrentados a unas difíciles opciones económicas y sabedores de que los educandos no están adquiriendo habilidades elementales de lectura, retiran a sus hijos de la escuela. En muchos países resulta evidente que esta tendencia ya se está observando: mientras que muchos niños están siendo escolarizados, las tasas de supervivencia y de completar el periodo de la educación primaria (una medida de la propia eficiencia del sistema educativo, así como de la “supervivencia” de los educandos en el sistema) no han mantenido ese ritmo de escolarización.

Midiendo el aprendizaje: el instrumento y el manual

En el contexto de estos interrogantes referidos al aprendizaje de los alumnos y la continua inversión efectuada en educación para todos, los ministerios de educación y los profesionales del desarrollo en el Banco Mundial, la Agencia de Los Estados Unidos para el Desarrollo Internacional (USAID) y otras instituciones pidieron que se crearan medidas simples, efectivas y de bajo costo de los resultados de aprendizaje de los educandos (Abadzi, 2006; Center for Global Development, 2006; Chabbott, 2006; Banco Mundial: Independent Evaluation Group, 2006). Algunos analistas incluso han propuesto la creación de un estándar u objetivo global del aprendizaje, además de las metas ya existentes de Educación para Todos y los Objetivos de Desarrollo del Milenio (Filmer, Hasan y Pritchett, 2006). Si leer bien para cierto grado es materia de debate, la cuestión acerca de medidas del aprendizaje específicas y simples ha sido puesta en la agenda política.

“En algunos países, el 50 por ciento de los educandos en cuarto grado no entiende el significado de los textos que lee (en una clase estatal hallé 20 alumnos que no leían, en una clase de 29), pero la mayoría de ellos asisten a escuelas que atienden a familias en la ‘mitad inferior del nivel de ingresos’. Esto significa que el 90 por ciento de los estudiantes en esta mitad de la población no entiende lo que lee (aunque muchos completan su educación primaria). En esta situación, un buen programa de alfabetización (en los primeros dos grados de la escuela primaria) puede tener un enorme impacto sobre el desempeño del sistema educativo”.

—Ernesto Schiefelbein, ex ministro de educación, Chile

En respuesta a esta demanda se inició el trabajo sobre la creación de un instrumento denominado Early Grade Reading Assessment (EGRA; la Prueba Diagnóstica de Lectura Inicial). Lo que se necesitaba era un instrumento simple que pudiera medir los niveles fundacionales del aprendizaje estudiantil, incluyendo una evaluación de los primeros pasos que dan en el aprendizaje de la lectura: reconocer las letras del alfabeto, leer palabras simples y entender oraciones y párrafos (texto vinculado). El diseño de EGRA se inició en octubre de 2006 cuando USAID, a través de su proyecto EdData II, contrató a RTI International para que diseñara un instrumento con el cual evaluar la lectura en los primeros grados. El objetivo era ayudar a los países socios de USAID iniciar el proceso de medir—sistemáticamente—cuán bien los niños de los grados iniciales de primaria vienen adquiriendo las habilidades de lectura, y en última instancia estimular esfuerzos más efectivos para mejorar el desempeño de esta habilidad nuclear del aprendizaje.

A partir de una revisión de las investigaciones y de las herramientas y evaluaciones de la lectura existentes, RTI diseñó un protocolo para una evaluación oral individual de las habilidades básicas de lectura de los educandos. Para conseguir observaciones a este protocolo y confirmar la validez del enfoque global, RTI convocó una reunión de científicos cognitivos, expertos en la enseñanza de la lectura en los primeros grados, y expertos en metodología de la investigación y en evaluación, para que revisaran los componentes claves propuestos del instrumento. En el taller se encargó a los participantes que cubrieran la brecha

existente entre investigación y práctica, esto es que fusionaran los avances en los campos de la lectura e investigación de la ciencia cognitiva, con prácticas de evaluación alrededor del mundo. Los investigadores y profesionales presentaron evidencias sobre cómo medir la adquisición de la lectura en los primeros grados de primaria. Además se les pidió que identificaran los puntos claves a considerar al diseñar un protocolo de evaluación de la lectura en los primeros grados, para múltiples países y lenguajes. El taller, organizado por USAID, el Banco Mundial y RTI en noviembre de 2006, incluyó a más de una docena de expertos de un grupo diverso de países, así como unos 15 observadores de instituciones tales como USAID, el Banco Mundial, la William and Flora Hewlett Foundation, la Universidad George Washington, el Ministerio de Educación de África del Sur y Plan International, entre otros. Puede encontrarse un resumen de las actas del taller en www.eddataglobal.org, bajo *News and Events*.

En 2007 el Banco Mundial apoyó una aplicación del instrumento de prueba en Senegal (francés y wolof) y Gambia (inglés), en tanto que USAID apoyó su aplicación en Nicaragua (español). Además, gobiernos nacionales, misiones de USAID y organizaciones no gubernamentales (ONG) en África del Sur, Kenia, Haití, Afganistán, Bangladesh y otros países más comenzaron a experimentar con la aplicación de ciertos componentes de la evaluación (con y sin la participación de RTI). Con miras a consolidar estas experiencias y diseñar un enfoque razonablemente estandarizado de la evaluación de la adquisición temprana de la lectura en los niños, el Banco Mundial solicitó que RTI diseñara un manual que guiara a los países que comenzaban a trabajar con EGRA, en áreas tales como la adaptación del instrumento, el trabajo de campo y el análisis de los resultados.

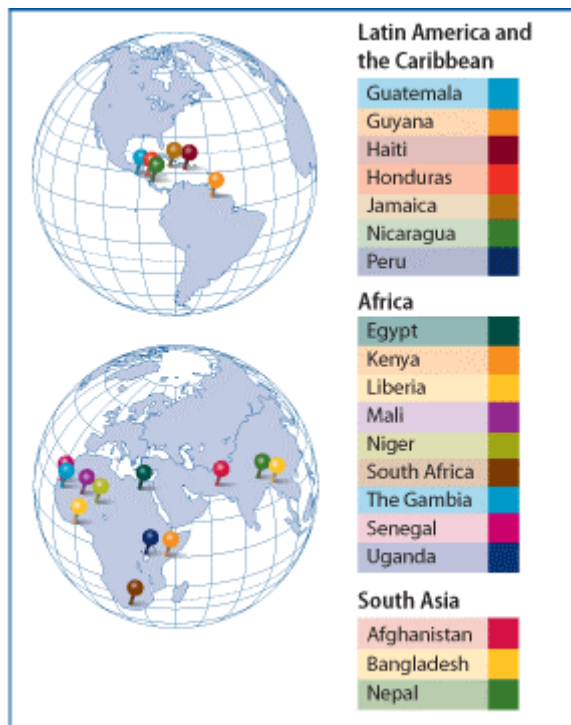
El objetivo de este documento es brindar una guía práctica a los ministerios de educación y sus agencias asociadas, para que respalden la aplicación de EGRA en español. Ello no obstante, con fines ilustrativos se hará referencias ocasionales a su uso en otros idiomas.

Aplicaciones mundiales

El instrumento EGRA que aquí presentamos, así como los resultados obtenidos, generaron una considerable discusión e interés en la comunidad de donantes y entre los ministerios de educación nacionales. En base a los resultados de la aplicación de EGRA en sus países, el personal del ministerio de Gambia y África del Sur diseñó manuales detallados para los docentes, con enfoques pedagógicos, y les capacitaron en su uso. Las ONG internacionales asimismo comenzaron a usar los instrumentos de prueba en sus actividades en el mundo en desarrollo. Plan International, que trabaja en el África francófona, diseñó enfoques de formación de docentes y de enseñanza, y les ha guiado en varios centros educativos de Mali y Níger. Save the Children usó la versión francesa diseñada para Senegal y la adaptó para su uso en Haití en criollo haitiano, y tiene planes para aplicarla en varios otros países.

Otros experimentos con EGRA le añadieron un rigor adicional al modelo. En Kenia, RTI y la Aga Khan Foundation están aplicando evaluaciones de línea de base y de seguimiento en centros educativos de tratamiento y de control. En cada una de las 20 escuelas de tratamiento se está capacitando a los docentes en técnicas de enseñanza de lectura temprana y evaluación continua. Recientemente, una versión adaptada al árabe se aplicó exitosamente en 60 escuelas de Egipto. Se vienen efectuando esfuerzos complementarios en varios otros países (véase el mapa de la gráfica 1, actualizado a febrero de 2009). (Para una lista de países que trabajan con EGRA, por favor véase el enlace “Documents and Data” en la página web www.eddataglobal.org).

Gráfica 1. Sedes de pruebas de EGRA a nivel mundial



Una de las razones del alto nivel del interés en la herramienta EGRA es el enlace directo con los adelantos en la investigación en el desarrollo tanto de la lectura como cognitivo. Si bien gran parte de esta investigación se deriva del trabajo efectuado con niños en países de altos ingresos, la base de tales investigaciones yace en los adelantos en la neurociencia y tiene, por ello, lecciones relevantes para los países de bajos ingresos.³ La importancia de “temprano” debe asimismo enfatizarse acá: las evidencias procedentes de diversos países indican la presencia de lo que Stanovich (1986) llama un “efecto Mateo” en la adquisición de la lectura.⁴ En otras palabras, si no se adquieren unas sólidas habilidades básicas tempranamente, las brechas en los resultados del aprendizaje (entre los ricos y los desposeídos) crecerán más a lo largo del tiempo.

Una segunda razón es la simpleza intuitiva que la medición tiene para el personal del ministerio, docentes y padres de familia. La mayoría de las personas coincidirá en que los niños matriculados en la escuela por 3 años deberían poder leer y entender un texto simple, sin importar el enfoque educativo. La experiencia de la ONG india Pratham, así como los esfuerzos conjuntos del Department for International Development británico (DFID) y el Banco Mundial en Perú, muestran que unas medidas simples pero razonablemente rigurosas de lectura temprana pueden tener un impacto sustancial sobre el diálogo nacional, en relación con la calidad escolar y el aprendizaje de los educandos.⁵

³ Para un resumen razonablemente accesible de los adelantos en la neurociencia y el desarrollo cognitivo, véase Abadzi (2006).

⁴ El término “efecto Mateo”, usado a menudo en el contexto de los estudios de la lectura, y resumido como “los ricos se enriquecen y los pobres se empobrecen”, se deriva de un enunciado que aparece en una parábola bíblica del libro de Mateo: “Porque al que tiene, le será dado, y tendrá más; y al que no tiene, aún lo que tiene le será quitado” (25:29).

⁵ El *Annual Status of Education Report* (2005) de Pratham documenta resultados provenientes de una simple evaluación de lectura y matemáticas, aplicada a 330,000 niños en 10,000 aldeas usando un personal íntegramente voluntario. Para el informe y más información, véase www.pratham.org. En Perú, los esfuerzos del Banco Mundial y DFID llevaron a la inclusión de la calidad escolar y las cuestiones de la lectura temprana en el debate presidencial nacional. En www.eddataglobal.org (página principal) podemos encontrar un enlace con un video preparado con

Por último, EGRA está diseñado para que sea un enfoque de la evaluación independiente del método: no importa cómo se enseña a leer; los estudios muestran que las habilidades evaluadas con EGRA son necesarias pero no suficientes para que los educandos se vuelvan lectores exitosos. Las “guerras” de lectura están vivas y coleando en muchos países de bajos ingresos, a menudo empantanando a los ministerios de educación y centros de enseñanza en debates aparentemente interminables entre los enfoques basados en el “lenguaje integrado” y “basado en el método fonológico”. Con todo, las evidencias sobre la adquisición de la lectura en inglés indican la existencia de cinco componentes esenciales como base, identificados por el U.S. National Reading Panel (National Institute of Child Health and Human Development, 2000): método fonológico, conciencia fonémica, vocabulario, fluidez y comprensión. EGRA utiliza cada uno de estos componentes, enfatizando las habilidades básicas de la adquisición de lectura.

Las ideas sobre cómo seguir mejorando y usando (o no) a EGRA vienen evolucionando. En marzo de 2008 casi 200 participantes, provenientes de unos 40 países, asistieron a un taller de 3 días de duración en Washington, D.C. Los participantes incluían a representantes de las organizaciones y fundaciones donantes, personal de ministerios y ONG, así como expertos internacionales en evaluación y enseñanza de lectura. Los objetivos primarios del taller eran dos. En primer lugar se quería seguir generando interés en, y conciencia de, las actividades de EGRA en la comunidad donante y los posibles países participantes. En segundo lugar se buscaba preparar a un grupo selecto de países interesados para que dieran inicio a aplicaciones concretas de EGRA. Los participantes ganaban un conocimiento técnico de la importancia de la enseñanza y la evaluación de la lectura temprana; comprendían las investigaciones que sustentan EGRA, incluyendo los pasos fundacionales de la adquisición de lectura; y los usos posibles de la información generada por EGRA. Para mayor información sobre el taller y enlaces de video de las presentaciones, véase por favor <http://go.worldbank.org/0SFS7PP330>.

Manual y contenido

Este manual está dividido en siete secciones y está pensado para que lo empleen el personal del ministerio de educación y profesionales en el campo del desarrollo educativo. Se indican públicos más específicos para ciertas secciones del manual, según sea apropiado.

El documento busca resumir un gran corpus de investigaciones en forma accesible, al mismo tiempo que se brindan consejos prácticos y detallados sobre el diseño y la realización de un EGRA a partir de una muestra que sirva de línea de base, con el cual elevar la conciencia y promover el diálogo político.

Este manual no pretende presentar un examen exhaustivo de todas las investigaciones referidas a la lectura. Con miras a la brevedad y la comprensión, no se cubren todos los aspectos y alternativas de la evaluación de la lectura. Debe asimismo indicarse que EGRA es una obra en marcha; los lectores de este manual deberían revisar los instrumentos más recientes y actualizaciones en la página web de EdData. Aún más, EGRA no es una evaluación ya lista; cada nueva aplicación por país requiere de un examen del vocabulario y el diseño de párrafos de lectura apropiados al contexto. El diseño de EGRA en lenguas locales, en particular cuando no se cuenta fácilmente con listas de palabras/vocabulario situadas en un lugar central,

propósito de diálogo político. Para información adicional sobre la evaluación y los resultados de Perú véase Abadzi, Crouch, Echegaray, Paco y Sampe (2005).

requiere de un considerable esfuerzo y debería realizarse conjuntamente con un experto en la lengua local (la traducción directa no es idónea, como se examina detenidamente en una sección posterior. Por último, EGRA está diseñado para complementar, no reemplazar, a las evaluaciones ya existentes que incluyen papel y lápiz con base en el currículo.

Siguiendo a esta introducción, la Sección II es un examen global de los fines y usos de la evaluación. La Sección III incluye el marco conceptual y las bases de la investigación (el sustento teórico de la evaluación). La Sección IV examina los pasos preparatorios de la aplicación de la evaluación, incluyendo un taller de diseño para la elaboración del instrumento EGRA. La Sección V da consejos sobre la selección de la muestra, la formación de los examinadores de EGRA, las realidades que los equipos-país pueden esperar en el campo, y los medios con que recolectar los datos. La Sección VI es un examen global de los análisis a realizar. La Sección VII brinda una guía de la interpretación de los resultados y algunas implicaciones concisas para el diálogo de políticas relacionadas con la mejora de la instrucción y la comunicación de resultados a los centros educativos.

Los anexos incluyen muestras de normas de fluidez en la lectura para inglés por grado, una muestra del cronograma del taller EGRA preparado para Kenia; y muestras de planes de lecciones de docentes basados en los resultados de EGRA. Además se presenta un ejemplo de cómo EGRA puede influir en la política del ministerio de educación con una carta del director de educación de África del Sur. Por último se incluyen anexos técnicos que describen consideraciones referidas al tamaño de la muestra y pruebas de confiabilidad.

Esta versión del manual fue creado para informar sobre el desarrollo y el uso de EGRA en español. USAID financió la preparación de las versiones separadas para francés y español (con una bibliografía apropiada específica a dichas lenguas).

II. Objetivo y usos de EGRA

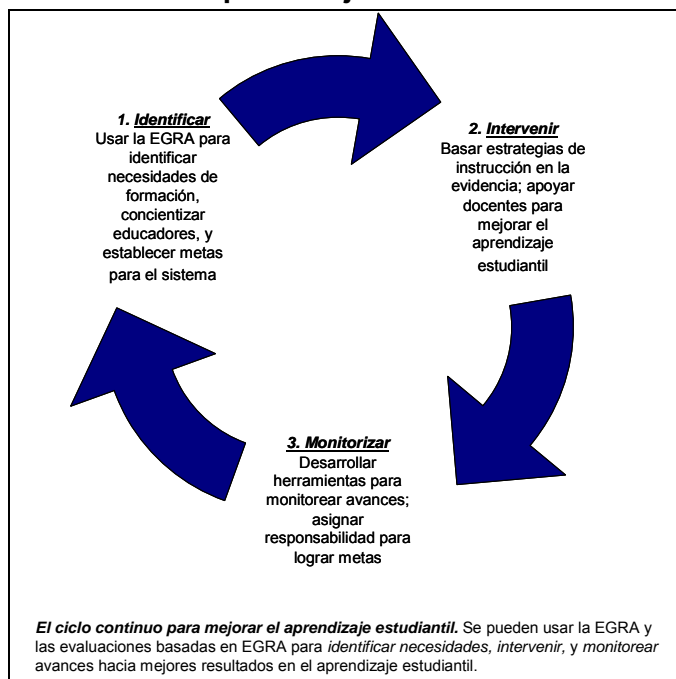
Si bien estuvo claro desde el principio que EGRA se concentraría en los grados iniciales y en las habilidades básicas de la lectura, los usos que se daría a los resultados quedaron, en cambio, más abiertos al debate. Los donantes interesados pidieron su compatibilidad entre países y medidas a nivel del sistema, que pudieran reportar la efectividad de sus inversiones. Los ministerios solicitaron un instrumento que pudiera decirles cómo apoyar a los docentes mediante una formación adicional y otros medios. Por último, los profesores querían una herramienta que les pudiera ayudar a identificar niños individuales que requieren de ayuda adicional, evaluando al mismo tiempo la efectividad de su propia enseñanza. ¿Podía acaso diseñarse un único instrumento para que satisficiera todas estas necesidades?

La respuesta es que “no”. El instrumento EGRA, tal como se desarrolló y se explica en este manual, fue diseñado para que fuera una medida “diagnóstica del sistema” basada en una muestra. Su finalidad es documentar el desempeño estudiantil en las habilidades de lectura en los grados iniciales, para así informar a ministerios y donantes de las necesidades del sistema para mejorar la instrucción. Para ser claros, tal como está diseñado actualmente, EGRA no está pensado para que lo usen los docentes directamente, ni tampoco como un medio con el que identificar alumnos individuales. Y ciertamente no está diseñado para ser una medida de rendición de cuentas de alto impacto con que tomar decisiones de financiamiento, o determinar la aprobación del grado por parte de un alumno. Pero eso no quiere decir que el diseño de una versión del instrumento no pueda informar a otra versión (con una finalidad y uso distintos). Las medidas incluidas en esta versión de EGRA podrían adaptarse para que el docente las use en la identificación de alumnos individuales, y con múltiples formularios equivalentes podría usarse EGRA para supervisar el avance de los educandos dentro de un programa de enseñanza dado. Estos usos alternativos sólo se examinan tangencialmente en el presente manual.

Las medidas de la subprueba EGRA—entre ellas el reconocimiento de letras, la decodificación de pseudopalabras y la fluidez en la lectura oral—han sido usadas para cubrir una amplia gama de necesidades de evaluación, como fines de identificación, diagnóstico y monitoreo del progreso. Usando los resultados de fluidez oral provenientes de miles de alumnos en todo EE.UU. (véase el anexo A), los profesionales e investigadores de la educación examinaron a los alumnos en busca de problemas de aprendizaje, le diagnosticaron fortalezas y debilidades para guiar la enseñanza, y tomaron decisiones referidas a la efectividad de sus programas de formación de docentes y desarrollo profesional. En cada uno de estos casos, el instrumento (y el plan de la muestra) **debe ser** adaptado para que refleje la finalidad de la evaluación (un aspecto crítico a considerar al construir y usar toda herramienta de evaluación) (Hasbrouck y Tindal, 2006; Kame’enui et al., 2006; Kaminski et al., 2006). Se examinan brevemente a continuación las implicaciones de diseño de cada uno de estos enfoques adicionales y las modificaciones requeridas de EGRA.

El diagnóstico basado en el uso del EGRA, tal como se presenta en este manual, está diseñado para cubrir un ciclo completo de apoyo y mejora del aprendizaje. Como se observa en la gráfica 2, EGRA puede usarse como parte de un enfoque global que persigue mejorar las habilidades de lectura de los educandos, siendo el primer paso una *identificación* global, a nivel del sistema, de las áreas en qué mejorar. El establecimiento de criterios generales y la planificación de objetivos para futuras aplicaciones se puede también hacer durante la fase inicial de aplicación del EGRA. En base a los resultados, los ministerios de educación o sistemas locales pueden entonces *intervenir* para modificar programas existentes usando enfoques en enseñanza basados en las evidencias, para apoyar a los docentes en mejorar las habilidades básicas de la lectura. Los resultados de EGRA pueden así informar para el diseño de los programas de formación de docentes que aún no están ejerciendo como también para aquellos que ya están trabajando en el campo profesional.

Gráfica 2. El ciclo continuo para mejorar el aprendizaje estudiantil



Una vez que estas recomendaciones se implementan, pueden usarse formas paralelas de EGRA para seguir el progreso y las ganancias en el aprendizaje de los educandos a lo largo del tiempo mediante un *monitoreo* continuo, con la esperanza de que semejante proceso estimulará la responsabilidad del docente y el administrador de la educación, para asegurarse que los educandos progresen y adquieran las habilidades básicas.

Usando EGRA para identificar las necesidades del sistema

Cuando se trabaja a nivel del sistema, para alcanzar conclusiones sobre cómo éste (o los educandos en su interior) viene desempeñándose, los investigadores y administradores de educación a menudo comienzan con datos a nivel de los alumnos, reunidos en base a una muestra y a los que se asigna un peso apropiado. Usando el desempeño promedio de los educandos por grado, a nivel del sistema, los administradores pueden evaluar en qué parte del sistema educativo es que ellos típicamente tienen problemas, y usar dicha información para diseñar enfoques de enseñanza apropiados. Al igual que toda evaluación cuyo objetivo sea diagnosticar las dificultades y mejorar los resultados del aprendizaje, para que una medida sea útil: (1) la evaluación debe relacionarse con las expectativas y metas del desempeño ya existentes, (2) debe correlacionarse con las habilidades posteriores deseadas, y (3) debe ser posible modificar o mejorar las habilidades mediante una instrucción adicional (Linan-Thompson y Vaughn, 2007). EGRA cumple todos estos requisitos del siguiente modo.

En primer lugar, en muchos países de altos ingresos los docentes (y los administradores del sistema) pueden ver las distribuciones y estándares de desempeño nacionales ya existentes, para entender cómo es que sus alumnos rinden en comparación con otros. Al comparar el

desempeño de subgrupos de educandos en relación con las distribuciones y estándares de desempeño nacionales, los administradores del sistema en EE.UU. y Europa pueden decidir si los centros educativos y los docentes necesitan de un apoyo adicional. EGRA puede asimismo ser usado por países de bajos ingresos para identificar regiones (o centros educativos, si la muestra lo permite) que merecen respaldo adicional, incluyendo la formación de docentes u otras intervenciones.

El problema para los países de bajo ingreso es que aún no cuentan con resultados generados localmente que sirvan de referencia. A partir del criterio basado en el número de palabras leídas correctamente por minuto (PCPM), desarrollados por investigadores y docentes para ser usado en varios países, entre ellos Chile, España y EE.UU. (y como se señala en Banco Mundial: Independent Evaluation Group, 2006; Crouch, 2006), sería posible derivar algunos estimadores para inglés y español, para así efectuar comparaciones más amplias. Las implicaciones que esto tiene se examinan con mayor detenimiento en la Sección VII. En este momento sugerimos que los países trabajen para construir sus propias referencias durante el proceso de aplicación de EGRA.

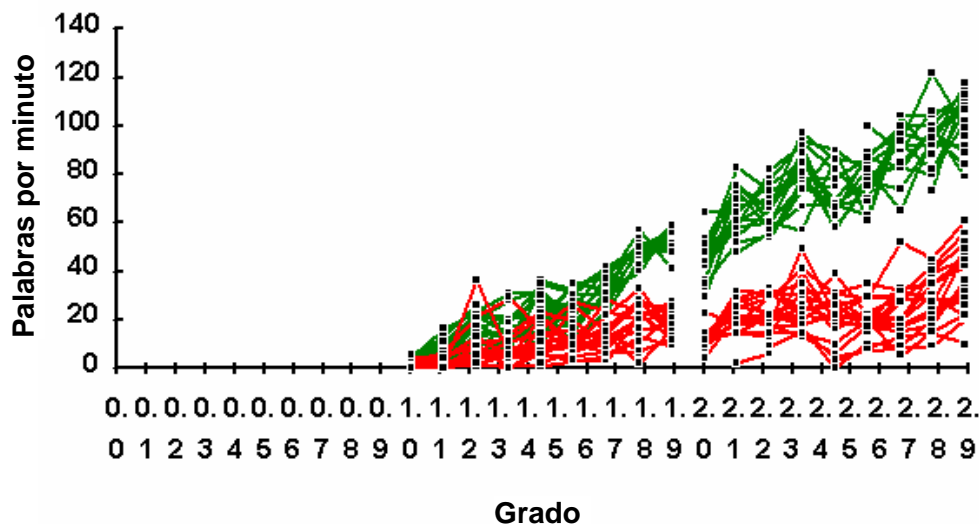
En segundo lugar, para que una medida sea útil para diagnosticar dificultades en la lectura temprana, debe correlacionarse con habilidades de lectura posteriores y deseadas. Por ejemplo, la capacidad de recitar los nombres de todos los presidentes de EE.UU. podría ser una habilidad útil, pero no es probable que se correlacione con las habilidades de lectura. Así como un médico no mide la longitud del pie de un paciente para ver si éste tendrá más adelante una predisposición al cáncer, así también no deseamos diagnosticar problemas en el desempeño de la lectura en base a una medida que no guarda relación alguna con el desempeño subsiguiente, o en los resultados de aprendizaje de los educandos.

Incluso sin investigaciones en los países de bajos ingresos, el amplio uso de las subpruebas o tareas de EGRA en otros países como medidas predictivas puede servir de referencia a aquellos países. Esto es, a pesar de la carencia actual de resultados para los países de bajos ingresos, sabemos lo suficiente sobre la capacidad de estas medidas como para predecir el desempeño posterior de la lectura y las habilidades de nivel más alto, de modo que podemos afirmar con bastante seguridad que el aspecto predictivo de las tareas de EGRA (incluyendo la identificación de letras, lectura de palabras y otras: véase la Sección IV) debería funcionar de igual modo en países de bajos ingresos. Como ejemplo del poder predictivo de las tareas que comprenden EGRA y herramientas similares, se demostró que la fluidez de la lectura oral predecía habilidades posteriores en lectura y comprensión (Fuchs, Fuchs, Hosp y Jenkins, 2001). La importancia de la fluidez como medida predictiva disminuye, sin embargo, con el paso del tiempo. A medida que los educandos se hacen más competentes y la lectura se automatiza, el vocabulario pasa a predecir mucho mejor el éxito académico posterior (Yovanoff, Duesbery, Alonzo y Tindall, 2005).

En tercer lugar, tiene poco sentido medir algo que no tenemos ninguna esperanza de cambiar mediante una enseñanza adicional. EGRA es valioso como herramienta diagnóstica precisamente porque incluye medidas de aquellas habilidades que pueden mejorarse (acompañadas por el apoyo de un docente para la instrucción). La gráfica 3 representa la trayectoria del desempeño estudiantil en la fluidez de la lectura oral, en un grupo de educandos de 1º y 2º grado en EE.UU., entre alumnos que no recibieron instrucción adicional orientada a la mejora de los procesos de lectura. Las líneas verdes en la parte superior del gráfico muestran los resultados mensuales de alumnos que podían leer al menos 40 palabras por minuto al finalizar el primer grado, en tanto que las líneas rojas en la parte inferior son los

resultados de aquellos alumnos que leen menos de 40 palabras por minuto al finalizar el mismo grado (cada unidad en el eje horizontal representa un mes del año escolar).

Gráfica 3. Puntajes de los educandos de palabras por minuto, 1º y 2º grado



Fuente: Good, R. H., III, Simmons, D. C., y Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review*, 27(1), 45–56.

Nota: los números en el eje horizontal se refieren al grado (hilera superior) y mes (hilera inferior).

Como vemos en la gráfica 3, en ausencia de una intervención adicional y de reforzamiento, la brecha entre los lectores tempranos hábiles y los menos competentes crece dramáticamente hacia finales del primer grado (y continúa creciendo con el tiempo). De este modo queda claro el uso de EGRA para motivar la necesidad de contar con enfoques de instrucción efectivos. Si los docentes y administradores no intervienen, es probable que las brechas iniciales en la adquisición de lectura crezcan con el tiempo. Si bien el examen de estrategias con que mejorar la instrucción en cada una de las áreas identificadas por EGRA cae fuera del alcance de este manual, pueden encontrarse guías del maestro y planes de lecciones en varias fuentes (véase “Cambiano la enseñanza de la lectura” en la Sección VII).

Usos adicionales de EGRA (con modificaciones)

Identificación

Podría usarse una versión modificada de EGRA a nivel del aula para indicar qué educandos tienen problemas, y permitir que los docentes cambien la trayectoria de los resultados de los alumnos mediante intervenciones específicas en las áreas requeridas. Para usar EGRA a nivel del aula en el examen individual de los educandos, deben hacerse tres cambios importantes.

En primer lugar se eliminaría el plan del muestreo descrito detalladamente en el anexo B, pues cada docente aplicaría el instrumento EGRA para cada educando (e identificaría a los que tienen dificultades, diseñando enfoques para la recuperación).

En segundo lugar tendría que dividirse EGRA en subpruebas o subtareas, a aplicarse (o no) dependiendo de la evaluación que el docente haga de las habilidades y el nivel del grado de

cada educando. Actualmente EGRA incluye ocho subtarefas y está diseñado para ser aplicado a una muestra de educandos en 1°-3° grado al finalizar el año escolar (o en 2°-4° grado al comenzar el año escolar). Para facilitar la aplicación, y como EGRA es llevado a cabo por examinadores que no conocen a los educandos individuales (y que por ende no tienen ningún conocimiento previo de su desempeño), el diseño actual del instrumento requiere que todos los alumnos intenten hacer todas las partes de la evaluación. En un instrumento aplicado por un docente sólo sería necesario aplicar algunas de estas tareas a cada educando. Por ejemplo, los docentes podrían comenzar evaluando a los alumnos con la tarea de las palabras y si éstos tienen éxito, pasarían entonces a la parte de los párrafos de texto de la evaluación. Alternativamente, si un alumno no puede leer letras, podría entonces probarse la comprensión oral y no habría necesidad de efectuar el resto de la prueba. Dicho esto, los diseñadores deberían ser juiciosos al decidir qué medidas usar; en varios de los países piloto, niños incluso del tercer grado no dominaban las habilidades básicas (como la identificación de letras).⁶

En tercer lugar, un EGRA aplicado por los docentes requeriría que se les capacite en la aplicación e interpretación de las tareas del mismo, en métodos de evaluación continua, el diseño de enfoques de instrucción para la recuperación, y estrategias para trabajar con educandos con problemas de lectura. Podría obviarse el uso de un instrumento EGRA, si se capacita a los docentes para que diseñen sus propias evaluaciones simples de las habilidades de lectura de los educandos.

Los experimentos realizados en Kenia, Mali y Níger, usando la formación explícita de los docentes en cómo usar una versión modificada de EGRA para entender cómo vienen desempeñándose los educandos, acompañado por planes de lección detallados y con una buena secuencia, tuvieron como resultado mejoras significativas tanto en la práctica de los docentes como en los resultados de lectura de los educandos. Para un ejemplo del enfoque basado en los docentes, aplicado en inglés, véanse los materiales preparados para Kenia en www.eddataglobal.org (Documents and Data>Kenya). Los materiales de Mali y Níger están disponibles en Plan International.

Evaluación de las intervenciones

Para que EGRA sea útil evaluando una intervención dada, debe estar diseñado de modo tal que se pueda evaluar a los educandos múltiples veces utilizando instrumentos paralelos (para evitar la memorización o que se aprendan los extractos del instrumento). Actualmente, este enfoque ha sido probado en Liberia (con formularios paralelos diseñados y efectuados como piloto en los mismos centros educativos, con planes para su uso en múltiples años). Sería de esperar que los educandos, evaluados al inicio de una intervención y luego de la intervención de enseñanza, mostraran una mejora mayor que la de un grupo similar de alumnos de control. Las intervenciones a evaluarse deben guardar relación con los resultados de lectura y aprendizaje; no tendría mucho sentido evaluar un programa al cual no se puede teorizar para mejorar así los resultados de lectura. Por ejemplo, uno que buscara mejorar la administración de base escolar probablemente no podría demostrar mejoras en la lectura debido a la ausencia de un vínculo causal directo. En otras palabras, EGRA puede usarse para evaluar un programa tal, pero dichas evaluaciones probablemente no revelarán mejoras demostradas en los resultados de lectura. Esto quiere decir que sólo se debería usar para evaluar programas que

⁶ Tal como está diseñado actualmente EGRA es para usarse con niños en los primeros grados, y las evaluaciones de la fluidez de la lectura oral han sido usadas en EE.UU. y otros lugares todavía en 6° grado. La fluidez oral en grados posteriores debe llevarse a cabo con textos de mayor longitud que aquellos incluidos en EGRA (para información adicional sobre la fluidez de la lectura oral en grados posteriores, por favor véase Espin y Foegen, 1996; Espin y Tindal, 1998).

busquen mejorar la enseñanza de la lectura y los resultados del aprendizaje en los grados de primaria.

Cómo no debe usarse EGRA

EGRA no es una herramienta de rendición de cuentas de alto impacto

No debe usarse EGRA para reforzar positivamente o negativamente al educando, es decir, para suministrar premios y castigos. Debe considerársele más bien como una herramienta diagnóstica cuyos principales clientes y usuarios son el personal del ministerio. La realidad es que una vez que comienza a usarse un instrumento de este tipo, y si las comunidades asumen que este tipo de instrumento que está fundamentado en la investigación científica puede contribuir a mejorar los procesos de comprensión de lectura en los educandos, será inevitable—y deseable—que los padres de familia y las comunidades se involucren en la supervisión el monitoreo del avance en la lectura. Es igualmente inevitable que los funcionarios, hasta el mismo ministro de educación, adquieran cierto interés por saber cuán bien se vienen desempeñando los niños, y que surgirá alguna necesidad de presentar un informe. Para reiterarlo, sin embargo: Hay que evitar el uso de EGRA como una herramienta de rendición de cuentas de alto impacto, como los planes de premios para docentes. Tal como se halla diseñado actualmente (un diagnóstico a nivel del sistema), EGRA no debería identificar educandos o docentes para un seguimiento subsiguiente, siendo la única excepción posible el monitoreo confidencial de profesores y alumnos para los fines de identificación y de evaluaciones de proyectos.

EGRA no resulta idóneo para las comparaciones entre lenguas

La cuestión de efectuar comparaciones entre países resulta un desafío desde la perspectiva de una evaluación. Si bien todas las evaluaciones específicas a países diseñadas hasta la fecha parecen ser sumamente similares, incluso entre lenguas, y parecerían a primera vista ser comparables, lo cierto es que las diferencias en la estructura de la lengua y su tasa de adquisición desalientan las comparaciones directas. Las investigaciones indican que las diferencias entre lenguas podrían deberse fundamentalmente a la *tasa* a la cual los niños dan los primeros pasos hacia la adquisición de la lectura (Seymour, Aro y Erskine, 2003). Dejando de lado la lengua, todos los niños que aprenden a leer avanzan de ser no-lectores (incapaces de leer palabras) a lectores parciales (pueden leer algunas cosas pero no otras) y finalmente a lectores (pueden leer todas las cosas o la mayoría de ellas). La progresión a través de estos niveles en las lenguas con ortografías transparentes o “superficiales” (llamadas a menudo lenguas escritas fonéticamente) es muy rápida (a menudo de apenas unos cuantos meses de aprendizaje); este proceso puede tomar varios años en lenguas con ortografías más complejas o “profundas”. En el inglés, por ejemplo, completar la fase alfabética requiere de dos o más años, con una tasa de ganancia de apenas unas cuantas cosas nuevas por mes de aprendizaje; en comparación, las lenguas regulares y transparentes como el español, italiano, finlandés y el griego sólo requieren alrededor de un año de enseñanza a los educandos para que alcancen un nivel comparable (Seymour et al., 2003).

De este modo, no debería usarse EGRA para comparar resultados entre lenguas. Como éstas tienen distintos niveles de transparencia ortográfica, no sería justo decir que el País A (en el cual todos los niños están leyendo con automatismos para el 2º grado) tiene un mejor desempeño que el País B (donde los niños sólo alcanzan este nivel para el 3º grado) si la lengua del primer país cuenta con una ortografía más transparente que la del segundo. Ello no obstante, encontrar a qué grado los niños usualmente “saltan” a la lectura en diversos países, y

comparando dichos grados, sería un útil ejercicio analítico y de política, siempre y cuando no se le use para formar un “ranking” o “tablas de ligas”, o para establecer un único estándar universal para, digamos, fluidez o automatismo de la lectura. De este modo, si la población de un país habla una lengua con una ortografía transparente y el automatismo se adquiere dos grados después que en países con lenguas similarmente transparentes ortográficamente, esto debería ser motivo de análisis y de examen. El afinamiento de las expectativas y los objetivos en distintos países, pero dentro de una misma lengua, forma parte de las finalidades de EGRA y es algo que probablemente captará la atención a nivel mundial. En efecto, ya se están llevando a cabo algunos intentos tempranos no sólo para medir, sino también para establecer algunos estándares simples (Banco Mundial, 2007). En esta etapa seguimos explorando las implicaciones que tendrían las comparaciones entre países (pero dentro de una lengua), lo que requeriría de investigaciones y debates adicionales.

Dentro de una misma lengua, el desafío de efectuar comparaciones entre países no es tanto una cuestión del diseño de un instrumento comparable (una dificultad importante pero superable), como de diferencias entre los dialectos y vocabularios locales. Por ejemplo, se diseñaron instrumentos para Perú y Nicaragua usando virtualmente la misma versión del instrumento (con modificaciones locales menores en las instrucciones del administrador). La comparación de dichos resultados ayudará a establecer si tales comparaciones son posibles o deseables.

Si bien los expertos no recomiendan comparar entre lenguas usando algún tipo de estándar universal, digamos las palabras leídas correctamente por minuto para finales del 2º grado, sí parecen ser posibles las comparaciones aproximadas dentro de las lenguas; y en todo caso las comparaciones elementales sí deberían ser posibles para cosas tales como la fluidez del reconocimiento de las letras en lenguas que usan una escritura latina.

III. Marco conceptual e investigaciones de base

El marco conceptual del aprendizaje de la lectura que subyace al diseño del EGRA, tiene su fundamento en el trabajo realizado por el U.S. National Reading Panel (National Institute of Child Health and Human Development, 2000), el National Literacy Panel (2004) y el Committee on the Prevention of Reading Difficulties in Young Children (Snow, Burns y Griffin, 1998), entre otros. Cada uno de estos informes revisa los principales componentes involucrados en el aprendizaje de la lectura y la instrucción.

En muchos de estos informes se revisó la investigación realizada hasta ese entonces en el campo de la lectura y la instrucción, pero dichos exámenes correspondían fundamentalmente a estudios en inglés. El National Reading Panel asimismo recabó información en todos aquellos sectores involucrados en esta área: profesores, padres de familia, estudiantes, universidades, expertos en política educativa, científicos, asociaciones, etc. La publicación de este informe sirvió como precedente para que en muchos estados de Norteamérica se comenzaran a implementar programas de enseñanza de la lectura basados en investigaciones científicas (v.g., *Texas Reading First Initiative*,⁷ *Oregon Reading First*⁸).

Los dos principios fundamentales que se derivan de la bibliografía científica revisada, y que sirven de fundamento al desarrollo del EGRA, son los siguientes. En primer lugar, la evaluación de la lectura (y sus componentes, o la instrucción) es compleja, pero se cuentan con suficientes investigaciones como para apoyar o fundamentar el desarrollo de instrumentos de evaluación específica, que nos permiten identificar qué tipo de habilidades son las que los estudiantes necesitan desarrollar para así convertirse en buenos lectores, *prescindiendo del método de enseñanza*. En segundo lugar, las habilidades de la lectura temprana se adquieren a lo largo de distintas fases; el nivel de complejidad de una lengua afecta el modo en que los estudiantes adquieren estas habilidades a temprana edad.

Evaluación inicial del aprendizaje de la lectura

La evaluación inicial de la lectura es un proceso complejo, pero gracias a la bibliografía científica sabemos cuáles son las principales habilidades en donde deberíamos concentrar más nuestra atención.

El lenguaje oral y el escrito se hallan estrechamente relacionados, como lo muestra el “modelo simple de lectura” (Gough, 1996). Según este modelo, la comprensión es el resultado de la descodificación, la comprensión oral y la velocidad. El modelo simple de lectura establece que la comprensión requiere: (1) comprensión general a nivel lingüístico y (2) habilidad para identificar las palabras impresas con exactitud y fluidez. Sin embargo, las expresiones más recientes de este modelo subrayan la importancia que la velocidad o la fluidez tienen como factores explicativos independientes (véase Carver, 1998; Catts, Hogan y Fey, 2003; Chiappe, 2006; Cutting y Scarborough, 2006; Ehri, 1998; Gough y Tunmer, 1986; Hoover y Gough, 1990; Joshi y Aaron, 2000; Share, 1999; Sprenger-Charolles, Siegel, Béchenec y Serniclaes, 2003). Por otra parte, y tal como Torgesen (1998) lo sugiere, una comprensión lectora eficiente sería el producto final de una instrucción en lectura que ha sido efectiva.

Según esto, para conseguir una adecuada comprensión lectora es importante que el alumno consiga identificar las palabras escritas con cierta fluidez y automaticidad. Hasta ahora, la

⁷ Para más información consúltese <http://www.texasreading.org/utclra/pd/trfi.asp>.

⁸ <http://oregonreadingfirst.uoregon.edu/>.

investigación experimental sobre cómo identificamos las palabras escritas en español sugiere que las palabras más frecuentes se identifican más rápido que las menos frecuentes, que la longitud también influye, y que la estructura ortográfica o la existencia de vecinos ortográficos o fonológicos influye también sobre la duración y la precisión del proceso de identificación (Jiménez y Hernández-Valle, 2000). Para explicar estos resultados se ha sugerido que cuando identificamos una palabra escrita, lo hacemos a través de una conexión entre su forma ortográfica y los conocimientos con que el lector cuenta en su memoria léxica, referidos a esta palabra (i.e., conocimientos acerca de su significado, su pronunciación, etc.). Y esta conexión la podemos establecer de dos formas distintas: una, que sería directa y se produciría cuando la palabra que tenemos que identificar forma parte de nuestro repertorio de palabras ya memorizadas, es la que denominamos *léxico ortográfico*, en tanto que la otra forma sería la del *ensamblaje fonológico*, que consiste en traducir los elementos ortográficos de la palabra (i.e., letras o grupos de letras) a la forma fonológica correspondiente. El procesamiento de las palabras escritas es una actividad automatizada en la cual las cualidades intelectuales del lector poco o nada tienen que ver con este proceso (Jiménez, Siegel, O'Shanahan y Ford, 2009).

Llegados a este punto, ¿qué habilidades se debe instruir para que el alumno prelector se convierta en un lector experto? A partir de la revisión hecha de las investigaciones y de consultas efectuadas con expertos, etc., los miembros del National Reading Panel (National Institute of Child Health and Human Development, 2000) describieron cuáles son los principales componentes o habilidades que la instrucción de la lectura debería incluir, para conseguir tales objetivos:

- (1) Conciencia fonológica (CF) – La instrucción en CF consiste en enseñar a los niños a segmentar el lenguaje oral en fonemas, sin contar con el apoyo de las letras del alfabeto.
- (2) Conocimiento alfabético y aprendizaje de las correspondencias grafema-fonema (CGF). La instrucción en el uso de reglas de CGF es una forma de enseñar a leer que enfatiza el conocimiento de las correspondencias letra-sonido de manera explícita, sistemática y temprana o de forma incidental. En el primer caso, desde que se inicia la enseñanza el profesor ayuda al alumno a comprender la relación que existe entre grafemas y fonemas, sin esperar a que éste la descubra de forma espontánea, y para ello sigue un orden preciso. En cambio, en el segundo caso, al enseñar el profesor no sigue una secuencia de instrucción planificada de las reglas de CGF, sino que éstas son enseñadas a medida en que van apareciendo en los textos.
- (3) Fluidez – La instrucción está orientada a estimular la lectura con velocidad, precisión, y respetando los signos de puntuación, ya que facilita la comprensión del texto. Si la lectura es muy laboriosa y poco precisa, será más difícil recordar y relacionar las ideas expresadas en el texto con los conocimientos previos que posee el alumno.
- (4) Vocabulario – La instrucción se centra en dos tipos de vocabulario, el oral y el escrito. Cuando un lector encuentra una palabra en el texto puede descodificarla, es decir, convertirla en habla. Si se encuentra dentro del repertorio lingüístico que posee el lector, éste será capaz de entenderla; en cambio, si no está, el lector tendrá que deducir el significado por otros medios y otras estrategias. En consecuencia, cuanto mayor sea el vocabulario del alumno (ya sea oral u escrito) tanto más fácil será la comprensión del texto.
- (5) Comprensión - La instrucción ha de procurar que los alumnos sean capaces de relacionar las ideas que están representadas en el texto con sus propios conocimientos y experiencias, así como representaciones mentales construidas en la memoria. Se sugiere que la comprensión podrá mejorarse en la medida en que enseñemos a los

alumnos a usar estrategias cognitivas específicas o a razonar de manera estratégica, cuando se topen con barreras para comprender lo que están leyendo. Los alumnos hasta cierto punto adquieren estas estrategias de manera informal, ya que una enseñanza formal y explícita de las mismas conduce a una mejora del proceso de comprensión.

A estas habilidades habría que sumar otras no menos importantes, como la conciencia general del lenguaje escrito. Hay que tener en cuenta que esta revisión realizada por el National Reading Panel corresponde a estudios realizados en lengua inglesa, razón por la cual hay aspectos que podrían no ser directamente extrapolables a otras lenguas en donde la escritura refleje la fonología de manera más fiel, como sería el caso del español. En otras palabras, aunque también se ha demostrado que estas habilidades son importantes cuando se aprende a leer en español, hay, sin embargo, algunos matices que conviene tener en cuenta dependiendo del contexto idiomático, y que exponemos a continuación (para un examen más exhaustivo véase Jiménez y O'Shanahan, 2008).

Conciencia general del lenguaje escrito

La conciencia general del lenguaje escrito incluye diferentes aspectos tales como el reconocimiento de las actividades de lectura y escritura, la comprensión de las funciones del lenguaje escrito y la conciencia de las características específicas de este lenguaje. Es cierto que los niños pueden *desarrollar concepciones propias acerca del lenguaje escrito* antes de iniciar la experiencia formal del aprendizaje y enseñanza de la lectura, y que se debería enfatizar el significado de lo que se aprende. De hecho, existen ciertas evidencias empíricas que demuestran la importancia que tienen las actividades orientadas hacia el significado en la enseñanza de la lectura. Por ejemplo, Stahl, McKenna y Pagnucco (1994) encontraron que algunas actividades desarrolladas en las aulas, como el diálogo sobre los propósitos de la lectura y la escritura, el familiarizarse con los relatos a través de los cuentos e historias que les leen los adultos, etc., pueden mejorar las actitudes para con, y orientación hacia, la lectura. Además, un número creciente de estudios ha demostrado que los niños tienen cierto conocimiento de los objetivos de la lectura y escritura, incluso antes de aprender a leer. Sin embargo, también se ha demostrado la presencia de una tendencia temprana en los niños que les hace buscar regularidades entre fonología y ortografía, y de algunos aspectos relacionados con la naturaleza convencional de la escritura, como la direccionalidad:

Los niños aprenden la forma gráfica de las palabras y su disposición en las páginas; también son conscientes de ciertas características de las letras y son capaces de nombrar algunas de ellas. Ellos pueden decir la diferencia que existe entre una palabra, una letra y un número. (Gombert, 1992, p. 152)

Ortiz y Jiménez (2001) analizaron la naturaleza, evolución y estructura de los conocimientos previos sobre el lenguaje escrito en niños prelectores hispanohablantes. Para ello administraron la prueba de Conocimientos sobre el Lenguaje Escrito (CLE) (Ortiz y Jiménez, 1993) a niños prelectores. Encontraron que el máximo rendimiento de los alumnos se daba en las tareas referidas al reconocimiento de las actividades de lectura y escritura, y al reconocimiento de la lectura como un instrumento de diversión y ocio. De todo ello se sigue la conclusión de que la motivación del aprendizaje de la lectura y escritura ha de ser fomentada procurando que los niños descubran las funciones y usos de la lengua escrita.

Conocimiento alfabético y aprendizaje de las correspondencias grafema-fonema

Numerosas investigaciones han demostrado que el conocimiento de las letras del alfabeto, antes de su enseñanza formal, es también un excelente predictor del aprendizaje de la lectura, a corto y a largo plazo (Muter, Hulme y Snowling, 1997; Vellutino y Scanlon, 2002; Whitehurst y Lonigan, 1998). El conocimiento espontáneo de las letras revela que los niños pueden haberlas aprendido implícitamente sin una intervención escolar explícita, probablemente por influencia de su entorno. Conocer su nombre o pronunciación en un sistema alfabético implica identificarlas mentalmente como unidades de fonemas. Riley (1999) demostró la importancia que el conocimiento ortográfico alcanzado por los niños antes de la instrucción escolar formal, tiene para identificar las letras y para escribir su propio nombre. Los datos de su investigación señalan que con este medio es posible identificar a los niños que aprenderán a leer al finalizar el primer año, con un 80% de precisión.

La importancia del conocimiento de las letras para el aprendizaje del lenguaje escrito al inicio del primer grado, ha quedado corroborada en varios estudios realizados con niños chilenos. Allí se encontró que los niños que conocían al menos 5 letras del alfabeto cuando ingresaron al primer curso, tuvieron posteriormente un nivel lector significativamente superior al del resto de sus compañeros (Bravo-Valdivieso, Villalón y Orellana, 2001, 2002). Más recientemente se llevó a cabo un estudio de tres años de seguimiento con 262 niños chilenos de estatus socioeconómico bajo, los cuales fueron evaluados en algunos procesos psicolingüísticos cuando ingresaron al primer año. Los resultados mostraron que el reconocimiento del primer fonema de las palabras, la identificación de algunos nombres propios y el conocimiento de algunas letras del alfabeto predecían significativamente el rendimiento de la lectura en el primer, segundo y tercer año, y además eran útiles para discriminar entre niños con un rendimiento alto y bajo (Bravo-Valdivieso, Villalón y Orellana, 2006). Según los autores, esta persistencia de la posibilidad de predecir durante tres años estaría indicando también que dichas habilidades continuaron ejerciendo su influencia sobre el aprendizaje del lenguaje escrito durante las etapas posteriores al proceso de descodificación inicial.

Por lo tanto, aprender a leer en un sistema alfabético consiste en la adquisición de las correspondencias entre la forma ortográfica y la forma fonológica. Una vez que se aprenden las reglas de CGF se puede leer cualquier palabra, al menos en aquellos sistemas alfabéticos en los cuales existe una alta regularidad entre la forma ortográfica y la forma fonológica, como es el caso de la lengua española.

Aquí han sido numerosos los estudios que han intentado analizar si los métodos de enseñanza de la lectura pueden influir sobre el desarrollo de estos procesos de lectura (para una revisión véase Jiménez y Guzmán, 2003). Los niños y niñas que aprenden a leer siguiendo métodos globales usan una estrategia global de lectura, y los que aprenden a leer con métodos fonéticos usan estrategias fonológicas. Sin embargo, la mayoría de estos estudios fueron realizados en lenguas que tienen una ortografía opaca (i.e., el inglés). Esto significa que los resultados no deberían ser extrapolables a otras lenguas en lo que respecta al reconocimiento de las palabras, puesto que hay diferencias en la manera de codificar la fonología a través de la ortografía. Por tal motivo son de especial relevancia los estudios realizados en las lenguas que tienen ortografía transparente, como la lengua española. Por ejemplo, Jiménez y Guzmán (2003) llevaron a cabo una investigación con la finalidad de examinar las estrategias de lectura usadas por los niños hispanohablantes que aprendían a leer a través de diferentes métodos de instrucción. Se utilizó un diseño transversal y se llevaron a cabo tres estudios con una muestra de 202 niños, que aprendían a leer siguiendo diferentes métodos (fonético vs. global). Los estudios 1 y 2 fueron diseñados para analizar los efectos que los métodos de enseñanza tienen

sobre los tiempos de reacción (TR) y sobre los tiempos de latencia (TL), a través de tareas de decisión léxica y nombrado. Las palabras estaban agrupadas en función a la frecuencia, la longitud y la frecuencia silábica posicional (FSP) (i.e., el número de veces que una sílaba se repite en determinada posición de la palabra), y las pseudopalabras en función a la longitud y la FSP. El estudio 3 consistió en el análisis de los efectos que los métodos de enseñanza tienen sobre los errores en la lectura oral de palabras y pseudopalabras. La predicción en dicho estudio fue que las diferencias en TR así como en los errores—en función a las variables que permiten evaluar las estrategias de lectura de palabras (i.e., frecuencia, FSP y longitud)—serían mayores en aquellos niños que aprenden a leer siguiendo métodos globales. Ello significa que estos niños quedarían afectados más por palabras que no son familiares, largas, de baja FSP y pseudopalabras, en comparación con los que aprenden a leer con el método fonológico. Esto apoyaría la teoría de que los niños que aprenden a leer con el método global tienen dificultades para nombrar palabras bajo condiciones que exigen un procesamiento fonológico. Se encontraron los efectos que los métodos de enseñanza tienen tanto sobre los tiempos de reacción como sobre los tiempos de latencia, y también sobre los errores en la lectura de palabras y pseudopalabras. Se halló un efecto de superioridad en el análisis subléxico en niños que aprenden a leer a través del método fonológico. Sin embargo, los niños que aprenden a leer con el método global experimentaron mayores dificultades para nombrar palabras cuando intervenían los procesos de mediación fonológica.

Por otra parte, Jiménez, Guzmán y Artilés (1997) diseñaron un experimento para analizar el efecto de la frecuencia silábica posicional, en el cual participó una muestra de 252 sujetos que aprendían a leer con distintos métodos de lectura (global vs. fonético). De este modo se halló un efecto consistente de la FSP tanto sobre los tiempos de reacción como de latencia, así como sobre los errores en la lectura oral de las pseudopalabras. Se registró un mayor número de errores en la lectura de pseudopalabras de baja FSP, por parte de alumnos que aprendían a leer con el método global. Ello significa que dichos niños no eran tan eficientes en el procesamiento fonológico, puesto que las pseudopalabras que contienen sílabas de baja frecuencia se leen por la ruta fonológica.

En líneas generales, los hallazgos aquí comentados sugieren que la instrucción debería estar orientada hacia el conocimiento fonológico y la codificación alfabética. En este sentido la enseñanza de lo fonológico, al enfatizar las correspondencias letra-sonido, produce un aceleramiento en la adquisición de correspondencias subléxicas. El reconocimiento visual de las palabras constituye una habilidad importante en el desarrollo de la lectura, ya que explica un alto porcentaje de varianza en el desarrollo de la habilidad lectora temprana. Estos resultados asimismo nos sugieren que el reconocimiento de palabras es un proceso modular altamente automatizado, que no requiere de información contextual (entiéndase de naturaleza lingüística) para su ejecución. Si el contexto tuviera alguna influencia en la enseñanza de la lectura, no tendrían por qué aparecer diferencias entre los distintos métodos de lectura en el procesamiento léxico.

En definitiva, tanto los estudios realizados en lenguas opacas (v.g., Byrne y Fielding-Barnsley, 1991; Foorman, Francis, Novy y Liberman, 1991) como aquellos llevados a cabo en sistemas ortográficos transparentes (v.g., Jiménez y Guzmán, 2003; Jiménez, Guzmán y Artilés, 1997; Jiménez y Rumeu, 1989; Wimmer y Hummer, 1990), pusieron de manifiesto una superioridad en el análisis subléxico cuando se empleó una instrucción basada en la enseñanza de reglas. Sin embargo, si bien se encontró en la ortografía opaca que los niños adoptan estrategias de lectura diferentes en función al método de lectura (v.g., Barr, 1974; Bradley y Bryant, 1985; Sowden y Stevenson, 1994), en los estudios realizados en español no se hallaron tales diferencias.

En definitiva, todos estos resultados sugieren que la instrucción fónica sistemática aumenta el éxito de los niños en el aprendizaje de la lectura, y resulta significativamente más efectiva que aquella que incluye poca o ninguna instrucción fónica.

Conciencia fonológica

Es evidente que si el sistema de escritura exige atribuir fonemas a las letras o grupos de letras, entonces para aprender a hacer esta operación el alumno tiene que comprender el código alfabético, lo que exige haber aislado mentalmente los fonemas. En la última década se han venido acumulando suficientes evidencias empíricas que demuestran la enorme importancia que la conciencia fonológica tiene al inicio del aprendizaje de la lengua escrita. Se ha puesto de manifiesto, tanto en estudios longitudinales como de corte transversal, que estas habilidades facilitan la adquisición de la lectura y la escritura. En cierto modo, el análisis de la estructura sonora del habla lo demanda la propia naturaleza de nuestro sistema alfabético, ya que no debemos olvidar que *la escritura representa los sonidos del habla*, esto es, cada grafema es utilizado para representar a un fonema. Los fonemas están ocultos en el habla y para que los niños puedan acceder a estas unidades se requiere una instrucción sistemática.

La formación de representaciones fonológicas no se da por el mero hecho de estar expuestos a la lengua oral. Contamos con evidencias empíricas en diferentes lenguas—portugués (Bertelson, De Gelder, Tfouni y Morais, 1989; Morais, Bertelson, Cary y Alegría, 1986; Morais, Cary, Alegría y Bertelson, 1979), chino (De Gelder, Vroomen y Bertelson, 1993; Read, Zhang, Nie y Ding, 1986), francés (Gombert, 1994) y español (Adrian et al., 1995; Jiménez y Venegas 2004)—de que los adultos iletrados tienen dificultades para resolver tareas que requieren la manipulación de fonemas a pesar de tener experiencia con el habla. Jiménez, García y Venegas (en prensa) recientemente concluyeron también que al estudiar adultos iletrados, la experiencia con el habla no bastaba para tomar conciencia de la estructura sonora del lenguaje. El rendimiento de los adultos estaba afectado más por el tipo de tarea y el tipo de estructura silábica, en comparación con niños que habían adquirido el código alfabético. Además, son precisamente los *problemas de conciencia fonológica y de procesamiento fonológico* los que caracterizan la dificultad de aprendizaje que muchos niños experimentan cuando no han conseguido dominar el código alfabético.

Cuando hablamos de una conciencia fonológica podemos estar nos refiriendo a distintos niveles, según la unidad fonológica que estemos considerando. Así, por ejemplo, se entiende por conciencia silábica la habilidad para segmentar, identificar o manipular conscientemente las sílabas que componen una palabra (v.g., contar las sílabas que ésta contiene). La conciencia intrasilábica se refiere a la habilidad para segmentar las sílabas en sus componentes intrasilábicos de principio y rima. El principio es una parte integrante de la sílaba conformada por la consonante o el bloque de consonantes inicial (v.g., /fl/ en flor). La otra parte de la sílaba es la rima, formada por la vocal y la consonante siguientes (v.g., /or/ en flor). La rima a su vez está constituida por un núcleo vocálico (v.g., /o/ en flor) y la coda (v.g., /r/ en flor). Y la conciencia fonémica es la habilidad metalingüística que implica la comprensión de que las palabras habladas están conformadas por unidades sonoras discretas, que son los fonemas (Ball, 1993); esto es la habilidad de prestar una atención consciente a los sonidos de las palabras como unidades abstractas y manipulables (Adams, 1990).

Los estudios realizados en español pusieron de manifiesto que los niños españoles tienen cierto grado de conciencia de las unidades fonológicas del lenguaje antes de aprender a leer, en particular de las unidades silábicas. Analizando el rendimiento de niños prelectores se

demuestra que éstos son capaces de resolver tareas de segmentación silábica con un alto porcentaje de aciertos, algo que no ocurre en otras tareas de conciencia fonológica (v.g., con las unidades de rima y fonemas) (Carrillo, 1994; Carrillo y Marín, 1996; Defior y Herrera, 2003; Domínguez, 1996a, 1996b; González, 1996; Herrera y Defior, 2005; Jiménez, 1992; Jiménez y Ortiz, 2000). Este tipo de hallazgo guarda relación con el carácter marcadamente silábico del español, que es una lengua *con una estructura silábica bien definida*, y también con el hecho de que las sílabas están marcadas acústicamente y se pueden articular de forma independiente. Por lo tanto, la fuerte delimitación silábica del español y la prominencia de la sílaba como unidad articuladora hacen que los niños alcancen un buen rendimiento en tareas de conciencia silábica desde muy temprano.

En cambio, tanto la detección de rimas como del primer sonido de las palabras conllevan un mayor nivel de conciencia segmental, ya que en ambos casos se trata de un nivel subsilábico. En español nos encontramos con que la conciencia de las unidades intrasilábicas aparece antes que la de los fonemas, ya que implica detectar partes de la sílaba (Herrera y Defior, 2005; Jiménez, 1992).

Otro resultado a destacar es que desde muy temprano se manifiesta una asociación entre el conocimiento lector y el conocimiento fonológico, como lo muestran las correlaciones obtenidas en distintos estudios. La correlación suele ser más fuerte entre el conocimiento prelector y las tareas de clasificación de palabras por su sonido inicial (Herrera y Defior, 2005).

También es importante señalar el valor de las distintas pruebas de procesamiento fonológico para diferenciar entre normolectores y niños con dificultades de aprendizaje en la lectura. Entre las pruebas de conciencia fonológica, la de conciencia silábica no resulta muy útil para este objetivo puesto que tanto unos como otros obtienen puntuaciones muy altas y, por lo tanto, no discrimina entre ambos grupos. Sin embargo, las pruebas de clasificación de palabras por su sonido inicial tienen un mayor valor de discriminación (Herrera y Defior, 2005).

Por lo tanto, el EGRA incluye tareas basadas en la identificación del primer segmento fonológico y en la clasificación de palabras por su sonido inicial, ya que estas tareas mantienen una mayor relación con la adquisición de la lectura en español.

Por último, el desarrollo de la conciencia fonológica se ve afectado por las propias características del lenguaje (Defior, 2002; Jiménez y Haro, 1995). El acceso a las unidades fonológicas depende de algunas variables lingüísticas. Las palabras monosilábicas presentan un mayor nivel de dificultad que las bisilábicas y trisilábicas, de modo que se produce una relación inversa, y cuanto menor es el número de sílabas, tanto peor es la ejecución. Por otra parte, las palabras bisilábicas y trisilábicas, que son las más frecuentes en español, son más fáciles de segmentar si la estructura silábica inicial es de CV, que si está conformada por una única vocal. Ambos hechos se podrían explicar en términos de la influencia que las características del sistema lingüístico español tienen sobre el propio desarrollo de las habilidades fonológicas. El español se caracteriza por el escaso número de palabras monosílabas, especialmente de palabras monosílabas de contenido; las más abundantes son las palabras bisílabas y trisílabas. Por otro lado, la estructura silábica más frecuente es CV, y es por lo tanto plausible que la mayor familiaridad de los niños con estas regularidades de la lengua influya en su capacidad de detección y manipulación de estos elementos lingüísticos. Por último, Jiménez y Haro (1995) diseñaron un experimento para analizar si el acceso a las unidades fonológicas se halla mediatizado por variables lingüísticas (v.g., el tipo de estructura silábica, la longitud de la palabra, la posición de la sílaba tónica y las propiedades

fonoarticulatorias). Estos autores encontraron que cuando se manipula la estructura silábica, los niños prelectores tienen dificultades para manipular fonemas que aparecen en grupos consonánticos (i.e., sílabas con una estructura consonante-consonante-vocal [CCV]), pero más facilidad para identificar el primer fonema de las palabras, en especial cuando se trata de fonemas fricativos.

Fluidez

Otro componente importante de la lectura lo constituye la fluidez. Aunque no existe una sola definición, sin embargo, cuando hablamos de lectura fluida nos estamos refiriendo a la habilidad de leer rápido, de modo preciso y con expresividad (National Reading Panel, NICHD, 2000).

La automatización del reconocimiento de las palabras es un componente importante de la lectura hábil, que se traduce en velocidad, autonomía y liberación de recursos cognitivos (Schwanenflugel, Hamilton, Kuhn, Wisenbaker y Stahl, 2004), y que contribuye de modo indirecto a la comprensión.

En este sentido se ha demostrado que la fluidez lectora es uno de los principales elementos que predicen el rendimiento en lectura y comprensión (Fuchs et al., 2001). La fluidez debe ser considerada tanto en términos del reconocimiento automático de las palabras como de la lectura prosódica o expresiva (Kuhn y Stahl, 2003). Se pueden distinguir dos niveles en los conocimientos prosódicos, en relación con la lectura: (a) a nivel léxico, que se traduce en el patrón de acentuación de las palabras; esto enlaza con una de las características del código español, que establece el uso de una marca escrita (la tilde) en determinadas condiciones; y (b) a nivel supraléxico, que se traduce en el ritmo, la entonación y las pausas en las oraciones y sus componentes. Esto se enlaza con las marcas utilizadas para traducir la prosodia al texto, mediante el uso de los signos ortográficos de puntuación y de las mayúsculas al inicio de la frase.

La lectura fluida aparece en la mayoría de los niños—aunque no siempre—cuando automatizan las habilidades del reconocimiento de palabras a través de la práctica, algo que sabemos que ocurre entre 1º y 3º grado (Defior, Jiménez-Fernández y Serrano, 2007).

Competencia lingüística (vocabulario y comprensión oral)

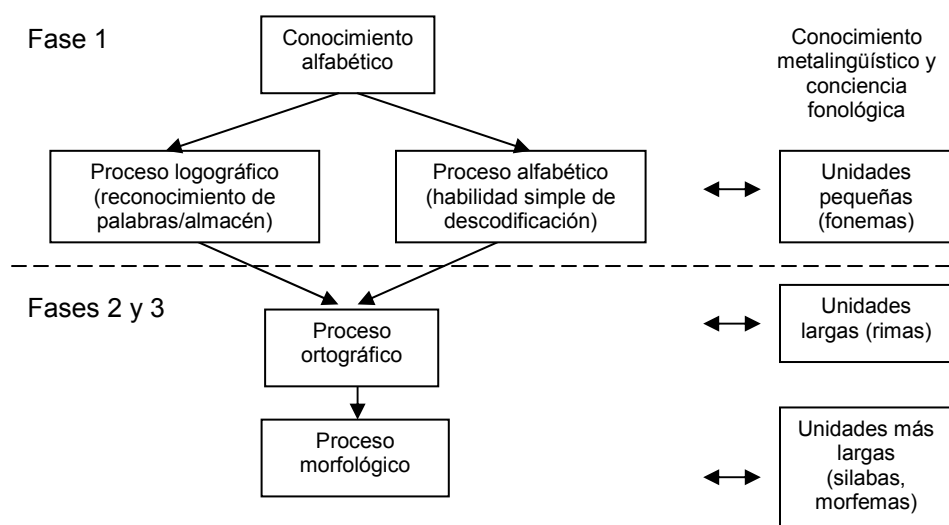
La competencia lingüística (oral language proficiency) es una de las habilidades críticas que los niños también deben desarrollar cuando aprenden a leer. Ella se refiere a las habilidades de comprensión y expresión oral, así como al conocimiento y dominio de la lengua oral, incluyendo el componente fonológico, vocabulario, morfología, gramática y habilidades pragmáticas. El conocimiento del vocabulario es, por ejemplo, una habilidad crítica que afecta a los procesos de comprensión lectora y en particular a los procesos superiores del lenguaje, como serían el procesamiento gramatical, la construcción de esquemas y las estructuras textuales (Adams y Collins, 1979; Chall, 1987). En definitiva, los normolectores pueden tolerar una pequeña proporción de palabras desconocidas cuando leen un texto, lo que no afecta a su comprensión ya que generalmente extraen el significado de estas palabras a partir del contexto. Sin embargo, sabemos que si la proporción de palabras desconocidas es muy alta, ello dificultará la comprensión de lo que leemos (Carver, 1994). En general, la competencia lingüística es una capacidad que ha sido evaluada en muchos estudios a partir de los propios informes o valoraciones de los profesores, pero es también recomendable su evaluación a través de pruebas de vocabulario o de comprensión oral.

Como ya hemos señalado, el diseño del EGRA, así como el énfasis dado a las habilidades de lectura temprana y en particular a la conciencia fonológica y la ruta fonológica, significa que no es conveniente evaluar estos componentes en niños de grados superiores (esta recomendación asume que dichos educandos ya han adquirido las habilidades de lectura temprana; por lo tanto, las decisiones referidas al uso del EGRA deberían basarse en la habilidad más que en la edad, y la atención debería centrarse en el conocimiento de la tasa de progresión en el aprendizaje del alumno, de ser posible). Una vez que la conciencia fonológica y la ruta fonológica quedan incorporadas al repertorio de habilidades lingüísticas del estudiante, no resulta ya apropiado concentrarse en dichas habilidades a través de la evaluación o la instrucción. En otras palabras, la enseñanza de lo fonológico está limitada temporalmente, en tanto que la enseñanza del vocabulario y la comprensión son prácticas que se deben prolongar en el tiempo, debiendo ambas ser evaluadas y enseñadas. El balance de la enseñanza también cambiará una vez que los niños pasen de “aprender a leer” a “leer para aprender” (Espin y Tindal, 1998; International Reading Association, 2007).

El desarrollo de la lectura como una sucesión de etapas

Aprender a leer es un proceso que se da a través de distintas etapas y que conlleva más tiempo en algunas lenguas, en comparación con otras. La gráfica 4 muestra una representación

Gráfica 4. El modelo dual del desarrollo ortográfico



Fuente: Adaptado de Seymour, P. H. K., Aro, M., y Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174.⁹

⁹ Es importante señalar que la adquisición puede ser simultánea. Hay evidencias de que el desarrollo de la conciencia fonológica en inglés comienza con unidades fonológicas más largas, y que gradualmente se pasa hacia unidades más pequeñas. Prescindiendo del orden o secuencia, el modelo subraya la necesidad de adquirir todas las competencias.

de las distintas etapas en el aprendizaje de la lectura. Según Seymour et al. (2003),

Aprender a leer es un proceso que tiene lugar en distintas etapas, de tal modo que la base fundamental tiene lugar en la primera etapa, desarrollándose luego los procesos morfológicos y ortográficos en las fases segunda y tercera. La primera etapa incluye dos procesos fundamentales, un proceso logográfico que se encarga de la identificación y el almacenamiento de palabras familiares, y un proceso alfabético que se apoya en la descodificación. (p. 144)

Para la etapa logográfica los niños ya han adquirido cierto nivel en el desarrollo del lenguaje oral, observan y participan en muchas actividades tales como escuchar cuentos, y comienzan a familiarizarse con las letras. En este contexto los niños adquieren conceptos funcionales acerca de la lectura y la escritura, aprenden las formas y nombres de las letras del alfabeto, y cómo ellas difieren de los números. También aprenden a identificar logotipos (v.g., Coca-Cola, McDonald's,...) y practican la escritura copiando lo que ven escrito, o inventando su propia escritura para etiquetar dibujos o mensajes creados por ellos mismos (Mason y Allen, 1986).

Ahora bien, ¿cómo cambian los prelectores de la etapa logográfica a la alfabética? En otras palabras: ¿qué tipo de conocimientos o habilidades permiten a los niños comenzar a leer procesando claves gráficas sin ninguna ayuda de dibujos u otras claves contextuales? A juicio de Ehri (1991) caben dos posibles explicaciones:

- Una supone que esta habilidad surge espontáneamente y de forma natural a partir de las experiencias en la prelectura, del mismo modo que se desarrolla el lenguaje oral. Así se adquieren asociaciones significado-lenguaje impreso, lo que se convierte en el fundamento con el cual se aprenden las funciones del lenguaje escrito. Además, gracias a la exposición repetida a etiquetas y señales impresas (v.g., McDonald's, Coca-Cola), lo escrito en sí mismo llega a ser descontextualizado gradualmente y puede ser reconocido a partir de claves gráficas solamente. Sin embargo, Masonheimer, Drum y Ehri (1984) llevaron a cabo un estudio para comprobar si la experiencia con lo impreso capacitaba a los niños prelectores para procesar claves gráficas. Los resultados demostraron que no bastaba la experiencia con lo impreso.
- La otra explicación supone que para comenzar a procesar las claves gráficas que permiten la lectura de palabras, los niños deben adquirir ciertos prerrequisitos tales como el conocimiento de las letras y una conciencia fonémica. Los niños prelectores necesitan aprender el alfabeto para que se produzca en ellos un cambio de atención, de las claves ambientales a la concentración en lo impreso por sí mismo. Poder asimilar que lo impreso traduce el habla, supone ser consciente de los sonidos que forman las palabras (esto es la conciencia fonológica). Esta última es concebida como una habilidad metalingüística que se refiere a la habilidad para reflexionar conscientemente sobre los segmentos fonológicos del lenguaje oral (Sinclair, Jarvella y Levelt, 1978). Así pues, se debe instruir a los niños en este tipo de prerrequisitos para que puedan comenzar a leer las palabras y el texto de forma autónoma. Esto quiere decir que la habilidad lectora no se alcanza simplemente a través de una mera exposición al lenguaje impreso. Lo que los niños necesitan es dominar el alfabeto, esto es conocer las formas de las letras y los nombres o sonidos, lo que constituye el fundamento a partir del cual procesar las claves gráficas en las palabras impresas. Conocer las formas de las letras ayuda a los niños a distinguir y recordar los componentes visuales de las palabras, y conocer los nombres de las letras les ayuda a asociar sonidos relevantes

con ellas, dado que muchos de los nombres contienen esos sonidos (v.g., t, d). Ahora bien, conviene puntualizar que el nombre de las letras sólo es un punto de partida ya que no todos estos nombres contienen sonidos relevantes (v.g., la h), y hay sonidos que no se encuentran en los nombres (v.g., m, n).

Por último, la etapa ortográfica se caracteriza por el uso de estrategias basadas en la comparación de la forma ortográfica de la palabra escrita con las representaciones de palabras almacenadas en el léxico ortográfico. Los diferentes modelos de desarrollo de la lectura establecieron un vínculo entre las etapas evolutivas y las teorías del procesamiento léxico. Entre estas teorías encontramos diferencias en la manera en que se explican los mecanismos psicológicos que intervienen en la identificación de la palabra escrita, una vez aprendidas las reglas de CGF. Mientras que unos modelos postulan la intervención de un solo mecanismo de lectura basado en la analogía, otros sugieren la existencia de diferentes vías con que poder acceder al léxico interno.

Causas que explican el desarrollo de la lectura en lengua española

Las teorías o modelos propuestos para explicar el desarrollo de la lectura como una sucesión de etapas, suelen incluir la conciencia fonológica y la conciencia general del lenguaje escrito como factores que influyen en el aprendizaje de la lectura.

Jiménez y Ortiz (2000) llevaron a cabo un estudio para contrastar empíricamente el modelo propuesto por Lundberg y Høien (1991) en una ortografía transparente como el español, que pretende explicar la dirección de la relación de la conciencia fonológica y la conciencia general del lenguaje escrito con diversos indicadores de la habilidad lectora: lectura de palabras, lectura de pseudopalabras y comprensión lectora. El modelo propuesto incluye siete componentes: conciencia general del lenguaje escrito, conciencia silábica, conciencia intrasilábica, conciencia fonémica, descodificación de palabras, descodificación de pseudopalabras y comprensión lectora.

Uno de los supuestos básicos del modelo es que los distintos niveles de conciencia fonológica (conciencia silábica, conciencia intrasilábica y conciencia fonémica) están más relacionados con la descodificación, mientras que el componente de conciencia general del lenguaje escrito se relaciona directamente con la comprensión lectora. Esta idea de tratar la descodificación y la comprensión lectora por separado, así como su relación con distintas habilidades metalingüísticas, ha sido sugerida mas no comprobada empíricamente.

Con el objeto de probar el modelo teórico propuesto, se realizó un análisis de vías mediante el programa estadístico EQS, V3.0 ("a structural equations program"). En dicho estudio participaron inicialmente 136 prelectores (70 niños y 66 niñas), que tenían entonces una edad promedio de 5.6 años y que al finalizar el trabajo, dos años más tarde, tenían 7.6 años. Se realizó, pues, un estudio longitudinal con tres momentos de medición: al inicio del último curso de educación infantil, al finalizar éste y al inicio del segundo curso de educación primaria. Los resultados demostraron que la direccionalidad de la relación entre la conciencia fonológica y la lectura dependía del nivel de conciencia fonológica estudiado, así como del momento en que se medía cada uno de los niveles de esta habilidad: antes o después de recibir la instrucción lectora.

En lo que respecta a los niveles de conciencia fonológica, la conciencia silábica de los sujetos que aún no habían recibido una enseñanza lectora tenía una relación causal con la descodificación de palabras y pseudopalabras. Durante el transcurso de dicha enseñanza se

producía un aumento en el desarrollo de la conciencia silábica. Después de transcurrido un año de instrucción lectora, la conciencia silábica mantenía una relación indirecta con la lectura de pseudopalabras a través de la conciencia fonémica. La conciencia intrasilábica de los sujetos que habían recibido la instrucción lectora no mantenía una relación causal con la descodificación de palabras y pseudopalabras. Por último, la conciencia fonémica de los sujetos que habían recibido la enseñanza lectora tenía una relación causal directa con la descodificación de palabras y pseudopalabras. Otras investigaciones realizadas también en lengua española, que intentaron relacionar los diferentes niveles de conciencia fonológica con la adquisición de la lectura, encontraron resultados similares (Carrillo, 1994; González, 1996). Esto significa que la conciencia fonémica es mucho más relevante que la conciencia silábica e intrasilábica, y efectivamente se ha apuntado que una vez que los niños españoles conocen el código alfabético y tienen una conciencia fonémica, ya no es necesaria la categorización de palabras en sus componentes intrasilábicos para ser capaces de leer (Jiménez y Muñeton Ayala, 2002; Serrano, Defior y Jiménez, 2005). Asimismo se demostró en nuestra lengua que la conciencia fonológica es susceptible de entrenamiento en el marco de la instrucción escolar, incluso antes de iniciar el aprendizaje formal de la lectura, y que dicho entrenamiento tiene efectos positivos sobre el desarrollo de la lectura (Domínguez, 1996a, 1996b).

En cuanto a la conciencia general del lenguaje escrito, los resultados demostraron que “la conciencia de las características específicas del lenguaje escrito” tenía una relación bidireccional con la comprensión lectora. Esto es, en los sujetos prelectores, esta habilidad metalingüística era un factor causal de la comprensión lectora y, con la instrucción lectora, se producía un desarrollo considerable de dicha habilidad metalingüística. Además, una vez transcurrido el primer año de enseñanza lectora, esta habilidad metalingüística ejercía una gran influencia sobre la comprensión lectora. También contamos con algunas evidencias empíricas en español, de que un programa de entrenamiento en este tipo de habilidades, que ofrezca a los alumnos la oportunidad de poder interactuar con materiales impresos, contribuye a una mayor estabilidad de la conciencia del lenguaje escrito en el tiempo (Ortiz, 1994).

La transparencia de la ortografía y el aprendizaje de la lectura

El aprendizaje de la lectura exige el desarrollo de habilidades muy específicas (v.g., el desarrollo de la conciencia fonológica, el dominio de reglas de conversión grafema-fonema, etc.) que dependen en gran medida del tipo de sistema de escritura que corresponde aprender. No son las mismas en un sistema logográfico (v.g., chino) que otro alfabético (v.g., inglés, francés, español) o silábico (v.g., japonés). En este punto es importante señalar que en lo que a los sistemas alfabéticos se refiere, “la escritura alfabética representa a la lengua en el nivel fonémico o fonológico y no fonético, puesto que los grafemas reflejan los fonemas, como sonidos abstractos, y no a los fonos o producciones concretas de aquellos” (Clemente y Domínguez, 1999, p. 29). La escritura es también una representación ortográfica, ya que las palabras contienen información que sobrepasa el principio fonémico. Por ejemplo, no siempre existe una correspondencia biunívoca entre grafemas y fonemas. La habilidad lectora, en particular el reconocimiento visual de las palabras, implica un dominio de la descodificación que involucra procesos fonológicos (es decir, se refiere a los sonidos que están incluidos en las palabras) y ortográficos (esto es, se refiere al procesamiento de las letras y de memoria). Para ello se requiere una enseñanza formal que no se asemeja en absoluto a la que tiene lugar cuando se aprende el lenguaje oral.

Los sistemas alfabéticos varían, por lo tanto, en la consistencia de sus relaciones con el lenguaje oral a lo largo de un continuo de opacidad y transparencia. En un extremo se sitúan los sistemas de escritura denominados transparentes (v.g., italiano, finlandés, español), en

donde existe una correspondencia más estrecha entre la forma fonológica y gráfica, y en el otro extremo los sistemas denominados opacos (v.g., inglés, francés), en los cuales hay muchas palabras que no se ajustan a las reglas de correspondencia grafema-fonema (CGF). Estos sistemas presentan irregularidades en las correspondencias fonológicas en mayor o menor medida, debido en parte a que también representan unidades morfológicas.

Un ejemplo concreto de las diferencias lingüísticas sería la regularidad en francés, español e inglés. Uno de los principales problemas que la ortografía de la lengua francesa presenta no son precisamente las vocales (como sí sucede en inglés) sino la ortografía del final de palabra ya que ella a menudo es silenciosa. Además, la consistencia de las correspondencias grafema-fonema en francés y español (i.e., reglas de CGF usadas para leer) es mayor a las que existen en las reglas de correspondencia fonema-grafema en la escritura. La tabla 1 muestra las correspondencias grafema-fonema y fonema-grafema para la lectura y escritura de vocales en inglés y francés.

En inglés, “seguir las reglas” de la lectura y escritura de palabras monosilábicas da como resultado una lectura correcta del 48 por ciento en la primera y 67 por ciento en la segunda. En la lectura en francés, en cambio, hay un alto nivel de regularidad en las correspondencias grafema-fonema, pero en la escritura se producen tantos cambios como ocurre con la escritura en inglés (ver para una revisión Sprenger-Charolles, Colé y Serniclaes, 2006), aunque esto es menor en el caso del español.

Tabla 1. Consistencia de las correspondencias grafema-fonema (CGF) y fonema-grafema (CFG) en las vocales, en inglés y francés (palabras monosílabas)

Lengua	CGF (lectura)	CFG (escritura)
Inglés	48%	67%
Francés	94%	68%

Fuente: Peereman, R., y Content, A. (1999). LEXOP: A lexical database providing orthography-phonology statistics for French monosyllabic words. *Behavioral Methods, Instruments and Computers*, 31, 376–379.

En el caso de esta última lengua (i.e., el español), el alfabeto cuenta con aproximadamente 29 grafemas que se pueden representar mediante 27 caracteres o letras diferentes. Algunos fonemas se pueden representar con más de un grafema y algunos grafemas pueden representar a más de un fonema (i.e., grafemas *inconsistentes*). En este sentido se ha sugerido que el español es más transparente en la dirección grafema-fonema que en la dirección fonema-grafema. Por ejemplo, el inglés tiene más de una decena de fonemas vocálicos y sólo utiliza las letras “a, e, i, o, u”, que han de cumplir funciones múltiples solas o combinadas, lo que da como resultado numerosas inconsistencias. En cambio el español posee sólo cinco fonemas vocálicos, lo que le permite establecer una relación biunívoca entre fonemas y letras. Esto no ocurre con las consonantes. Hay un total de ocho fonemas que se pueden representar con más de un grafema: el sonido /x/ se puede representar con j, g; el sonido /k/ con c, k, q; el sonido /g/ con g, gu; el sonido /b/ con b, v, w; el sonido /r/ con r, rr; el sonido /θ / con z, d final; el sonido /N/ con ll, y; y el sonido /i/ como i e y; luego estaría el caso de algunas letras como la h, a la que no corresponde ningún fonema (/Ø/). El hecho de que existan fonemas que tienen más de una representación gráfica hace que para muchos alumnos, la escritura ortográficamente correcta de algunas palabras resulte una tarea difícil de realizar. Por ello en español tenemos palabras en las cuales la correspondencia fonema-grafema es unívoca (i.e., al fonema siempre le corresponde el mismo grafema), y palabras que tienen algún fonema que

se puede representar con más de un grafema. En este último caso hablaríamos de palabras de ortografía arbitraria, donde la elección del grafema para el fonema inconsistente podría estar dependiendo del conocimiento y aplicación de una regla ortográfica (i.e., palabras de ortografía arbitraria reglada) o no, dependiendo de dicha circunstancia (i.e., palabras de ortografía arbitraria no reglada). Para la escritura correcta de las palabras de ortografía arbitraria el alumno, además de conocer la correspondencia fonema-grafema, tiene también que tener una representación interna de la palabra a nivel ortográfico. Por consiguiente, cuando leemos en la lengua española, que tiene una ortografía transparente, el mismo fonema le suele corresponder siempre a cada grafema. Esto no ocurre en una lengua como la inglesa, que tiene una ortografía opaca, ya que las relaciones existentes entre grafemas y fonemas no llegan a ser consistentes; hay así una mayor irregularidad entre la grafía y el sonido.

Los sistemas ortográficos transparentes son más fáciles de aprender que los opacos, es decir que la consistencia del código es un factor importante a tener en cuenta en los modelos explicativos de la adquisición lectora (Wimmer y Hummer, 1990). La inmensa mayoría de los datos con que contamos sobre la lectura y su aprendizaje provienen de estudios realizados en lengua inglesa (Share, 2008), y el nivel de opacidad de una lengua tiene una incidencia sobre el aprendizaje de la lectura (Ziegler y Goswami, 2005, 2006). Los resultados obtenidos en muchos estudios españoles muestran un control automático del procedimiento fonológico a temprana edad, con tiempos de reacción mucho más bajos con respecto a otros sistemas lingüísticos. Entre los factores lingüísticos que se pueden considerar como posibles causas de las diferencias entre las distintas lenguas están el número de vocales (mucho menor en español que en las otras lenguas), la transparencia del código (la del español es mucho mayor) y el fenómeno de reducción vocálica (muy alto en portugués, algo menor en francés y bajo en español). Las implicaciones que estos hallazgos tienen para la enseñanza de la lectura son obvias. Cuando se tiene una lengua que es bastante predecible al nivel de las reglas de CGF, el factor lingüístico que parece decisivo en la velocidad de adquisición es el número de vocales. Dada la superior inconsistencia de otras lenguas (v.g., inglés, francés) al nivel de estas reglas, en comparación con el español, la escritura resulta particularmente afectada por el grado de transparencia ortográfica.

Estudios comparativos

La investigación que comparó el aprendizaje de la lectura en trece países europeos diferentes ofrece evidencias empíricas que apoyan la hipótesis antes expuesta, esto es que la regularidad de una lengua influye en la velocidad de adquisición de la habilidad lectora (Seymour et al., 2003). El nivel de complejidad de una lengua afecta el aprendizaje de la lectura desde el principio. En este caso, la adquisición inicial de la lectura es incluso más lenta en lenguas con estructuras lingüísticas más complejas. Es decir, se da el caso de que el nivel de complejidad de una lengua afecta al aprendizaje incluso en las habilidades más básicas, como por ejemplo la identificación de letras (para una revisión de las diferencias entre distintas lenguas, véase Sprenger-Charolles et al., 2006; Ziegler y Goswami, 2005). Wimmer y Goswami (1994) también demostraron que la lectura de los nombres de los números en inglés (i.e., los niños tenían que leer *three, four, six*, etc.) es lenta y plantea problemas incluso a escolares de tercer grado. En cambio, en alemán esta habilidad ha quedado bien establecida para finales del primer año de escolaridad. Estos resultados no se debieron a problemas de pronunciación, puesto que los niños tuvieron que leer también las cifras correspondientes (i.e., 3, 4, 6, etc.). En lo que a estas últimas condiciones respecta, no se hallaron diferencias entre anglófonos y germanófonos. El trabajo sobre la aplicación del EGRA en Gambia y Senegal también parece apoyar esta conclusión, a saber, que el rendimiento de niños que aprenden a leer en inglés es mucho más

bajo que el de los que aprenden a leer en francés, que presenta una mayor regularidad en la dirección grafema-fonema.

En cuanto a los niños que aprenden a leer en lenguas con ortografía transparente o regular tales como español, italiano, finlandés y griego, solamente requieren de aproximadamente un año de enseñanza para alcanzar cierta automatización (Seymour et al., 2003). Así, por ejemplo, el rendimiento en la lectura de palabras y pseudopalabras por parte de niños que aprenden a leer en distintos contextos idiomáticos (v.g., inglés, francés y portugués) es sistemáticamente más alto en español que en otras lenguas (Genard, Alegría, Leybaert, Mousty y Defior, 2005). El conocimiento de las complejas reglas de conversión grafema-fonema (CGF) es más temprano en español que en inglés, francés y portugués. Los niños hispanohablantes asimismo alcanzan el techo en la lectura de palabras y pseudopalabras muy temprano en comparación con otras lenguas,¹⁰ lo que indicaría que la utilización adecuada del procedimiento fonológico ocurre más pronto en español que en inglés, francés y portugués. Por lo tanto, estos resultados vienen a reforzar la idea de que la consistencia de las relaciones grafémico-fonémicas tiene un impacto positivo sobre el aprendizaje de la lectura.

Ahora bien, las diferencias debidas a la consistencia de la ortografía no son únicamente cuantitativas. Por ejemplo, la existencia de una etapa logográfica ha sido puesta en tela de juicio en el caso de niños que aprenden a leer en una lengua que tiene una ortografía transparente. En general estos niños, al ser estimulados a leer fonológicamente por las propias exigencias de la ortografía, son más eficientes en el procesamiento subléxico y no suelen usar la estrategia logográfica de lectura. Los niños anglófonos utilizan mucho más sus conocimientos lexicales para leer, probablemente debido a las dificultades de asociación entre grafemas y fonemas en esta lengua. Esto es lo que sugieren algunos de los resultados hallados en numerosos estudios (Bruck, Genesee y Caravolas, 1997; Goswami, Gombert y de Barreira, 1998; Seymour et al., 2003), que muestran que las dificultades de los lectores anglófonos aprendices son menos marcadas en la lectura de palabras que en la de pseudopalabras.

La opacidad de la ortografía también incide sobre las unidades ortográficas utilizadas por el lector aprendiz. En la lengua inglesa existen patrones ortográficos en muchas palabras que corresponden a unidades lingüísticas mayores que el fonema, como sería el caso de la rima (v.g., *right*, *sight*, *flight*, etc.); por lo tanto, el enfoque basado en analogías sería más apropiado para los niños que aprenden a leer en inglés. En cambio, cuando la ortografía es transparente, se utilizará una sola estrategia apoyándose en las correspondencias grafema-fonema. Los estudios en español han demostrado, por ejemplo, que los niños no se apoyan en este tipo de unidad lingüística (i.e., principio-rima) en el reconocimiento visual de palabras (Jiménez, Álvarez, Estévez y Hernández-Valle, 2000). Este mismo resultado se ha observado incluso en estudios comparativos, por ejemplo, en un estudio en el que se comparó el rendimiento basado en la utilización de la “rima” en niños anglófonos, francófonos, e hispanohablantes de 7, 8 y 9 años (Goswami et al., 1998). Estos niños debían leer pseudopalabras que rimaban [+] o no [-] con palabras (v.g., en francés: *voile* (vela) vs. *roile* [+] o *loave* [-]; en español: *mientes* vs. *lientes* [+] o *teslien* [-], en inglés: *tape* vs. *fape* [+] o *faish* [-]). En los sujetos más jóvenes (7 años) se encontró que la presencia de rimas facilitaba la lectura de los niños anglófonos y en menor medida la de los francófonos (una mejora del orden del 15% a 20%, contra 5%), pero no la de los hispanohablantes. La misma tendencia fue también observada en una comparación entre niños anglófonos y germanófonos (Goswami, Ziegler, Dalton y Schneider, 2003).

¹⁰ Véase la Sección VIII al inferior para comparaciones entre lenguas de fluidez esperada.

En definitiva, estos estudios comparativos entre sistemas ortográficos diferentes indican que la adquisición de la lectura es más rápida en un sistema transparente.

Por otra parte, dada la asimetría de las correspondencias grafema-fonema y fonema-grafema, siendo las primeras más consistentes que las segundas en inglés y sobre todo en francés y español, podemos esperar observar también una inferioridad de la escritura sobre la lectura. Es lo que sugieren los resultados encontrados una vez más. Así, mientras que las palabras regulares son leídas mejor y más rápidamente que las pseudopalabras, particularmente en francés (Sprenger-Charolles, Siegel, Bonnet, 1998), no son escritas mejor (en francés: Alegría y Mousty, 1996; Leybaert, y Content, 1995; en inglés: Bruck y Waters, 1990; Foorman, Jenkins y Francis, 1993; Stage y Wagner, 1992; en español: Jiménez, O'Shanahan, Pérez et al., 2008; Jiménez, Pérez, O'Shanahan, Muñetón y Rojas, 2009). También se ha sugerido con respecto a este punto, que el léxico ortográfico se desarrolla con mayor rapidez en español que en francés, y que la ventaja de los hispanohablantes, en comparación con los francófonos, se debe probablemente a que los primeros aprenden a leer antes que los segundos, y los niños que leen antes pueden leer más, en tanto que el mecanismo de autoaprendizaje desarrolla la lectura y la ortografía lexical con mayor rapidez (Alegría, Carrillo, y Sánchez, 2005).

IV. Taller de adaptación e investigación de EGRA

Una vez que un ministerio o alguna otra organización dentro de un país decide iniciar el proceso de adaptar y aplicar el instrumento EGRA, el primer paso es organizar un taller en el país que usualmente dura unos cinco días, y que tiene los siguientes objetivos:

- dar una base, tanto a los funcionarios del ministerio como a los especialistas locales en el currículo y la evaluación, en las investigaciones que sustentan los componentes del instrumento.
- revisar los procedimientos del consentimiento informado y discutir la ética de la investigación de, y el trabajo con, sujetos humanos, niños en particular.
- revisar los componentes del instrumento y su vínculo con el currículo.
- adaptar el instrumento a las condiciones locales usando las pautas de construcción de ítems proporcionadas en este manual (lo que incluye la traducción de sus instrucciones; el diseño de una versión en la lengua local de ser necesario; y la modificación de los componentes de palabras y párrafos de lectura, para que reflejen las palabras y conceptos local y culturalmente apropiados).
- capacitar supervisores de campo en la supervisión del proceso de aplicación del instrumento, lo que incluye la evaluación de la confiabilidad entre evaluadores y la preevaluación del nuevo instrumento adaptado.

La sección examina los pasos en la preparación y dictado de un taller EGRA, y presenta un examen general de los temas a cubrir. Después del taller se recomienda realizar una segunda semana de formación para los examinadores y llevar a cabo un piloto del instrumento (en equipos de cuatro o cinco por centro educativo). La formación de los examinadores y el trabajo de campo se examinan en la Sección V.

Taller de adaptación e investigación

El número de participantes en el taller de adaptación e investigación quedará determinado por el de los centros educativos a muestrear (véase el Anexo B) y la disponibilidad del personal del ministerio para que participe tanto en el diseño como en el trabajo de campo concreto en dichos centros. Se recomienda usar al personal del ministerio para así construir capacidades y ayudar a asegurar la sostenibilidad de estas evaluaciones. De ser posible, los participantes pueden incluir también a profesionales que no forman parte del ministerio, académicos, docentes y expertos en el diseño de currículo, evaluación, supervisión y apoyo escolar, y en la lengua local. La situación ideal es que el personal del ministerio participe en todo el proceso de adaptación y del piloto (más de 1 mes en total, dependiendo del número de centros educativos a muestrear). En los países en donde no pueda contarse con personal del ministerio durante todo este periodo de tiempo, debe incorporarse al grupo supervisores y examinadores de campo contratados.

Los supervisores de campo deben participar en todo el proceso de adaptación y de justificación de la investigación, para que así entiendan mejor los principios subyacentes de cada uno de los componentes de evaluación. Si todos los participantes no estuvieran disponibles durante todo el proceso de adaptación, entonces examinadores adicionales pueden asistir a una semana de formación posterior a este taller, que incluye una aplicación piloto en varios centros educativos

para prácticas y modificaciones finales al instrumento. Un número de examinadores mayor al necesario debería participar en la formación de seguimiento, para así permitir escoger a los que muestren ser más hábiles en la aplicación del instrumento (establecido mediante cálculos de confiabilidad entre evaluadores; véase el Anexo C).

Los grupos conformados por personal del ministerio, formadores de docentes, maestros retirados y otros en formación brindan una buena combinación de experiencia, entusiasmo y energía, todos elementos importantes en el proceso de evaluación.

Si bien el personal del ministerio en particular debería escogerse en base a su capacidad para contribuir al diseño y adaptación de los instrumentos de EGRA, es posible que no todos ellos sean seleccionados como examinadores o supervisores. Los dos requisitos más importantes de los examinadores son:

- **La capacidad de interactuar con niños pequeños de modo no amenazante.** Dado que el instrumento se aplica oral e individualmente, la calidad y exactitud de los datos depende mayormente de la capacidad del examinador para animar y calmar a los educandos, de tal modo que se desempeñen lo mejor posible. Los resultados negativos falsos de educandos que tienen un buen desempeño pero que están nerviosos o tienen miedo, pueden minimizarse con el enfoque apropiado. Si bien uno puede practicar y entrenarse algo en esto, gran parte viene a ser actitud y personalidad, dos cosas difíciles de modificar con un ejercicio de formación de 1–2 semanas.
- **Habilidades organizativas.** El segundo requisito importante es la habilidad de manejar varias tareas al mismo tiempo, entre ellas escuchar al educando, puntuar los resultados y manejar un cronómetro o temporizador.

Como parte del proceso de selección, los jefes de taller deben efectuar pruebas de confiabilidad entre evaluadores usando grabaciones de audio de educandos participando en evaluaciones piloto. Los examinadores escuchan y puntúan la evaluación, y luego informan de sus resultados. Aquellos examinadores cuyos resultados caen a más de una desviación estándar de la media deberían recibir oportunidades adicionales de formación y práctica. Si no mejoran no deberían ser seleccionados para que participen en el proceso de recolección de datos. Las pruebas adicionales de confiabilidad entre evaluadores se examinan en la Sección V.

El taller debería ser conducido por un equipo de al menos dos expertos. El primero de ellos—responsable en liderar la adaptación del instrumento, presentar el proceso de diseño del instrumento EGRA y guiar los procesos de recolección e ingreso de los datos—debería contar con antecedentes en investigación educativa por encuestas y diseño de evaluaciones/pruebas. Esta experiencia debería incluir estadística básica y un conocimiento operativo de Excel y de un programa estadístico como SPSS o Stata. El segundo experto—responsable para la presentación de las investigaciones de la lectura y de procesos pedagógicos de instrucción—debería tener antecedentes en herramientas de evaluación de lecturas e instrucción. Ambos jefes de taller deben estar bien documentados en los componentes y las justificaciones de la evaluación, y ser expertos en el trabajo en una diversidad de países y contextos.

De ser posible, el respaldo logístico (la sede, dietas, visitas escolares, impresiones, laminado de los instrumentos escolares, etc.) debería brindarlo una ONG local o una empresa acostumbrada a realizar encuestas en los centros educativos locales.

Materiales que el taller debe incluir:

- Papel y lápices con borrador para los participantes
- Cronómetros o temporizadores (encuentre uno de cocina de ser posible, que cuente de un minuto hacia abajo)¹¹
- Proyector LCD, pizarra y rotafolio (de ser posible, el proyector LCD debe poder proyectar a la pizarra para los ejercicios de puntuación simulados)
- Copias de las presentaciones, manual del supervisor y borrador de los instrumentos
- Presentación del proceso de diseño de EGRA, finalidad, usos y antecedentes de la investigación
- Presentación de investigaciones en instrucción de lectura relacionadas con EGRA.

La tabla 2 muestra un ejemplo de agenda para el taller de adaptación e investigación.

Tabla 2. Ejemplo de agenda: Taller de adaptación e investigación de EGRA

Lunes	Martes	Miércoles	Jueves	Viernes
<ul style="list-style-type: none"> • Presentar facilitadores y asistentes • Revisar la logística • Revisar el proceso de diseño de EGRA • Revisar las investigaciones que subyacen a EGRA 	<ul style="list-style-type: none"> • Revisar el borrador del instrumento y del cuestionario de los educandos • Revisar el borrador del instrumento 	<ul style="list-style-type: none"> • Capacitar en el uso de la aplicación, y practicar 	<ul style="list-style-type: none"> • Aplicar pre-prueba en 2–6 centros educativos • Ingresar datos como ejercicio de práctica • Aplicar prueba de confiabilidad entre evaluadores • Efectuar análisis simples usando Excel 	<ul style="list-style-type: none"> • Discutir resultados • Discutir implicaciones para la instrucción • Revisar el borrador del instrumento a partir de los resultados

Nota sobre el diseño del instrumento y la estrategia de codificación

Hay un tema global referido a los componentes del instrumento y otros específicos vinculados al diseño, que en última instancia tendrá un efecto significativo sobre la codificación y el análisis de los datos. Este tema es que el sistema de codificación de los ítems del instrumento debe permitir a los evaluadores distinguir entre diversos tipos de respuestas de aspecto similar. Podría, por ejemplo, haber situaciones en las cuales (a) una pregunta no se hizo, (b) un educando no sabía la respuesta, (c) el educando no quería o no podía responder, o (d) su respuesta era del todo incorrecta (un cero auténtico). Un error común es no lograr distinguir todas estas posibilidades. Sin embargo, ningún conjunto universalmente válido de valores está estipulado para estas respuestas, de modo tal que decisiones como ésta tendrán que ser adaptadas al instrumento bajo desarrollo. Las versiones actuales del instrumento EGRA reflejan esta estrategia de codificación, incluyendo los códigos de las reglas de detención y de no respuesta.

¹¹ Aunque algo costoso, el mejor cronómetro que hallamos para este fin se consigue en la página web de Dynamic Indicators of Basic Early Literacy (DIBELS): <http://dibels.com/merch.html>. Puede encontrarse otra opción significativamente menos costosa en <http://www.cutleryandmore.com/details.asp?SKU=7499>.

Nota sobre la ética de la investigación y la *Institutional Review Board* (IRB, Junta Revisora Institucional)

Como centro de investigación que recibe subsidios federales, RTI sigue las regulaciones federales de EE.UU en lo que respecta a efectuar investigaciones éticas. Tal como se indica en la descripción que RTI hace del proceso: toda organización que lleve a cabo investigaciones que involucren a sujetos humanos debe utilizar *Institutional Review Boards* (IRB). Las IRB usan el conjunto de principios básicos esbozados en el “Informe Belmont”, un informe presentado en 1978 por la United States National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, como guía de su examen de los protocolos de investigación propuestos. El Informe Belmont esboza tres principios básicos:

- **Respeto por las personas.** Los posibles sujetos de una investigación deben ser tratados como agentes autónomos, que tienen la capacidad de considerar alternativas, hacer elecciones y actuar sin la influencia o interferencia indebida de terceros.
- **Beneficio.** Los dos principios básicos del beneficio son: (1) no hagas daño, y (2) protege del daño maximizando los posibles beneficios y minimizando posibles daños.
- **Justicia.** Este principio ético requiere de equidad en la distribución de las cargas y beneficios de la investigación (página web interna de RTI, consultada el 12 de enero de 2009).

En cada una de las evaluaciones efectuadas hasta la fecha, RTI incluyó un consentimiento verbal de los sujetos humanos que participaban en ellas. Antes de efectuar la evaluación, los examinadores describen los objetivos del estudio e informan a los educandos que ella es anónima, que no afectará su calificación en la escuela y que se la usará para hacer mejoras en cómo los niños de su país aprenden a leer. De efectuarse encuestas de directores de colegios o de docentes como parte del estudio, se completa un proceso escrito de consentimiento similar. Aunque este proceso de consentimiento a menudo no resulta familiar a las contrapartes del país local, el proceso frecuentemente es aplaudido por educandos y docentes, quienes informan sentirse orgullosos al tener la opción de participar en la evaluación. Pocos alumnos y profesores rehusaron participar (en el caso de una muestra reciente de Nicaragua, con más de 6,000 estudiantes, sólo 8 no quisieron hacerlo). En estos casos se elige otro educando al azar. Para información adicional sobre las IRB y la investigación ética con sujetos humanos, niños inclusive, por favor véase (en inglés) <http://www.hhs.gov/ohrp/faq.html>.

Revisión de los componentes del instrumento

Para diseñar el Early Grade Reading Assessment completo, el equipo de diseño de EGRA revisó más de una docena de instrumentos de evaluación, entre ellos el Dynamic Indicators of Basic Early Literacy (DIBELS), el Peabody Picture Vocabulary Test (PPVT) e instrumentos aplicados en España, Perú, Kenia, Mongolia e India. El instrumento EGRA asimismo aprovecha lecciones de esfuerzos en marcha para desarrollar formas “más pequeñas, más rápidas y más baratas” de evaluar el alfabetismo (adulto) (International Literacy Institute y UNESCO, 2002; Wagner, 2003).

Como ya vimos, para tener comentarios al diseño inicial de EGRA, USAID, el Banco Mundial y RTI organizaron una reunión de expertos (puede encontrarse un resumen de las actas y una lista de los participantes del taller en www.eddataglobal.org, bajo News and Events). En base a esta y otras consultas con expertos se diseñó un Early Grade Reading Assessment completo

para su aplicación en inglés. El instrumento resultante contiene ocho tareas o subpruebas, como sigue:

1. Conocimiento del nombre de las letras
2. Conocimiento de los sonidos de las letras
3. Conciencia fonológica
4. Conocimiento de palabras simples
5. Descodificación de palabras sin sentido (pseudopalabras)
6. Lectura y comprensión de un párrafo
7. Comprensión oral
8. Dictado.

Cada uno de estos componentes ha pasado por estudios piloto en árabe, inglés, francés, español y varias otras lenguas más gracias a iniciativas financiadas tanto por el Banco Mundial como por USAID. Los comentarios de profesionales y contrapartes locales incluyeron pedidos de que se redujera el número de habilidades evaluadas con EGRA. Uno de los objetivos del instrumento, como ya se indicó, es evaluar un conjunto de habilidades básicas de lectura, para así poder identificar qué áreas requieren de instrucción adicional. Si EGRA sólo probara la fluidez oral, los resultados de muchos países de bajos ingresos tendrían problemas considerables con el efecto suelo (esto es, la mayoría de los niños de los grados iniciales no podría rendir a un nivel de habilidad suficiente como para permitir el análisis). Por ejemplo, el alumno promedio de tercer grado evaluado en un país africano, se hallaba por debajo del 10º percentil de un estudiante de primer grado en EE.UU. (evaluados ambos al finalizar el año escolar). Esto indica que probar sólo la fluidez de la lectura oral no arrojaría suficientes datos como para informar al personal del ministerio qué está pasando, o no, en lo que respecta a qué habilidades de pre-lectura requieren ser mejoradas (reconocimiento de letras, etc.).

Es asimismo importante señalar que se ha demostrado que el instrumento y los procedimientos aquí presentados constituyen un punto de partida razonable para la evaluación de la lectura en los primeros grados. No debe considerarse sagrado al instrumento en lo que a sus partes componentes respecta. Pero se recomienda que se justifiquen y documente las variaciones, ya sea en las tareas componentes o en los procedimientos. Si bien RTI, el Banco Mundial y USAID entienden que distintos donantes y países adaptarán el instrumento a sus propias necesidades, es importante que dichos cambios se justifiquen y expliquen en función a la finalidad y el uso de la evaluación.¹²

Para una explicación detallada de la calidad y confiabilidad técnica del instrumento EGRA, que incluye pautas con que efectuar revisiones elementales de la calidad del instrumento y de confiabilidad, véase por favor el Anexo C de este manual.

Para resumir el enfoque de evaluación global de EGRA a los asistentes en el taller, se podría compartir la tabla 3 durante el examen de cada uno de los componentes individuales del

¹² RTI y sus patrocinadores asimismo piden que las nuevas adaptaciones de los instrumentos se compartan a través del sitio web de EdData II (www.eddataglobal.org, en el vínculo de EGRA), de modo tal que toda la comunidad educativa pueda aprender de los resultados. Se puede acceder libremente a todos los instrumentos diseñados hasta hoy por RTI en el sitio web; RTI espera que otros tengan interés por compartir sus instrumentos y procesos de aprendizaje.

instrumento. Los participantes deben también entender la diferencia existente entre (1) la evaluación de las habilidades de los educandos y (2) las técnicas de instrucción con que mejorar el rendimiento de los alumnos en dichas habilidades. En otras palabras y como ya se vio, no se debería enseñar los componentes de la prueba a los educandos; deberían, más bien, diseñarse enfoques de instrucción en base a los resultados de EGRA. Puede encontrarse un examen adicional de estos enfoques en la Sección VII.

Tabla 3. Revisión de los componentes del instrumento

Componente	Habilidad temprana de lectura	Habilidad demostrada por la capacidad del educando para:
1. Conocimiento del nombre de las letras	Reconocimiento de las letras	<ul style="list-style-type: none"> • Dar el nombre de las letras en mayúscula y minúscula, en orden aleatorio
2. Conocimiento de los sonidos de las letras	Método fonológico	<ul style="list-style-type: none"> • Dar el sonido de letras en mayúscula y minúscula, distribuidas en orden aleatorio
3. Conciencia fonológica	Conciencia fonémica	<ul style="list-style-type: none"> • Segmentar palabras en fonemas • Identificar los sonidos iniciales en distintas palabras
4. Conocimiento de palabras simples	Lectura de palabras	<ul style="list-style-type: none"> • Leer palabras simples y comunes de una y dos sílabas
5. Descodificación de palabras sin sentido	Principio alfabético	<ul style="list-style-type: none"> • Efectuar correspondencias entre grafemas y fonemas (CGF) mediante la lectura de pseudopalabras simples
6. Lectura y comprensión de un párrafo	Fluidez de la lectura oral	<ul style="list-style-type: none"> • Leer un texto con exactitud, con poco esfuerzo y a un ritmo suficiente
	Comprensión de lectura	<ul style="list-style-type: none"> • Responder correctamente a distintos tipos de preguntas, incluyendo preguntas literales e inferencias sobre el texto que leyeron
7. Comprensión oral	Comprensión oral	<ul style="list-style-type: none"> • Responder correctamente a distintos tipos de preguntas, incluyendo preguntas literales e inferencias sobre el texto que el examinador les leyó
8. Dictado	Principio alfabético	<ul style="list-style-type: none"> • Escribir, deletrear y usar la gramática correctamente, mediante un ejercicio de dictado

1. Conocimiento del nombre de las letras

La prueba del conocimiento del nombre de las letras es la más elemental de las evaluaciones de la preparación (y riesgo) en la lectura del educando. El conocimiento del nombre de las letras es un predictor consistente del desarrollo de la lectura en los hablantes nativos de inglés, francés y otras lenguas alfabéticas (Chiappe, Siegel y Wade-Woolley, 2002). También ha resultado un indicador útil para hablantes no nativos (Chiappe, 2006).

En esta evaluación del conocimiento del nombre de las letras se pide a los alumnos que den el nombre (no el sonido) de todas las letras que puedan, en un periodo de tiempo de un minuto. Todas las letras del alfabeto son listadas en orden aleatorio, 10 letras por hilera, usando una fuente clara, grande y familiar (por ejemplo, Century Gothic de Microsoft Word es sumamente similar a los libros de texto estándares de los niños) y en hileras horizontales, presentándose cada una múltiples veces. Las letras se seleccionan en base a la frecuencia con que se dan en la lengua en cuestión (véase tabla 4). Para impedir que los educandos reciten un alfabeto

memorizado se recurre al azar, esto es, para probar el automatismo real del reconocimiento de las letras y su traducción de una impresión al sonido. Se presenta el alfabeto completo (tanto en mayúscula como en minúscula) en base a evidencias de que las habilidades de lectura en lenguas europeas de los educandos sólo avanzaban después que se conocía alrededor del 80 por ciento del alfabeto (Seymour et al., 2003).

Tabla 4. Letras en la lengua española: Frecuencia de uso

e	13.68%	c	4.68%	q	0.88%
a	12.53%	t	4.63%	h	0.70%
o	8.68%	u	3.93%	f	0.69%
s	7.98%	m	3.15%	z	0.52%
r	6.87%	p	2.51%	j	0.44%
n	6.71%	b	1.42%	ñ	0.31%
i	6.25%	g	1.01%	x	0.22%
d	5.86%	v	0.90%	w	0.02%
l	4.97%	y	0.90%	k	0.01%

Nota: En la tabla se cuenta la frecuencia de las vocales con acento junto con ellas sin acento (p. ej., en “Panamá”, “á” = 3). Se miden las “ch,” “ll,” y “rr” como si fueran letras individuales. EGRA normalmente incluye estas letras al menos una vez cada una.

Fuente:

http://en.wikipedia.org/wiki/Letter_frequencies#Relative_frequencies_of_letters_in_other_languages (consultado el 12 de octubre de 2009).

Las tablas de frecuencia de letras dependen del texto analizado (un informe de rayos x o xilófonos necesariamente mostrará una frecuencia más alta de la letra x que el texto promedio). Se cuenta con estas tablas para español, francés y otras lenguas alfabéticas comunes.¹³ Los diseñadores de pruebas que construyen instrumentos en lenguas locales deberían analizar textos electrónicos para así preparar tablas de frecuencia de letras similares. Para diseñar una de ellas, tome un documento grande y representativo en Microsoft Word y use el comando “Buscar” (en el menú Edición en Office 2003; o Ctrl+B). Ingrese la letra “t” en el cuadro “Buscar” y marque el recuadro “Resaltar todos los elementos encontrados” en el documento principal. Haga clic en “Buscar todos” y Microsoft Word resaltará cada vez que la letra “t” aparezca en el documento, e informará del número de veces que apareció (por ejemplo, en el caso de esta versión de la manual en español, la letra “t” aparece más de 12,000 veces). Repita este proceso para cada letra del alfabeto, registrando el número total para cada letra hasta que puede calcular la proporción en que cada una de ellas aparece como parte del número total de letras en el documento.

Las cuestiones de pronunciación deben manejarse con tacto en ésta y otras tareas. Aquí no se trata de probar la pronunciación “correcta”, donde lo “correcto” es interpretado como la adhesión a algún estándar que indica un estatus socioeconómico privilegiado. De lo que se trata es de probar el automatismo en el uso de una pronunciación que podría ser común en una

¹³ Se cuenta con ellas para español, francés, alemán, italiano, esperanto, turco y sueco en http://en.wikipedia.org/wiki/Letter_frequencies#Relative_frequencies_of_letters_in_other_languages (consultado el 10 de febrero de 2009).

región o una forma de español dada. Así, los acentos regionales son aceptables al juzgar si se nombra correctamente una letra.

Normas de corrección y puntuación. El puntaje del niño en esta subprueba debe calcularse como el número de letras leídas correctamente por minuto. De completar todas las letras antes de finalizar el tiempo estipulado, deberá registrarse el tiempo de ejecución y los cálculos deberán basarse en dicho periodo de tiempo. Los examinadores deberían marcar toda letra incorrecta con una barra oblicua (/), colocar un corchete (]) después de la última letra nombrada, y registrar el tiempo que resta en el cronómetro a la terminación del ejercicio (las variables son como sigue: total de letras leídas, total incorrectas, tiempo restante en el cronómetro). Estos tres indicadores se usan entonces para calcular el total de letras leídas correctamente por minuto (LCPM):

$$\text{LCPM} = \frac{(\text{Total letras leídas} - \text{Total incorrectas})}{[(60 - \text{Tiempo restante en el cronómetro}) / 60]}$$

Cada uno de estos indicadores puede asimismo usarse para análisis adicionales. Por ejemplo, la información sobre el número total de letras nombradas permitirá distinguir entre un educando que nombra 50 letras en un minuto pero sólo la mitad de ellas correctamente, y otro que nombra 25 en igual periodo de tiempo, pero todas correctamente.

Nótese que esta tarea, al igual que muchas de las siguientes, no sólo está cronometrada, sino también limitada por el tiempo (i.e., se la detiene después de un periodo de tiempo especificado, haya sido completada o no). La limitación temporal es útil para hacer que la evaluación sea más corta, además de menos estresante tanto para el niño como para el evaluador, puesto que el primero no tiene que seguir intentando hacer toda la tarea a un ritmo lento. Además, el cronometrado ayuda a evaluar el automatismo.

Para cada una de las tareas cronometradas que figuran en una sección posterior, los examinadores deberían además registrar únicamente la información indicada arriba. Hacer que calculen en el campo resultados tales como las LCPM les distrae del proceso de evaluación, y puede llevar a errores significativos.

Construcción de ítems. Las letras del alfabeto deberían distribuirse al azar, 10 por línea, y deberían estar distribuidas homogéneamente en mayúsculas y minúsculas. Los porcentajes deberían servir de guía de la frecuencia con que las letras aparecen en la hoja de tareas (i.e., en español, en una lista de 100 letras, la “E” debería aparecer aproximadamente 14 veces, la “A” 13, etc.).

Ejemplo de diseño de evaluación: conocimiento del nombre de las letras

[Muestre al estudiante la hoja plastificada de letras. Diga:] Te voy a mostrar una serie de letras para que las leas. Empiezo con un ejemplo: esta letra [señala la C] es “C”.

Ahora inténtalo con esta otra letra. Nombra esta letra [señala la letra “T”]:

[si el estudiante responde correctamente, diga]: bien, esta letra es “T”.

[si el estudiante no responde correctamente, diga]: esta letra se llama “T”.

¿Comprendes lo que debes hacer? Cuando te diga “comienza”, lee las letras lo mejor que puedas. Yo me voy a callar y te voy a escuchar, a menos que necesites ayuda. Cuando pase un tiempo voy a decir “alto” para que te detengas. Pon tu dedo debajo de la primera letra. ¿Listo(a)? Comienza por favor.

Active el cronómetro cuando el niño(a) comience a leer. Es importante que marque las letras incorrectas con una barra diagonal. Dé por correctas las auto-correcciones. **Permanezca en silencio**, excepto si el estudiante duda por 3 segundos, léale la letra, indique la siguiente letra y diga “Por favor continúa”. Marque la letra que Ud. le leyó como incorrecta. Después de un minuto diga “alto”. Marque con un corchete (|) la última letra intentada antes de que usted dijera “alto”. **Si el estudiante se equivoca en todas las letras de la primera línea, pare el ejercicio.**

DETENGA AL ESTUDIANTE A LOS 60 SEGUNDOS Y MARQUE CON UN CORCHETE (|) LA LETRA DONDE DEJÓ DE LEER.

V	I	e	m	S	y	h	ñ	L	N	10
I	K	T	D	H	T	a	d	z	w	20
r	ch	z	m	U	e	j	G	X	u	30
g	R	B	Q	I	f	J	Z	s	r	40
B	n	C	B	p	Y	F	c	a	E	50
y	s	L	P	M	v	O	t	n	P	60
Z	A	e	x	f	F	r	u	A	t	70
Ch	G	T	b	S	I	g	m	i	I	80
L	L	o	q	a	N	E	Y	p	x	90
N	k	c	D	d	y	b	j	R	v	100

2.1 Anotar el tiempo total transcurrido al detenerse el cronómetro: _____

2.2 Anotar total letras leídas en tiempo anotado: _____

2.3 Anotar total letras incorrectas en tiempo anotado: _____

2.4 Anotar total letras correctas en tiempo anotado: _____

2. Conocimiento de los sonidos de las letras

El conocimiento de cómo las letras corresponden a los sonidos, es otra habilidad crucial que los niños deben dominar para ser lectores exitosos. Las correspondencias letra-sonido se suelen

enseñar con enfoques basados en el método fonológico, que ganaron y perdieron popularidad en las últimas décadas. Al igual que el ejercicio de descodificación de pseudopalabras no familiares, es probable que esta subprueba resulte algo controversial entre algunos grupos de educadores. El conocimiento del sonido de las letras es un enfoque de evaluación bastante común y se le usa en varias evaluaciones de la lectura temprana, entre ellos la *Preschool Comprehensive Test of Phonological and Print Processing* (Lonigan, Wagner, Torgesen y Rashotte, 2002). La evaluación abarca una serie de letras.

Normas de corrección y puntuación. Al igual que en el ejercicio de nombrar las letras, el puntaje del niño se calcula en esta subprueba en base al número de sonidos correctos de las letras por minuto.

Construcción de ítems. Para evaluar el conocimiento del sonido de las letras, se debe usar la misma página laminada de letras, esto es, la que se usa en la primera subprueba del conocimiento del nombre de las mismas. En el caso de consonantes que pueden representar más de un sonido (i.e., c, g), ambas respuestas serán aceptables. Los alumnos pueden tener problemas para eliminar el sonido de la vocal frecuentemente asociada con las consonantes; en estos casos se acepta como respuesta correcta ya sea /b/ o /buh/. Durante la formación, los examinadores y supervisores deben revisar cuidadosamente las posibles pronunciaciones de cada letra.

Ejemplo de diseño de evaluación: Conocimiento de los sonidos de las letras

[Muestre al estudiante la hoja plastificada de letras. Diga:] **Aquí tienes una serie de letras para que me digas sus sonidos. Empiezo con un ejemplo: el sonido de esta letra [señala la F] es “fffff”. Ahora inténtalo tú con esta otra letra. Dime el sonido de esta letra [señala la letra “M”]:**

[si el estudiante responde correctamente, diga]: **bien, el sonido de esta letra es “/mmm/”.**

[si el estudiante no responde correctamente, diga]: **el sonido de esta letra es “/mmm/”.**

¿Comprendes lo que debes hacer? Cuando te diga “empieza”, dime los sonidos de las letras lo mejor que puedas. Yo me voy a callar y te voy a escuchar, a menos que necesites ayuda. Cuando pase un tiempo voy a decir “alto” para que te detengas. Pon tu dedo debajo de la primera letra. ¿Listo(a)? Empieza por favor.

Active el cronómetro cuando el niño(a) comience a leer. Es importante que marque las letras incorrectas con una barra diagonal. Dé por correctas las auto-correcciones. **Permanezca en silencio**, excepto si el estudiante duda por 3 segundos, dígame el sonido, indique la siguiente letra y diga **“Por favor continúa”**. Marque la letra como incorrecta en la que Ud le prestó ayuda. Después de un minuto diga “alto”. Marque con un corchete (|) la última letra intentada antes de que usted dijera “alto”. **Si el estudiante se equivoca en todas en la primera línea, pare el ejercicio.**

DETENGA AL ESTUDIANTE A LOS 60 SEGUNDOS Y MARQUE CON UN CORCHETE () LA LETRA DONDE DEJÓ DE LEER

V	I	e	m	S	y	f	ñ	L	N	10
l	K	T	D	n	T	a	d	z	w	20
r	ch	z	m	U	e	j	G	X	u	30
g	R	B	Q	I	f	J	Z	s	r	40
B	n	C	B	p	Y	F	c	a	E	50
y	s	LI	P	M	v	O	t	n	P	60
Z	A	e	x	f	F	r	u	A	t	70
Ch	G	T	b	S	I	g	m	i	l	80
L	L	o	q	a	N	E	Y	p	x	90
N	k	c	D	d	y	b	j	R	v	100

2.1 Anotar el tiempo total transcurrido al detenerse el cronómetro: _____

2.2 Anotar total letras leídas en tiempo anotado: _____

2.3 Anotar total letras incorrectas en tiempo anotado: _____

2.4 Anotar total letras correctas en tiempo anotado: _____

3. Conciencia fonológica

Primer enfoque: Identificación del sonido inicial. El primer enfoque para evaluar la conciencia fonológica es hacer que los educandos identifiquen el primer sonido en una selección de palabras comunes. El ejemplo que aparece abajo usa 10 grupos de palabras simples y le pide a los alumnos que identifiquen el sonido inicial en cada una de ellas. El examinador lee cada palabra en voz alta dos veces, antes de pedirle al educando que identifique el sonido.

Normas de corrección y puntuación. El examinador anota el número de respuestas correctas. Este no es un segmento cronometrado de la evaluación.

Construcción de ítems. Deberían escogerse palabras simples de no más de dos sílabas, de entre las listas de palabras de primer o segundo grado. Se recomienda usar sólo palabras de una o dos sílabas, para así no hacer trabajar más de la cuenta la memoria de trabajo de los educandos.

Ejemplo de diseño de evaluación: Conciencia fonológica

Sabemos que cada letra tiene un sonido, por ejemplo, la letra M suena así: /mmm/. Ahora, voy a leerte algunas palabras para que me digas su primer sonido.

Por favor escucha bien y dime el sonido con que comienza cada palabra. Empiezo con “mamá”; el primer sonido de “mamá” es /mmm/. Practiquemos juntos ¿Cuál es el primer sonido de mamá? /mmm/

[Si lo hace incorrecto, diga:] Practiquemos de nuevo la palabra “mamá”, cuyo primer sonido es /mmm/. [Marcar con énfasis el sonido /mmm/]. Dime cuál es el primer sonido de “mamá”. [Si dice, /mmm/ diga] ¡Muy bien!. El primer sonido es /mmm/.

Practiquemos con otra palabra, ¿Cuál es el primer sonido de la palabra “sí”?

[Si dice, /sss/ diga]: ¡Muy bien! , El primer sonido de “sí” es /ssss/ .

[Si lo dice incorrectamente diga]: El primer sonido de “sí” es /ssss/ .

¿Comprendes lo que debes hacer? ¿Listo(a)? Te voy a dar la primera palabra. ¿Cuál es el primer sonido de la palabra “__” ? [Repita cada vez estas instrucciones. En los 2 ejercicios marcar correcto o incorrecto para cada palabra. Si el estudiante se equivoca en las primeras cinco palabras, pare el ejercicio. Debe dar 15 segundos al máximo por palabra].

¿Cuál es el primer sonido de la palabra“	” ? [Marque correcto o incorrecto]		
sol	/s/	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
mal	/m/	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
dedo	/d/	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
tiza	/t/	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
cena	/c/	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
mesa	/m/	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
luna	/l/	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
foca	/f/	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
preso	/p/	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
tribu	/t/	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>

Segundo enfoque: Identificación de palabras que inician con el mismo sonido. El segundo enfoque para evaluar la conciencia fonológica es hacer que los educandos identifiquen cual de tres palabras comienza con un sonido distinto. El examinador lee las palabras en voz alta dos veces, antes de pedirle al educando que identifique qué palabra comienza con un sonido distinto.

Normas de corrección y puntuación. El examinador anota el número de respuestas correctas. Este segmento de la evaluación no está cronometrado.

Construcción de ítems. Deberían escogerse palabras simples de no más de dos sílabas, de entre las listas de palabras de primer o segundo grado. Se recomienda usar sólo palabras de una o dos sílabas, para así no hacer trabajar más de la cuenta la memoria de trabajo de los educandos.

Ahora, vamos a hacer otro ejercicio. Sabemos que las letras tienen sonido, por ejemplo, la letra R suena: /rrr/ y la S suena /sss/. Por favor escucha bien para que me digas cuál de estas palabras comienza con un sonido diferente a las otras dos. Practiquemos: “rana; rata; sala.” ¿Qué palabra comienza con un sonido distinto? [Si dice “sala” diga] ¡Muy bien!

[Si lo hace incorrecto, diga:] Practiquemos de nuevo, te voy a dar otras tres palabras para que me digas cuál comienza con un sonido distinto: “traste; tripa; preso.” [Si dice “preso” diga] ¡Muy bien! La palabra que comienza con un sonido distinto es “preso”.

[Si comete error, diga:] La palabra que comienza con un sonido distinto es preso. Inicie.

¿Cuál de las palabras comienza con un sonido distinto? [Marque correcto o incorrecto]

pato	pelo	luna	luna	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
rojo	masa	remo	masa	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
carro	foca	faro	carro	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
misa	mapa	beso	beso	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
lapa	cama	cola	lapa	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
plato	pluma	grado	grado	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
flaco	plomo	flecha	plomo	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
blusa	traje	triste	blusa	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
bravo	cráter	broma	cráter	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>
grasa	gruta	frita	frita	Correcto <input type="checkbox"/>	Incorrecto <input type="checkbox"/>

4. Conocimiento de palabras simples

Las capacidades descodificadoras de los niños a menudo son evaluadas usando listas de lectura de palabras no relacionadas entre sí. Esto permite tener una medida más pura del reconocimiento de palabras y las habilidades de descodificación que la lectura de párrafos completos, puesto que los niños no pueden adivinar la siguiente palabra por el contexto. Para esta evaluación, las palabras familiares deben ser palabras de alta frecuencia seleccionadas en materiales de lectura de los primeros grados, y libros de cuentos de primero, segundo y tercer grado (de dificultad creciente).¹⁴ En español contamos con algunos estudios normativos sobre frecuencia de las palabras que se pueden consultar—por ejemplo, ver Alameda y Cuetos (1995), Davis y Perea (2005), Guzmán y Jiménez (2001), Justicia (1995), o Sebastián, Martí, Carreiras y Cuetos, 2000.

Normas de corrección y puntuación. El examinador anota el número de palabras leídas correctamente por minuto. Si el niño completa todas las palabras antes de finalizar el tiempo, deberá anotarse el tiempo transcurrido y los cálculos deberán tener dicho periodo de tiempo como base. Deben anotarse y puntuarse las palabras leídas correctamente por minuto. Las mismas tres variables recogidas en el ejercicio del nombre de las letras, deben reunirse en éste y en los otros ejercicios cronometrados (i.e., total de palabras leídas, total de palabras incorrectas, tiempo restante en el cronómetro). Para los cálculos véase el examen descrito anteriormente.

Construcción de ítems. Las palabras deberán disponerse horizontalmente, con una buena separación y una letra clara y familiar (minúsculas) en 10 hileras, cinco por línea. La fuente

¹⁴ Para revisar ejemplos en español, ver [http://es.wiktionary.org/wiki/Wikcionario:Frecuentes-\(1-1000\)-Subt%C3%ADtulos_de_pel%C3%ADculas](http://es.wiktionary.org/wiki/Wikcionario:Frecuentes-(1-1000)-Subt%C3%ADtulos_de_pel%C3%ADculas); en francés, <https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=174>.

usada deberá ser similar en tamaño y estilo a la que se usa en los libros de lectura oficiales, o de no existir éstos, en los que se compran con más frecuencia.

Ejemplo de diseño de evaluación: Conocimiento de palabras simples

[Muestre al estudiante la hoja plastificada de palabras. Diga:] **Aquí tienes una serie de palabras para que las leas. Empiezo con un ejemplo: esta palabra es “mi”. Ahora inténtalo con esta otra palabra. [señale la siguiente palabra: el] Léela en voz alta.**

[Si el estudiante dice “el”, diga]: **“Muy bien: el”.**

[Si el estudiante no dice “el” correctamente, diga]: **Esta palabra es “el”.**

¿Comprendes lo que debes hacer? Cuando diga “empieza”, lee las palabras lo más rápido que puedas sin equivocarte. Cuando pase un tiempo voy a decir “alto” para que te detengas. Pon tu dedo debajo de la primera palabra. ¿Listo? Comienza por favor.

Active el cronómetro cuando el niño(a) comience a leer. Es importante que marque las palabras incorrectas con una barra diagonal. Dé por correctas las auto-correcciones. **Permanezca en silencio**, excepto si el estudiante duda por 3 segundos, léale la palabra, indique la siguiente palabra y diga **“Por favor continúa”**. Marque la palabra que Ud. le leyó como incorrecta. Después de un minuto diga “alto”. Marque con un corchete (|) la última palabra intentada antes de que usted dijera “alto”. **Marcar como incorrecto si el niño no respeta la tilde en las palabras acentuadas. Si tuvo que marcar como incorrectas todas las palabras en la primera línea, pare el ejercicio.**

DETENGA AL ESTUDIANTE A LOS 60 SEGUNDOS Y MARQUE CON UN CORCHETE (|) LA PALABRA DONDE DEJÓ DE LEER.

sol	pesa	rojo	he	come	5
masa	nos	al	niña	lápiz	10
la	lobo	solo	cama	toma	15
casa	las	tela	muy	ojo	20
en	hay	alto	una	parte	25
mano	ellos	si	amo	que	30
sal	techo	carro	mesa	pelo	35
niño	vaso	todo	vaca	malo	40
rana	azul	bajo	pata	corre	45
hoja	papá	plato	tiza	fue	50

4.1 Anotar el tiempo total transcurrido al detenerse el cronómetro: _____

4.2 Anotar total palabras leídas en tiempo anotado: _____

4.3 Anotar total palabras incorrectas en tiempo anotado: _____

4.4 Anotar total palabras correctas en tiempo anotado: _____

5. Descodificación de palabras sin sentido (pseudopalabras)

La lectura de pseudopalabras es una medida de la capacidad descodificadora y está diseñada para evitar el problema de que puedan ser reconocidas visualmente. Para ser lectores exitosos, los niños deben combinar tanto habilidades de descodificación como de reconocimiento visual; las pruebas que no incluyen un ejercicio de descodificación pueden exagerar la capacidad del niño para leer palabras no familiares (puesto que las palabras evaluadas pueden formar parte de su vocabulario visual).

Normas de corrección y puntuación. Se calcula el puntaje del niño como el número de pseudopalabras leídas correctamente por minuto. Si el niño completa todas las palabras antes de finalizar el tiempo estipulado, deberá anotarse el tiempo transcurrido y los cálculos deberán tener como base dicho periodo de tiempo. Las mismas tres variables recogidas en los ejercicios del nombre de las letras y de lectura de palabras, deben reunirse en éste y en los otros ejercicios cronometrados (i.e., total de pseudopalabras leídas, total de pseudopalabras incorrectas, tiempo restante en el cronómetro). Para los cálculos véase el examen descrito anteriormente.

Construcción de ítems. Esta parte de la evaluación debe incluir una lista de 50 pseudopalabras de dos sílabas, cinco por hilera, con el siguiente patrón de letras (C = consonante, V = vocal): CV, VC, CVC. (Esto se puede adaptar según la lengua.) Las formas deberían ser correctas para la lengua, usando letras en posiciones legítimas (v.g., “nd” no porque en español no se usa la “nd” como sonido final), deberían limitarse a combinaciones consonante-vocal típicas de la lengua, y no deberían ser homófonas de palabras reales (v.g., no “harbol”, homófona de “árbol”). Deberán estar dispuestas en hileras (cinco pseudopalabras en cada una), usando una letra clara y bien espaciada.

Ejemplo de diseño de evaluación: Descodificación de palabras sin sentido

[Muestre al estudiante la hoja plastificada de palabras sin sentido. Diga] **Aquí tienes una serie de palabras inventadas para que las leas. Empiezo con un ejemplo: esta palabra inventada es “ut”. Ahora inténtalo tu con esta otra. Léela en voz alta**
[señale: ut].

[Si el estudiante dice “ut”, diga]: **“Muy bien: ut”.**

[Si el estudiante no dice “ut” correctamente, diga]: **Esta palabra inventada es “ut”.**

¿Entiendes lo que debes hacer? Cuando diga “empieza”, lee las palabras lo más rápido que puedas sin equivocarte. Cuando pase un tiempo voy a decir “alto” para que te detengas. Pon tu dedo debajo de la primera palabra. ¿Listo? Empieza por favor.

Active el cronómetro cuando el niño(a) comience a leer. Es importante que marque las palabras incorrectas con una barra diagonal. Dé por correctas las auto-correcciones. **Permanezca en silencio**, excepto si el estudiante duda por 3 segundos, léale la palabra, indique la siguiente palabra y diga **“Por favor continúa”**. Marque la palabra que le dijo como incorrecta. Después de un minuto diga “alto”. Marque con un corchete (I) la última palabra intentada antes de que usted dijera “alto”. **Si tuvo que marcar como incorrectas todas las palabras en la primera línea, pare el ejercicio.**

DETENGA AL ESTUDIANTE A LOS 60 SEGUNDOS Y MARQUE CON UN CORCHETE (J) LA PALABRA DONDE DEJÓ DE LEER.

lete	quibe	bofa	mise	garo	5
cafa	celu	bede	lura	mesi	10
lluno	rite	duso	jata	fica	15
luma	alti	lufa	rate	dulte	20
ledo	fosu	gesa	lemo	golpa	25
bosa	rale	flano	rabu	bulo	30
luva	arcu	cince	llusia	firta	35
onti	zaca	queno	bana	juru	40
foba	lise	vodo	tuzi	listu	45
quira	cuto	ganco	rafo	duba	50

5.1 Anotar el tiempo total transcurrido al detenerse el cronómetro: _____

5.2 Anotar total palabras sin sentido leídas en tiempo anotado: _____

5.3 Anotar total palabras sin sentido incorrectas en tiempo anotado: _____

5.4 Anotar total palabras sin sentido correctas en tiempo anotado: _____

6. Lectura y comprensión de un párrafo (texto vinculado)

La fluidez en la lectura oral es una medida de la competencia global en la lectura: la capacidad de traducir letras en sonidos, unificar éstos en palabras, procesar conexiones, relacionar el texto con significados, y hacer inferencias para deducir información no explícita en el texto (Hasbrouck y Tindal, 2006). A medida que los lectores competentes traducen el texto en el lenguaje hablado, combinan estas tareas aparentemente sin esfuerzo; dado que la fluidez de la lectura oral capta este complejo proceso, podemos usarle para caracterizar la habilidad global de lectura. Se ha mostrado que las pruebas de fluidez de la lectura oral, medida por evaluaciones cronometradas de las palabras leídas correctamente por minuto, tienen una fuerte correlación (0.91) con la subprueba de Comprensión de Lectura de la Stanford Achievement Test (Fuchs et al., 2001). Un desempeño deficiente en una prueba de comprensión lectora sugeriría que el educando tiene problemas con la descodificación, o con leer con suficiente fluidez como para comprender, o con el vocabulario.

Normas de corrección y puntuación. Se asigna un puntaje a los educandos sobre el número de palabras leídas correctamente por minuto y el número de preguntas de comprensión respondidas aceptablemente. Habrán tres puntajes por alumno: la proporción de palabras leídas, el tiempo por palabra y la proporción de preguntas correctamente respondidas. Las mismas tres variables recogidas en los ejercicios—del nombre de las letras, de lectura de palabras y de lectura de pseudopalabras—deben reunirse en éste y en los otros ejercicios cronometrados (i.e., total de palabras leídas, total de palabras incorrectas, tiempo restante en el

cronómetro). Para los cálculos véase el examen descrito anteriormente. Debería, además, recogerse e introducir en la base de datos resultados para cada una de las preguntas de comprensión, con una variable de puntaje final calculada como parte del total de preguntas contestadas. Sólo deberían hacerse preguntas sobre el texto que el niño ha leído (véase la estructura de las preguntas y el párrafo en una sección posterior).

Construcción de ítems. Para construir esta prueba de evaluación, los examinadores deberían revisar párrafos de materiales de lectura para niños (no del libro de texto escolar). Una historia narrativa debería tener una sección de inicio en donde se presentan los personajes, una sección media que contiene cierto dilema y una sección final con una acción que resuelve al mismo. No debería ser una lista de oraciones con poca relación entre sí. Deberán evitarse los nombres típicos de personajes de los libros de texto, puesto que los educandos podrían dar respuestas automatizadas en base a los relatos con los cuales están familiarizados. Los nombres y lugares deberían reflejar la cultura local y las narrativas deben tener un personaje principal, comienzo, sección media y final. Los textos deberían contener algo de vocabulario (formas declinadas, derivaciones, etc.) y estructuras de oraciones complejas. Para facilitar la lectura por parte de los educandos, debería usarse una letra grande, clara y familiar, y un buen espaciado entre las líneas. No debería incluirse ninguna imagen. Las preguntas de comprensión deberían incluir preguntas de selección y con base en hechos, y al menos una que requiere efectuar una inferencia a partir del texto.

Ejemplo de diseño de evaluación: Lectura y comprensión de un párrafo

[Muestre al estudiante el pasaje en la última página del formulario plastificado. Diga,]

Aquí tienes un cuento para que lo leas. Quiero que leas esto en voz alta. Cuando termines, te haré algunas preguntas sobre el cuento. ¿Comprendes lo que debes hacer? Cuando te diga “comienza”, lee el cuento lo más rápido que puedas sin equivocarte. Cuando pase un tiempo voy a decir “alto” para que te detengas. Pon tu dedo debajo de la primera palabra. ¿Listo? Comienza por favor.

Active el cronómetro cuando el niño(a) comience a leer. Es importante que marque las palabras incorrectas con una barra diagonal. Dé por correctas las auto-correcciones. **Permanezca en silencio**, excepto si el estudiante duda por 3 segundos, léale la palabra, indique la siguiente palabra y diga **“Por favor continúa”**. Marque la palabra que le como incorrecta. Después de un minuto diga “alto”. Marque con un corchete (I) la última palabra intentada antes de que usted dijera “alto”. **Si tuvo que marcar como incorrectas todas las palabras en la primera línea, pare el ejercicio.**

DETENGA AL ESTUDIANTE A LOS 60 SEGUNDOS Y MARQUE CON UN CORCHETE (|) LA PALABRA DONDE DEJÓ DE LEER. HAGA LAS PREGUNTAS INMEDIATAMENTE DESPUÉS.

María tiene una gata. La gata es negra y gorda. Le 11
gusta jugar y brincar. Un día, María no pudo encontrar 21
a la gata. María y su mamá la buscaron por toda la 33
casa. La gata estaba debajo de la cama. La gata tuvo 44
tres gatitos. La mamá de María le dice – Yo también 54
voy a tener un bebé. Vas a tener un hermanito. 64

- 6.1 Anotar el tiempo total transcurrido al detenerse el cronómetro: _____
6.2 Anotar total palabras leídas en tiempo anotado: _____
6.3 Anotar total palabras incorrectas en tiempo anotado: _____
6.4 Anotar total palabras correctas en tiempo anotado: _____

Quitarle el texto al estudiante después de que lo haya leído. Solamente hacer preguntas hasta donde esté el corchete. Después de leer cada pregunta, dé al estudiante cuando mucho 15 segundos para responder cada pregunta. **Después de cada pregunta marque la casilla adecuada.**

Me gustaría hacerte algunas preguntas sobre el cuento que acabas de leer.

- 6.5 ¿Quién tiene una gata?**
[María] Correcto Incorrecto Rehusó responder
- 6.6 ¿A quién le gusta jugar y brincar?**
[A la gata, a la gata de María] Correcto Incorrecto Rehusó responder
- 6.7 ¿Qué le pasó a la gata de María?**
[Se perdió, tuvo gatitos] Correcto Incorrecto Rehusó responder
- 6.8 ¿Dónde estaba la gata?**
[Debajo de la cama] Correcto Incorrecto Rehusó responder
- 6.9 ¿Qué le dijo la mamá a María?**
[Un bebé o un niño] Correcto Incorrecto Rehusó responder

7. Comprensión oral

Una evaluación de la comprensión oral involucra la lectura de párrafos en voz alta por parte del examinador; los educandos responden luego a preguntas o enunciados de comprensión oral. La evaluación de la comprensión oral por separado de la de lectura es importante, debido a las distintas formas en que los alumnos enfocan, procesan y responden al texto. Las pruebas de comprensión oral ya existen hace cierto tiempo y se las ha usado en particular como una

evaluación alternativa para niños desfavorecidos, con un acceso relativamente reducido al lenguaje impreso (Orr y Graham, 1968). La finalidad de esta evaluación es ver si el educando es capaz de escuchar un pasaje oral y luego responder a varias preguntas correctamente con una palabra o con un enunciado simple. Un rendimiento pobre en una prueba de comprensión oral sugeriría que los niños simplemente no tienen el vocabulario básico que contienen los materiales de lectura, o que tienen dificultades para procesar lo que escuchan.

Normas de corrección y puntuación. Se puntúa a los educandos en función al número de enunciados correctos que dan como respuesta (fuera del número total de preguntas). Los diseñadores del instrumento deberían evitar preguntas que tengan respuestas únicamente de “sí” o “no”.

Construcción de ítems. Los párrafos deberían tener alrededor de 50 palabras de longitud y narrar una actividad o evento que sea familiar para los niños locales. Deberían incluirse preguntas de selección e inferencia.

Ejemplo de diseño de evaluación: Comprensión oral

[No hay una hoja plastificada para esta parte del diagnóstico. Diga:] **Voy a leer un párrafo breve una vez, y luego te haré algunas preguntas. Por favor presta mucha atención mientras leo y responde a las preguntas lo mejor que puedas. ¿Listo? Comenzamos.** [Lee en voz alta y clara UNA VEZ el texto a un ritmo constante y pausado.]

El sábado fue un día divertido. Juana y su familia fueron al partido de fútbol de Juana. Juana era la portera del equipo. El juego estaba empatado. En el último momento un jugador del otro equipo intentó un gol pero Juana lo paró. Sus amigas la felicitaron al final del partido.

- 7.1 ¿Cómo fue el sábado?
[Divertido] Correcto Incorrecto Rehusó responder
- 7.2 ¿A dónde fueron Juana y su familia?
[Al partido de fútbol] Correcto Incorrecto Rehusó responder
- 7.3 ¿Qué pasó al final del partido?
[El otro equipo intentó un gol, o Juana lo paró] Correcto Incorrecto
Rehusó responder
- 7.4 ¿Cómo quedó el partido?
[Empatado] Correcto Incorrecto Rehusó responder
- 7.5 ¿Por qué sus amigas felicitaron a Juana?
[Porque salvó el partido, porque paró el gol] Correcto Incorrecto
Rehusó responder

8. Dictado

Los docentes usan frecuentemente la evaluación del dictado para evaluar las habilidades tanto de comprensión oral como de escritura. Como ya vimos, el proceso de lectura puede también evaluarse en reversa: la habilidad de los educandos para oír sonidos y escribir correctamente letras y palabras correspondientes a los sonidos que oyen, demuestra su éxito con el principio

alfabético. Una serie de paquetes de evaluación ofrecidos por especialistas en el diseño de pruebas comerciales le dan al docente instrucciones sobre cómo diseñar y puntuar sus propias evaluaciones. Esta evaluación particular está inspirada por los modelos promovidos por el Educational Testing Service (2005) y la Children’s Literacy Initiative (2000), y respaldado por las investigaciones de la International Reading Association (Denton, Ciancio y Fletcher, 2006).

Normas de corrección y puntuación. Los educandos son puntuados en una escala simple que capta la exactitud de los sonidos de vocales y consonantes, la ortografía, el espaciamiento y la dirección del texto, las mayúsculas y la puntuación. Cada categoría tiene un total de 2 puntos posibles para una exactitud total, 1 para cierta exactitud y 0 para nada de exactitud (véase el criterio de puntaje en el ejemplo de diseño de evaluación que se presenta a continuación). Durante el análisis se suman estas variables para dar una variable de un único puntaje.

Construcción de ítems. La oración dictada debería tener cuando mucho 10 palabras e incluir al menos una palabra difícil o irregular.

Ejemplo de diseño de evaluación: Dictado

[No hay una hoja plastificada para este parte del diagnóstico. El estudiante debe escribir detrás de la página 6, usando la tableta del encuestador. Diga:] Vas a escribir **UNA VEZ** la oración que te voy a dictar. Escucha la primera vez, sin escribir, luego la voy a repetir dos veces más mientras la escribes. ¿Listo? Comenzamos. [Lea la siguiente frase lentamente en voz alta 3 veces, a un ritmo de una palabra por segundo.]

En la playa a Juan le cayó arena en la nariz y los ojos.

Criterio de Evaluación (PARA SUPERVISOR)	0=nada correcto	1=algo correcto	3= todo correcto
8.1 Escribió "playa" correctamente.			
8.2 Escribió "cayó" correctamente.			
8.3 Escribió "nariz" correctamente.			
8.4 Escribió "ojos" correctamente.			
8.5 Usó espacios y dirección de texto correctamente.			
8.6 Usó mayúsculas correctamente.			
8.7 Usó puntuación correctamente (punto final).			

Otros posibles componentes del instrumento y razones de su exclusión

Durante el diseño del instrumento, tanto el examen de la bibliografía como el proceso de revisión por parte de los expertos generaron numerosas sugerencias para que se incluyeran componentes y mediciones adicionales a las pruebas. A medida que se pasaba revista a cada una de estas sugerencias, se establecieron criterios de selección para la idoneidad de su inclusión en el instrumento. La principal consideración a tener en cuenta era la utilidad de cada habilidad medida en la prueba en la predicción del éxito futuro que el educando tendría en la lectura. Dado que hay poca bibliografía en el mundo en países en vías de desarrollo (y con lenguas múltiples) referida a la aplicación de estas medidas, RTI dependió de la bibliografía

existente, que proviene fundamentalmente de Estados Unidos y Europa, aunque sí se cuenta con algo de literatura para América Latina.

Destacaba entre estas sugerencias la inclusión de una subprueba basada en dibujos como la Peabody Picture Vocabulary Test, una prueba disponible comercialmente de Pearson Learning Group. Algunas variantes de las herramientas de evaluación de la lectura en los primeros grados (incluso una versión aplicada por Plan Internacional en francés en África occidental) incluyeron imágenes para identificar el conocimiento de un vocabulario común (como las partes del cuerpo: mano, cabeza, dedo del pie, etc.). Actualmente EGRA no incluye imágenes o pruebas de vocabulario en imágenes, por varias razones: (1) el vocabulario se mide indirectamente en los segmentos de lectura de palabras y párrafos, (2) el desarrollo de las imágenes frecuentemente se topa con el problema del copyright (por ejemplo, se descartó el uso de la PPVT porque tendría que obtenerse permiso del copyright cada vez que el instrumento se usase en otro país), y (3) una consideración importante fueron las dificultades que presenta el crear imágenes que sean universalmente apropiadas para todas las culturas y contextos culturales. Además, cuando se diseña y prepara las imágenes localmente a fin de evitar los problemas de copyright o hacer que sean culturalmente apropiadas, al menos dos problemas parecen surgir. En primer lugar, ellas a menudo son de muy baja calidad gráfica, lo que a veces hace que resulte difícil que hasta un adulto experimentado interprete la imagen y responda la pregunta. En segundo lugar, incluso asumiendo una alta calidad gráfica, el diseño de ítems apropiados en base a dibujos parece necesitar una competencia considerable.

Uno de los componentes inicialmente probados y posteriormente eliminados de la evaluación se derivó de Concepts About Print assessment, de Marie Clay (1993). En Gambia, Senegal y Nicaragua, el uso de tres de los ítems de Clay (3 a 5, reglas direccionales que incluían donde iniciar la lectura, en qué dirección leer y dónde leer luego) demostró efectos techo (casi todos los niños completaron la tarea exitosamente). Es más, derivando conclusiones de investigaciones tanto estadounidenses como internacionales, el *Handbook of Psychology* informa que la conciencia de la letra en forma escrita parece tener poco poder vaticinador de la posterior habilidad de lectura; sirve fundamentalmente como una medida sustituta de la exposición a los entornos de la letra y el alfabetismo (Paris y Paris, 2006). En base a estos resultados, así como a limitaciones de eficiencia y tiempo, la evaluación EGRA no incluye un segmento de “Conceptos sobre la letra escrita” [Concepts About Print], ya sea en forma reducida o completa (toda la batería contiene 24 ítems).

Traducción

El consenso que viene surgiendo entre expertos como los que se reunieron en la sesión de Washington de noviembre de 2006, así como en la sesión de diseño de Penny Chiappe con el departamento de educación de África del Sur, es que cuando los evaluadores buscan formas en que usar EGRA en las lenguas locales, no es una buena idea simplemente traducir las palabras o el párrafo, de una versión base en inglés (o en cualquiera otra lengua) a la lengua local. La traducción puede, por ejemplo, arrojar palabras muy largas en la lengua nativa. En lugar de ello se recomienda que se use un pasaje de dificultad aproximadamente igual a la del párrafo base. Deben usarse palabras simples de dos sílabas, oraciones cortas y una narrativa familiar. En la medida en que los textos de lectura han sido validados para que correspondan al currículo nacional, el uso de párrafos procedentes de los textos tenderá asimismo a evitar las cuestiones de la validez o idoneidad de la selección de textos de lectura. Una alternativa es pedir a docentes y expertos en currículo, versados en las reglas de las lenguas nativas, que armen un pasaje similar en nivel de dificultad al párrafo.

Como señalara Chiappe (memorando a RTI en base a la experiencia de África del Sur),

Dadas las diferencias lingüísticas (ortográficas y morfológicas), es crucial que los párrafos usados se escriban de manera independiente. No pueden establecerse equivalencias entre los párrafos traduciendo el de inglés a las distintas lenguas. Esto quedó claramente ilustrado con el estudio piloto inicial del pasaje isiZulu. Este último pasaje era una traducción del pasaje inglés. Aunque sería de esperar que el ritmo de lectura oral de los niños fuera similar en las listas de palabras/pseudopalabras libres de contexto y el pasaje, los estudiantes de isiZulu que podían leer 20-30 palabras correctamente por minuto en la lista no podían leer el pasaje en absoluto. Una inspección detenida del pasaje en isiZulu reveló que sus palabras eran mucho más largas que las de la lista en la misma lengua, o que las palabras usadas en el pasaje en inglés. Así, el pasaje en isiZulu era evidentemente demasiado difícil para que lo leyeran estudiantes al nivel de primer grado.¹⁵

Si los diseñadores de la prueba deciden usar pseudopalabras, es importante asegurarse de que la estructura silábica tenga sentido. En inglés, por ejemplo, las pseudopalabras con un patrón consonante-vocal-consonante (CVC), como “wub” o “dod”, son “legales” o consistentes con los patrones usuales de la lengua.

¹⁵ *Inglés*: “John had a little dog. The little dog was fat. One day John and the dog went out to play. The little dog got lost. But after a while the dog came back. John took the dog home. When they got home John gave the dog a big bone. The little dog was happy so he slept. John also went to sleep.” [“John tenía un perrito. El perrito era gordo. Un día John y el perro salieron a jugar. El perrito se perdió. Pero después de un rato el perro regresó. John llevó el perro a casa. Al llegar a ella, John le dio al perro un gran hueso. El perrito estaba feliz, de modo que se durmió. John también se fue a dormir”] *IsiZulu*: “USipho wayenenja encane. Inja yakhe yayikhuluphele. Ngolunye usuku uSipho wayehamba nenja yakhe ukuyodlala. Inja yalahleka. Emva kwesikhathi inja yabuya. USipho waphindela ekhaya nenja yakhe. Emva kokufika ekhaya, uSipho wapha inja ekhaya ukudla okuningi. Inja yajabula kakhulu yaze yacina ilele. NoSipho ngokunjalo wagcina elele.”

V. Formadores de EGRA y trabajo de campo

Como ya se indicó en la introducción de la Sección IV, se recomienda especialmente una semana de formación para los examinadores que llevarán a cabo el estudio piloto del instrumento. Esta sección está dirigida a los formadores, que llevarán a cabo la formación y supervisarán el trabajo de campo piloto. Se espera que muchos de estos formadores asimismo sean supervisores de los examinadores en el campo.

Idealmente, todos los participantes en la formación de los examinadores deberían también haber asistido al taller de adaptación e investigación descrito en la Sección IV, aunque de ser necesario, el equipo EGRA puede traer examinadores adicionales o aplicadores de la prueba en este punto, para complementar así al personal del ministerio y a los equipos de supervisores (véase el examen de las calificaciones de los examinadores en la Sección IV).

Esta sección presenta un examen global del trabajo de campo de EGRA y el proceso inicial de la aplicación piloto, incluyendo las lecciones aprendidas durante las múltiples aplicaciones piloto de los instrumentos EGRA. Sin embargo, no se trata de un manual exhaustivo del supervisor; este manual asume que un manual tal será diseñado por la asistencia técnica o los equipos-país, de establecerse que es necesario.

Centros de discusión de estos temas:

- Aplicando el piloto del instrumento
- Llegando al centro educativo
- Seleccionando a los educandos y llevando a cabo la evaluación
- Enseñando lecciones para el trabajo de campo

La experiencia piloto del instrumento

Como se describiera en la Sección IV, durante el taller de adaptación e investigación, un equipo del ministerio habrá examinado cada uno de los componentes del instrumento y las investigaciones subyacentes.

En esta formación de seguimiento, los participantes practicarán y tendrán una experiencia piloto de aplicación del instrumento en varios centros educativos (aproximadamente tres a seis, dependiendo del número de examinadores). Luego de la formación se llevará a cabo una aplicación a toda escala en los centros educativos escogidos de la muestra (para información sobre el muestreo véase el Anexo B). Las instrucciones en esta sección referidas a las evaluaciones con base en colegios, se aplican tanto a la prueba piloto como en la aplicación a toda escala. La tabla 5 muestra un ejemplo de agenda para esta etapa de la formación EGRA.

Tabla 5. Ejemplo de agenda: formación de examinadores y trabajo de campo piloto

Lunes	Martes	Miércoles	Jueves	Viernes
<ul style="list-style-type: none"> • Revisar los principios subyacentes • Revisar el borrador del instrumento • Realizar formación en el uso de la aplicación, y practicar 	<ul style="list-style-type: none"> • Realizar formación en el uso de la aplicación, y practicar • Revisar los papeles y responsabilidades de supervisores y examinadores 	<ul style="list-style-type: none"> • Llevar a cabo la aplicación piloto en 3 a 6 colegios (determinado por el número de equipos que vienen siendo instruidos) • Ingresar datos • Analizar resultados y modificar el instrumento 	<ul style="list-style-type: none"> • Imprimir versión final del instrumento • Realizar formación en el uso de la aplicación, y practicar • Cumplir prueba de confiabilidad entre evaluadores 	<ul style="list-style-type: none"> • Preparar y empaquetar los materiales • Finalizar y revisar la logística

Los objetivos de la formación incluyen:

- La revisión de los principios de EGRA a fin de entender mejor el razonamiento que subyace a los distintos componentes del instrumento.
- Solidifique la aplicación y práctica de formación del supervisor, así como sus papeles y responsabilidades (en este caso podría resultar útil el diseño del antedicho manual del supervisor).
- Haga que los examinadores escogidos lleven a cabo la aplicación piloto del instrumento en tres a seis centros educativos (como ejercicio).
- Continúe formando a los examinadores en la aplicación y la puntuación de EGRA.
- Finalice el instrumento y complete los preparativos logísticos (impresiones, etc.).
- Notifique a los centros educativos de la muestra su selección, la finalidad de su evaluación y las necesidades de logística (v.g., una habitación separada y silenciosa donde aplicar el instrumento).

Durante el taller, los participantes necesitarán:

- copias del borrador completo de los instrumentos de examinador y educando
- cronómetros o temporizadores (de ser posible uno de cocina, que cuente de 1 minuto hacia abajo)
- lápices con borrador y tablillas con sujetapapeles
- varias laptops con Excel para ingresar los datos (una por grupo de examinadores)

Los jefes de taller deben revisar y reforzar las habilidades evaluadas en el instrumento y las relaciones de cada componente con la instrucción. El instrumento debe revisarse cuidadosamente, prestando estrecha atención a la claridad de las instrucciones tanto para educandos como examinadores. En la medida de lo posible, las instrucciones para los alumnos deben ser consistentes (lo que incluye las formas de animar al niño, el número de ejemplos, etc.) a lo largo de todos los componentes de la prueba (para una lista de los componentes refiérase a la tabla 3).

El número de examinadores a contratar y formar dependerá del número de centros educativos a visitar y el marco temporal de la realización del ejercicio. Dado que la aplicación piloto en tres a seis de ellos es un ejercicio de preparación, todos los examinadores seleccionados para la aplicación global deberían tomar parte en el taller de la Semana 2. Los examinadores y supervisores deberían como mínimo visitar los centros educativos en grupos de cuatro: un supervisor y tres examinadores por equipo. De este modo los primeros pueden seleccionar los educandos y circular entre los examinadores durante el periodo de la prueba. En base a las aplicaciones de EGRA efectuadas en varios países hasta la fecha, se calcula que se requieren los resultados de 400 alumnos para cada grupo de comparación de interés (grado, tipo de centro educativo, etc.; para un examen detallado del tamaño de la muestra véase el Anexo B). De este modo, una simple línea de base nacional que compare educandos por grado en los grados 1 al 3, requerirá de 1200 de ellos. De necesitarse grupos de comparación adicionales (v.g., rural versus urbano, por grado), entonces serán necesarios 400 alumnos para cada grupo de comparación (en este ejemplo, 2400 alumnos).¹⁶ La tabla 6 que presentamos a continuación resume los cálculos para determinar el número de centros educativos y examinadores, necesarios para efectuar una encuesta que recoja resultados para los grados 1 a 3. En varios países donde hasta hoy se ha aplicado EGRA, el personal del ministerio que tomó parte en la Semana 1 del taller fue seleccionado como supervisor durante todo el ejercicio. Es importante contar con más supervisores y examinadores de lo necesario, para así poder escoger los mejores en ambos grupos.

Tabla 6. Estimados de educandos y centros educativos muestreados, y número de examinadores

Grupo de interés de comparación	Educandos de la muestra	Centros educativos de la muestra (60 alumnos por colegio)	Educandos por grado del colegio	Examinadores requeridos (4 equipos de 4)	Días de trabajo de campo (1 día por colegio por equipo)
Grados 1–3	1200	1200/60=20	20	16	5
Grados 1–3, Control vs. tratamiento	2400	2400/60=40	20	16	10
Grados 1–3, Control vs. tratamiento, urbano vs. rural	4800	4800/60=80	20	16	20

El número de examinadores puede subir o bajar, dependiendo de las necesidades del equipo del ministerio y del tiempo disponible para completar la evaluación. De este modo, una muestra de 40 centros educativos puede completarse en 10 días con 16 examinadores, o en 5 con 32. La aplicación en estos 40 colegios debería realizarse inmediatamente después de la aplicación piloto en la Semana 2.

Continuando con el ejemplo de la línea de base de una comparación simple de grados con 1200 alumnos, la aplicación piloto deberá llevarse a cabo en al menos cuatro centros

¹⁶ Véase en el Anexo B un respaldo estadístico adicional para esta estimación, usando resultados procedentes de diversas aplicaciones de EGRA.

educativos (uno por cada equipo de cuatro), dependiendo del número de examinadores que vienen siendo instruidos. Para la aplicación piloto, los participantes necesitarán:

- copias del instrumento final del examinador
- un conjunto laminado de formularios del educando por examinador (los mismos formularios laminados serán usados con cada alumno que el examinador evalúe)¹⁷
- cronómetros o temporizadores (de ser posible encuentre uno de cocina que cuente hacia abajo a partir de 1 minuto)
- lápices con borrador y tablillas con sujetapapeles
- lápices u otros materiales escolares menores para entregarlos a los educandos, como agradecimiento por su participación

Calculando la confiabilidad entre evaluadores

La confiabilidad entre evaluadores mide el grado en que distintos calificadores o examinadores coinciden en su puntuación de una misma observación. Por ello es muy importante asegurar bien el proceso de formación, para así mejorar el desempeño de los examinadores antes de que salgan al campo. También se la puede usar como ayuda en la selección de los examinadores de mejor desempeño.

Hay varias formas de generar datos para calcular la confiabilidad entre evaluadores:

1. Un examinador evalúa al educando mientras otro observa y puntúa al mismo tiempo. Ambos comparan entonces su puntaje y discuten. Los supervisores también pueden observar y puntuar con cada examinador, y discutir cualquier discrepancia.
2. En un entorno grupal pueden reproducirse grabaciones de audio o video de una evaluación de educandos, mientras que todos los examinadores la puntúan. Los formadores entonces pueden recoger las hojas de puntuación para revisarlas o comentarlas (verificar la codificación y la puntuación).
3. Los formadores o los examinadores adultos pueden actuar en entornos de grupos pequeños o grandes, y las hojas de puntuación pueden recogerse para revisarlas y comentarlas. El beneficio de este último escenario es que los adultos deliberadamente pueden cometer varios errores en cualquiera de las subpruebas dadas (v.g., saltarse o repetir palabras o líneas, variar el volumen de la voz, hacer pausa durante periodos de tiempo prolongados para conseguir algún apunte, etc.).

Con todas las estrategias antedichas, deben reunirse datos para así calcular la confiabilidad entre evaluadores. Los formadores-jefe deberían recoger las hojas de puntuación de cada subtarea, ingresarlas a Excel y calcular las medias y desviaciones estándar. Aquellos examinadores cuyos resultados sean mayores que una desviación estándar de la media podrían requerir de práctica o apoyo adicional. Si el análisis de la confiabilidad entre evaluadores revela un desempeño consistentemente deficiente de parte de un examinador, y si

¹⁷ Dado que los formularios de respuesta de los educandos se usarán con múltiples alumnos, la laminación, si bien no es del todo necesaria, sí prolongará su vida (también son útiles las láminas de plástico protectoras de páginas, insertadas en carpetas).

éste no mejora luego de una práctica y apoyo adicionales, no debería participar en el trabajo de campo.

Llegada al centro educativo

Antes de partir a los centros educativos, los examinadores y supervisores deberían:

- Verificar todos los materiales, incluyendo una copia del formulario laminado del instrumento del educando por examinador, y suficientes copias de este instrumento.
- Discuta los procedimientos de aplicación de la prueba y las estrategias con que hacer que los educandos se sientan cómodos, y haga un juego de rol con este ejercicio.
- Verifique que todos los aplicadores están cómodos usando un cronómetro o su propio reloj.

Al llegar al centro educativo, el supervisor debe presentar el equipo de examinadores al director. En la mayoría de los países se requerirá de una carta firmada del ministerio para poder efectuar el ejercicio; el supervisor debe presentarla (una copia de la cual habrá sido enviada por adelantado de ser posible; véase un ejemplo de esta carta en el Anexo D), explicando la finalidad y los objetivos de la evaluación, y agradeciéndole al director su participación en la evaluación de la lectura en los primeros grados. Se le debe recordar que ni los educandos ni los docentes serán identificados por su nombre en el proceso de recolección de datos.

De estar planeado, el director del centro educativo debe ser notificado del procedimiento para darle al colegio un comentario del desempeño global de los educandos. Por último, el supervisor debería preguntarle si se dispone de un aula, sala de profesores o un lugar tranquilo donde cada uno de los administradores pueda llevar a cabo las evaluaciones individuales. Los examinadores deberían pasar a cualquier espacio que se les indique y colocar dos sillas o carpetas, una para el educando y otra para él mismo.

Seleccionando educandos y llevando a cabo la evaluación

De contarse a nivel central con datos recientes y exactos sobre la matrícula escolar por centro educativo, grado y clase, se podrá usar una lista numérica aleatoria antes de llegar al centro elegido para así generar la muestra de educandos. Dado que esto resulta extremadamente improbable en casi todos los contextos de países de bajos ingresos, para elegir los alumnos del centro educativo se deberá más bien seguir el siguiente procedimiento.

Para evaluar a 20 educandos por grado, el supervisor deberá llevar a cabo el siguiente procedimiento para cada grado (1, 2 y 3, uno a la vez).

1. De ser posible consiga la lista de alumnos en cada salón o del director. Es una buena práctica comenzar con la selección de los niños de primer grado. A menudo sucede que cuanto más avanzado está el día escolar, tanto menos pueden estos niños pequeños interesarse y concentrarse.
2. Cuente el número total de alumnos matriculados en cada salón de dicho grado y súmelos (v.g., Salón A = 40, Salón B = 30, Salón C = 50, Total = 120).
3. Divida el total (120) por el número de alumnos a entrevistar (en este ejemplo se seleccionarán 20 de cada grado $[20 \times 3]$, de modo que la respuesta es 6).

4. Use esta respuesta para contar en las listas de salón y seleccione cada “equis-ésimo” alumno para que forme parte de la muestra. En este ejemplo la respuesta es 6, de modo que los alumnos 6, 12, 18, 24, 30 y 36 de la lista tomarán parte en el Salón A; los alumnos 2, 8, 14, 20 y 26 en el Salón B; y los alumnos 2, 8, 14, 20, 26, 32, 38, 44 y 50 en el Salón C.
5. Si un alumno está ausente o rehúsa a participar, seleccione el siguiente número en la lista del salón. De estar ausente dicho alumno o rehusarse, escoja al siguiente. Esto dará la muestra de 20 alumnos distribuidos entre los salones del 1° grado. Regístrese en el informe del centro educativo el número de negativas y de ausencias.
6. Saque a los alumnos de sus salones en grupos pequeños, 1 por examinador, para así minimizar problemas organizativos de las clases. Llévelos a los examinadores y preséntelos por su nombre. Tome nota del tamaño de la clase del alumno e información sobre su edad y/o fecha de nacimiento, y comuníquela a cada uno de los examinadores al inicio de la entrevista del educando.
7. Una vez que los aplicadores hayan completado la evaluación de todos los alumnos de 1° grado, repita el mismo procedimiento para 2° y 3° grado.
8. Asegúrese de que los aplicadores siempre tengan un alumno que evaluar, para así no perder mucho tiempo durante la aplicación. En la medida de lo posible, todas las entrevistas deben completarse dentro del día escolar. Si la escuela sólo tiene un turno y la evaluación no ha sido completada antes del final del mismo, busque a los alumnos restantes y pídale que esperen después del final del día escolar. En este caso el director de la escuela o los docentes deben tomar medidas para notificar a los padres que algunos niños regresarán tarde a casa. Este punto deberá discutirse por adelantado con los examinadores y supervisores, en lo que respecta a la práctica más apropiada dadas las condiciones locales.

Enseñando lecciones para el trabajo de campo

A lo largo de la formación, los participantes deberían reflexionar y compartir experiencias sobre la aplicación piloto del instrumento. Las instrucciones deberían mejorarse y aclararse en base a la experiencia de los examinadores en los centros educativos. El trabajo de campo concreto debería tener lugar inmediatamente después de la formación. De ser posible, cada equipo debería tener un automóvil para transportar los materiales y llegar a los colegios muestreados antes del inicio del día escolar. Hasta ahora la experiencia ha mostrado que la aplicación de EGRA requiere de unos 15 a 20 minutos por niño. Esto significa que un equipo de tres examinadores y un supervisor puede completar unas nueve o diez aplicaciones por hora, esto es unos 30 niños en tres horas ininterrumpidas.

En base al trabajo llevado a cabo por su socio local, el Centro de Investigación y Acción Educativa Social (CIAS) en Nicaragua, RTI diseñó un manual del supervisor para que acompañe al instrumento y guíe el proceso de recolección de datos. Puede encontrarse el manual, junto con las instrucciones más recientes, en el sitio web de EdData (www.eddataglobal.org), bajo EGRA > Current EGRA Instruments.

VI. Analizando los datos de EGRA

Esta sección cubre algunos enfoques fundamentales y de bajo contenido tecnológico, que el equipo de ingreso y análisis de los datos puede usar al trabajar con la información de EGRA.

Durante todo el desarrollo de EGRA, el enfoque de RTI fue trabajar con herramientas y enfoques de bajo costo y que son ampliamente accesibles. Las contrapartes de EGRA han preferido trabajar con Excel puesto que los paquetes estadísticos son bastante costosos y requieren de una formación especial o de una formación autodidacta. Con esto en mente, la presente sección del manual fue diseñada usando el “Data Analysis ToolPak” de Excel (un complemento disponible para la mayoría de los usuarios de Excel) y las *Pivot Tables* (véase el examen debajo). Cuando se cuente con paquetes estadísticos comunes como SPSS y Stata, y no se requiera de mucha formación adicional para su uso, se deberían emplear dada su mayor capacidad para el análisis de los datos. Esta sección se ocupa de los siguientes temas:

- Limpiando e ingresando datos
- Usando Excel para analizar los datos

En el Anexo B se incluye un examen completo del tamaño de la muestra.

Limpiando e ingresando datos

La experiencia con EGRA y otras encuestas sugieren que los siguientes apartados son importantes para asegurar un buen ingreso de los datos (libre de errores) y su limpieza.

1. Asegúrese de que en cada hoja del cuestionario o formulario de evaluación se incluya al menos una pieza clave de datos de identificación por cada niño, para que los cuestionarios o formularios queden separados durante el transporte.
2. Asegúrese de que al final de cada día se revise que todos los formularios de datos estén completos y sean consistentes, siendo deseable que lo hiciera otra persona que no sea la que llevó a cabo la evaluación. Esto implica que haya una proporción razonable entre supervisores y evaluadores, de modo que los primeros puedan revisar todos los formularios al finalizar el día. De modo alternativo, los evaluadores pueden revisar su trabajo mutuamente al terminar el día.
3. Dado el tamaño relativamente pequeño de las muestras usadas en las evaluaciones de EGRA, el ingreso de los datos en muchos casos se ha llevado a cabo en Excel. Esto maximiza la transparencia y la facilidad para compartirlos. Puede usarse este programa para crear funciones simples y comparaciones que permiten tener consistencia interna y efectuar verificaciones del rango de los datos, para así detectar y prevenir errores en su ingreso. Naturalmente que es posible usar otros métodos de ingreso de datos, pero para el tamaño de las muestras consideradas en la mayoría de las evaluaciones de EGRA basta con Excel. Se ha desarrollado un sistema de interfase de ingreso de datos basado en Microsoft Access, al cual se viene probando en varios países. Este sistema estandarizado de ingreso de datos reduce enormemente los errores en su ingreso y puede programársele para que genere informes simples.

4. Pero como ya se indicó, es probable que una vez que se ingresen datos en Excel, uno desee transferirlos a un paquete estadístico como SPSS o Stata para el análisis mismo. Es por ello importante ingresarlos con gran cuidado para proteger así la integridad del registro. En otras palabras, es importante asegurarse de que los datos de un niño dado sean ingresados cuidadosamente en una sola hilera, y tener gran cuidado de no mezclarlas. Las hileras mixtas o múltiples son particularmente peligrosas si los datos se ordenan electrónicamente. Excel tiene una muy débil capacidad de integridad del registro, de modo que el personal de ingreso y análisis de los datos debe tener mucho cuidado al manipularlos.
5. Como se anotó en la Sección IV, al codificar los datos es de extrema importancia que el personal que los ingrese registre las respuestas correctamente y cuente con una estrategia de codificación que distinga entre los siguientes tipos de respuesta: (a) pregunta no formulada, (b) el alumno no sabía, (c) el alumno no quería o no podía responder, y (d) un cero auténtico (respuesta del todo incorrecta).
6. De igual modo, si los datos van a ser compartidos con otros o ser importados a un paquete estadístico, es importante crear nombres de variables que sean completos y mnemotécnicamente útiles (v.g., PCPM para palabras leídas correctamente por minuto).
7. Para algunas variables claves, como la fluidez (el número de palabras leídas correctamente por minuto), los datos, tal como salen directamente de los formularios de evaluación, denotan el tiempo que una tarea dada tomó y el número de palabras o letras leídas. En otras palabras, aunque la tarea esté programada para que tome 1 minuto, unos cuantos niños podrían terminarla en menos tiempo, y en algunos casos podría permitirse que tome más de 1 minuto. Se recomienda por ello que ambos valores—esto es, el número de palabras o letras leídas y el tiempo en segundos—se ingresen en la base de datos, pero que se aplique la funcionalidad de Excel para crear la variable “por minuto”. Así, uno ingresa que un niño leyó correctamente 55 palabras en 40 segundos, y luego se usa una fórmula de Excel para calcular que esto significa que leyó 82.5 palabras por minuto, donde la fórmula sería palabras correctas) segundos x 60 (véase el examen descrito anteriormente).
8. El equipo de análisis de datos debe crear un libro de claves o diccionario de datos para describir el nombre de cada variable. Podría muy fácilmente colocársele en una página separada en la hoja de cálculo de Excel, y contendría información similar a la de la tabla 7:

Tabla 7. Ejemplos de entradas en el libro de claves

Nombre de la variable	Descripción de la variable	Notas de codificación
Pcleídas	Palabras leídas correctamente	En blanco quiere decir que la tarea no se continuó
Pcsegundos	Tiempo en segundos que toma leer las letras correctamente	En blanco quiere decir que la tarea no se continuó
Pcpm	Fluidez en letras leídas correctamente por minuto (variable creada)	Cero significa que ninguna palabra fue leída, puesto que la tarea no se continuó

Nombre de la variable	Descripción de la variable	Notas de codificación
Pctvleído	Palabras leídas correctamente en el párrafo	En blanco quiere decir que la tarea no se continuó
Pctvsegundos	Tiempo en segundos que toma leer el párrafo	En blanco quiere decir que la tarea no se continuó
Pcpmtv	Palabras correctas por minuto en el párrafo (variable creada)	Cero significa que ninguna palabra fue leída, puesto que la tarea no se continuó

9. Al ingresar los datos y crear las variables, el equipo de análisis de los datos debe crear dos variables para cada uno de los conceptos importantes, como palabras leídas correctamente por minuto en el párrafo, o letras leídas correctamente por minuto. Por ejemplo, para calcular promedios podría resultar útil crear una que incluya sólo a los niños que intentaron efectuar la tarea y leyeron lo suficiente como para que se les permitiera continuarla, y otra que también incluya a aquellos que simplemente no leyeron lo suficiente como para registrarlo. De incluirse sólo a los primeros, se crearía una imagen algo distorsionada del centro educativo puesto que se exagera qué tan bien leen los niños. Pero de incluirse en el promedio a los niños que no pueden leer en absoluto, no se da una buena idea de la fluidez de lectura de los que sí pueden leer.

No existe una solución simple a este problema, en particular porque la línea que separa “puede leer” de “no puede leer” es en realidad algo arbitraria, lo que podemos ver en el hecho que la fluidez calculada típicamente varía bastante y mostrará muchos casos cercanos a cero. Por ello, una buena solución es permitir ambos tipos de cálculos: un promedio que incluya a los niños que no pueden leer en absoluto como de fluidez 0, y otro promedio que excluya a los que se juzga no leen, incluyendo así sólo a aquellos con una fluidez mayor que 0. Esto resulta simple de usarse un programa como Stata o SPSS, pues hay comandos simples para excluir los casos con un valor de 0 en la variable cuyo promedio se está calculando.

Análisis de datos: Usando Excel para analizar los datos

La mayoría de los análisis necesarios para producir un informe básico de los resultados de EGRA pueden efectuarse con Excel. Esta sección sugiere cómo debe hacerse la mayoría de los cálculos. Debe además usarse pesos para una ponderación correcta de los resultados (para información adicional véase el Anexo B).

Las siguientes sugerencias deberían permitir efectuar todos los análisis básicos necesarios.

Para proteger la integridad de los datos se sugiere especialmente que la mayor parte posible del análisis se efectúe usando la utilidad “*Pivot Table*” de Excel. Esto permite que el equipo de análisis de datos calcule promedios para las variables claves según cualquier subconjunto necesario de datos, como por edad, por grado, por género o por centro educativo. Uno puede calcular fácilmente los resultados usuales para cualquier variable y todo subgrupo, como la media, la cuenta o el número de casos, y la desviación estándar.

Los resultados típicamente de interés incluirán el promedio de palabras (o letras) correctas por minuto, desagregados por edad o grado, o algún otro factor. Un ejemplo típico generado usando Excel sería como sigue:

Variable	Grado		
	1	2	3
Letras leídas correctamente por minuto	12.8	25.4	36.1
Letras leídas correctamente por minuto, excluyendo a los no lectores	19.8	28.6	37.7
Palabras familiares leídas correctamente por minuto	1.2	2.3	4.3
Palabras familiares leídas correctamente por minuto, excluyendo a los no lectores	6.2	8.3	9.2
Palabras leídas correctamente por minuto, el párrafo	2.2	4.0	9.2
Palabras leídas correctamente por minuto, el párrafo, excluyendo a los no lectores	11.0	11.6	17.3

Podría ser de interés ver si hay diferencias en los resultados por grado, por edad o por género. Por ejemplo, en este caso parece haber cierta progresión por grado en la fluidez de nombrar las letras: una ganancia de unas 10 letras por grado. ¿Pero cuán seguro podemos estar de que las diferencias por grado, edad o género sean significativas?

Podemos derivar una idea razonablemente rigurosa de cuán significativas resultan las diferencias por grado (o edad, género, público-privado, urbano-rural, o cualquier otro atributo) pidiéndole a Excel—vía el comando *Pivot Table*—las medias, desviaciones estándar y conteos por grado, calculando luego un simple intervalo de confianza. Un ejemplo simple para el caso de la variable de letras leídas correctamente por minuto (ver arriba), es como sigue:

	Grado		
	1	2	3
Conteo	419	389	392
Promedio letras leídas correctamente por minuto	12.8	25.4	36.1
Desviación estándar	18.2	21.9	23.2
Error estándar	0.89	1.11	1.17

Sin embargo, un cuadro generado de este modo aún no da los intervalos de confianza. El cuadro nos dice que hay una variabilidad sustancial *dentro* de los grados: algunos niños se desempeñan mucho mejor que otros. Dado que la mayoría de los niños no están en el promedio, claro está, el “niño promedio” se hallará a cierta distancia, arriba o abajo del promedio. La desviación estándar nos dice que en general, el desempeño de un niño promedio puede fluctuar hasta 20 o más palabras por encima o por debajo del promedio del desempeño de los niños. En otras palabras, la desviación estándar es una medida de la diferencia promedio que existe entre cada individuo y el promedio.

La variabilidad resulta interesante por sí misma puesto que nos dice algo de la desigualdad o el desnivel del desempeño. Pero también se puede usar para dar una idea de cuán confiables estadísticamente son los promedios. La intuición de por qué la desviación estándar podría darnos una guía en lo que toca a cuán confiables son los promedios—o cuán confiables son las diferencias entre ellos—tiene como base la noción de que si hay bastante variación dentro de los grados, entonces tal vez las diferencias observadas entre ellos sean accidentales. Dado que todo con lo que contamos es una muestra y no la población real, podríamos estar

cometiendo un margen de error amplio en lo que toca a las letras leídas correctamente por minuto que los niños de la población pueden leer. Toda muestra dada podría presentar un margen de error del lado alto o bajo, y cuanto mayor sea la variabilidad, tanto mayor será la probabilidad de que una muestra presente el margen de error ya sea por lo alto o por lo bajo. ¿Pero cuán probable es que uno haya cometido ese margen de error tanto como para llegar a la conclusión errónea de que hay cierta progresión por grado cuando realmente no la hay?

Podemos responder a esta pregunta con cierto rigor construyendo algo a lo que se conoce como “intervalos de confianza”. Hacer esto toma dos pasos usando comandos o fórmulas muy simples de Excel. Primero calcule el “error estándar”, que es la desviación estándar dividida por la raíz cuadrada del conteo. Por ejemplo, el error estándar de las letras leídas correctamente por minuto en 1º grado es 0.89, o $18.2 / \sqrt{419}$. En segundo lugar, y aplicando un principio general simplificado, sume dos veces el error estándar a la media para tener un “límite superior” del intervalo de confianza, y reste dos veces el error estándar a la media para así obtener su “límite inferior”. Esto puede tabularse también con Excel, como sigue.

	Grado		
	1	2	3
Promedio de letras leídas correctamente por minuto	12.8	25.4	36.1
Límite inferior del intervalo de confianza	11.0	23.1	33.7
Límite superior del intervalo de confianza	14.6	27.6	38.4

Por ejemplo, en esta tabla 14.6 es sólo $12.8 + 2 * 0.89$, o el promedio más el doble del error estándar. Esta tabla debe interpretarse como sigue: “El promedio de letras leídas correctamente por minuto en 3º grado es de 36.1 en nuestra *muestra*, y podemos estar 95 por ciento seguros de que el promedio subyacente en la *población* cae en algún lugar entre 33.7 y 38.4”. Si el límite superior de un grado, digamos 2º, resulta inferior al límite inferior del siguiente grado, digamos 33.7, entonces podemos estar plenamente seguros de que hay una progresión real por grado.¹⁸

Tamaño de la muestra

En este punto se puede discutir el tamaño de la muestra. Podría parecer que éste ya se discutió antes con mayor detenimiento. Sin embargo, resulta más fácil entender las referencias al tamaño después de haber visto los tipos de intervalos de confianza que podemos generar usando ejemplos reales de tamaños de muestras. En la tabla anterior vimos que los intervalos de confianza de, por ejemplo, las letras leídas correctamente por minuto, tendían a ser más o menos 1.8 alrededor del promedio si el tamaño de la muestra era de unos 400 niños por grado, y que esto basta para detectar la progresión por grado. Esto depende, claro está, de la desviación estándar. En base a las experiencias de EGRA hemos comenzado a descubrir que la desviación estándar es razonablemente constante entre un país y otro.

En base al análisis de datos de Perú, Jamaica, Kenia, Gambia y otros, los resultados fundamentales de otros conceptos, como el de palabras leídas correctamente por minuto en un párrafo, no diferían mucho entre sí, mostrando más bien desviaciones estándar de alrededor de 15 a 20 en el lado alto. La siguiente tabla muestra los resultados para el párrafo de Kenia.

¹⁸ Consúltense el Anexo B para un examen del acto de crear un equilibrio requerido en las decisiones referidas al enfoque del muestreo, el tamaño de la muestra y el nivel deseado de precisión.

	Kiswahili	Inglés
Conteo	400	400
Media de palabras leídas correctamente por minuto, texto vinculado	8.7	9.3
Desviación estándar	13.1	14.2
Límite inferior del intervalo de confianza	7.4	7.9
Límite superior del intervalo de confianza	10.0	10.7

En base a estos tipos de resultados, parece adecuado el tamaño de muestras de aproximadamente 400 niños por grupo (género, grado, lengua, público-privado, urbano-rural), cuyos valores vale la pena considerar desagregar. Naturalmente que se necesitan unos 400 niños para cualquier *combinación* (varón-urbano requeriría de unos 400, al igual que mujer-urbano, varón-rural y mujer-rural; distinguir por género y localidad requeriría así que el tamaño de la muestra sea de unos 1600, en tanto que una línea de base simple por grado sólo necesitaría de 400 niños por grado). Para un examen detallado del tamaño de la muestra y los procedimientos de asignar pesos, consúltese por favor el Anexo B.

VII. Usando EGRA: implicaciones para el diálogo de políticas

En las aulas del mundo hay una desconexión entre las prácticas de instrucción y el desempeño de los alumnos. Los docentes a menudo escriben largos párrafos en la pizarra para que los educandos los copien; en estas mismas aulas, muchos alumnos no pueden reconocer todas las letras del alfabeto. Muchos de los avances en la enseñanza de la lectura, que numerosas investigaciones demostraron son efectivos, no están siendo usados por los docentes. En efecto, en muchos países es muy poca la enseñanza sistemática de la lectura que se viene llevando a cabo. En muchos casos los cinco componentes de la enseñanza efectiva de la lectura, esbozados por el National Reading Panel (National Institute of Child Health and Human Development, 2000) (conciencia fonémica, método fonológico, fluidez, vocabulario y comprensión), brillan por su ausencia. Si bien hay una miríada de explicaciones de esta desconexión, al menos parte del problema radica en la brecha existente entre la comunidad investigadora y lo que se está enseñando en los programas de formación docente en países de bajos ingresos.¹⁹

Como ya se viera en la Sección II, uno de los objetivos primarios de EGRA es diagnosticar las áreas a mejorar, a nivel del sistema, en la enseñanza de la lectura temprana. La siguiente sección brinda un breve cuadro general de las cuestiones de política e instrucción que los ministerios y donantes deben enfrentar al usar los resultados de EGRA con este fin.

Usando resultados para informar el diálogo de políticas

La finalidad máxima de EGRA es informar los cambios en la enseñanza. Hasta ahora, en nuestra experiencia, el impacto que el diálogo de políticas tiene para informar la enseñanza parece tener dos pasos distintos.

Influyendo en la toma de decisiones de políticos y funcionarios

En primer lugar, los resultados tienden a despertar interés en políticos y funcionarios. Una de las virtudes de EGRA es que la ciencia que le sustenta corresponde bastante bien al concepto que el ciudadano promedio tiene de qué significa leer: la noción de “saber las letras”, de ser capaz de leer sin vacilaciones y a un ritmo razonable, y de ser capaz de responder unas cuantas preguntas sobre lo que uno ha leído, son lo que la mayoría de los ciudadanos intuitivamente considera es la lectura. Por ende, ser capaz de informar que los niños no reconocen las letras, o que sólo pueden leerlas en forma extremadamente lenta, es algo que la mayoría de la ciudadanía puede entender. El uso de grabaciones de audio o video, que dramatizan las diferencias entre un lector de muy bajo rendimiento (un niño que lee digamos 10–15 palabras por minuto, sin comprensión alguna) y un buen lector (uno que lea digamos 60 palabras por minuto, comprendiéndolas) resulta instantáneamente obvio y conmovedor. (Para un ejemplo de un video tal, diseñado por DFID y el Banco Mundial en Perú, véase www.eddataglobal.org, página principal).

¹⁹ Actualmente, en el programa Educational Quality Improvement Program (EQUIP1) financiado por USAID, la International Reading Association y el American Institute for Research vienen revisando una muestra de currículos nacionales y materiales de enseñanza, para así evaluar el énfasis (o no) en la enseñanza de la lectura.

Es más, los ciudadanos y funcionarios, en particular aquellos que aplican la prueba EGRA por sí mismos (o simplemente le piden a niños que les lean), rápidamente se dan cuenta de que los niños no están leyendo y comunican esto a otros funcionarios. Los funcionarios de diversos países parecen estar tomando conciencia de un problema serio de lectura entre los niños de

sus centros educativos. EGRA ayudó a inducir esto en algunos casos, pero en otros se trata en realidad de una respuesta a los tipos de preocupación que los funcionarios ya estaban manifestando.

En algunos contextos, las reacciones a una evaluación de la lectura tipo EGRA no son tan directas. Algunos comentaristas en algunos países cuestionan la utilidad de la fluidez de la lectura oral como marcador o indicador predictor del aprendizaje en general, o incluso de la lectura. Por esta razón es importante tener acceso a bibliografía de referencia que explique los problemas, parte de la cual es citada en este manual

“Para prevenir los problemas de lectura, se debe dar a los niños:

- Oportunidades para que exploren los diversos usos y funciones del lenguaje escrito, y desarrollar un aprecio y comando de los mismos.
- Oportunidades para que capten y dominen el uso del principio alfabético para leer y escribir.
- Oportunidades para que desarrollen y mejoren las habilidades de lenguaje y meta-cognitivas, para satisfacer lo que demanda entender textos impresos.
- Oportunidades para que experimenten contextos que promueven el entusiasmo y el éxito al aprender a leer y escribir, así como el aprendizaje *a*/leer y escribir.
- Oportunidades para que se identifique a los niños *que es probable tengan* problemas para ser lectores fluidos y participen así en programas de prevención eficaces.
- Oportunidades para que los niños *que tienen* dificultades para convertirse en lectores fluidos sean identificados y participen en programas efectivos de intervención y recuperación, bien integrados con la enseñanza en curso en el aula”.

Snow, C.E. et al. (1998), *Preventing reading difficulties in young children* (pág. 278)

de herramientas. Pueden encontrarse otras referencias útiles en www.reading.org, www.nationalreadingpanel.org, y <http://dibels.uoregon.edu/>.

En otros casos parecería existir una sensación de que los esfuerzos de EGRA intentan transmitir la noción de que “la lectura es todo lo que importa”. En estos casos es importante señalar que la lectura efectivamente es una habilidad fundacional importante que influye en el éxito académico a lo largo del currículo escolar, y que ella es un buen marcador de la calidad global de un centro educativo, pero que EGRA no se basa en el supuesto de que la lectura sea todo lo que importa.

En general, todo intento de medir la calidad, representada por el aprendizaje, queda sujeto a este tipo de debates muy conocidos. En la experiencia que se va acumulando con la aplicación de EGRA o de herramientas semejantes, parecería que los docentes, las personas a quienes les interesa prestar un apoyo directo a los maestros, y los funcionarios de alto nivel tienden todos a ver las virtudes de EGRA, en tanto que algunos teóricos del currículo o de la lectura parecieran tener algunas inquietudes o preocupaciones por temor a posibles simplificaciones excesivas. Resulta crucial entender que el uso práctico de EGRA y las estrategias de mejora derivadas deberían verse sólo como un punto de entrada, y como un ejemplo de qué puede conseguirse concentrándose y monitoreando resultados específicos. La lección básica puede entonces aplicarse a otros aspectos de la enseñanza y el aprendizaje.

Al concentrar la atención en las decisiones de políticos y funcionarios en este tema, es útil poder fijar de algún modo un punto de referencia o criterio sobre los resultados. Dos formas de hacer esto resultaron útiles en los ejercicios de EGRA en cualquier país dado: una comparación con estándares o metas internacionales de algún tipo, y algún análisis del desempeño de los centros educativos en el país en cuestión. La tabla 8 que figura en la parte inferior muestra

resultados promedio reales en un reciente ejercicio de EGRA, un posible punto de referencia internacional y un posible punto de referencia nacional.

Los datos en la última columna, esto es los puntos de referencia internacionales, se tomaron de los Dynamic Indicators of Basic Early Literacy Skills.²⁰ En el Anexo A pueden encontrarse indicadores adicionales de textos vinculados para estudiantes de EE.UU.

Tabla 8. Un ejemplo de referencias entre países con fines comparativos

	Promedio entre niños en un país dado, mediados de año, <u>2º grado</u>	Promedio máximo a <u>nivel escolar</u> en el mismo país, 2º grado	Referencias de países desarrollados con fines comparativos
Letras correctas por minuto	22.7	41	40 para finales de kindergarten*
Palabras sin sentido correctas por minuto	7.5	25	50 a mediados de 1º grado
Palabras correctas por minuto	11.4	36	20 a mediados de 1º grado
Puntaje de comprensión	0.4	2	<i>No hay datos</i>

* Esta habilidad ni siquiera es seguida o tiene una referencia más allá de kindergarten, asumiéndose que será dominada en dicho año de estudios. Nótese en el caso presentado en la tabla, el bajo nivel del promedio para 2º grado.

La variable más importante para el cual hay que fijar un criterio—puesto que su poder predictor con respecto a las restantes habilidades es el más alto—es la fluidez en leer los párrafos (palabras correctas por minuto). Por ello la mayoría de los investigadores coincide en la idea de que para finales del 1º o 2º grado, los niños que están aprendiendo a leer en español deben estar leyendo aproximadamente 60 palabras correctamente por minuto, respectivamente. En base a nuestra experiencia hasta la fecha en aproximadamente 10 países, sería tal vez razonable suavizar estos criterios aproximadamente a unas 45 palabras leídas correctamente por minuto en los países pobres con complejidad lingüística u ortografías particularmente difíciles. A partir de estos resultados los investigadores podrían entonces especificar la proporción de niños que alcanzan la exactitud y la fluidez de lectura del nivel fundacional, además del grado promedio al cual los niños “dan el salto” al conocimiento de la lectura. El grado en que se “da el salto” podría entonces ser aquel en el cual el 90 por ciento de los niños están satisfaciendo algún estándar o criterio.²¹

Es muy importante tener en cuenta que un país puede fijar sus propios puntos de referencia examinando el rendimiento de aquellos centros educativos que se sabe tienen un buen desempeño, o que puede mostrarse que lo tienen en una evaluación de tipo EGRA, pero que no poseen en cambio ninguna ventaja socioeconómica particular, o un nivel insostenible de uso de recursos. Tales colegios típicamente arrojarán referencias que serán razonablemente

²⁰ Puede encontrarse los DIBELS en <http://dibels.uoregon.edu/benchmark.php>.

²¹ Si los grados en los cuales se prueba el instrumento—en particular en los primeros casos en que se les puso a prueba—no incluyen un grado en el cual por ejemplo el 90 por ciento de los niños esté dando el salto a la capacidad de lectura, podríamos utilizar técnicas de regresión y extrapolación para estimar este grado. La implementación subsiguiente tendría entonces que asegurarse de incluir dicho grado.

exigentes pero que se puede demostrar que son alcanzables por niños sin mayores ventajas socioeconómicas o centros educativos sin una gran ventaja en los recursos, siempre y cuando se cuente con una buena instrucción.

Las preguntas de comprensión de lectura asimismo indican si los niños están leyendo “entendiendo”. Aquí también debe especificarse un criterio que tenga como base resultados empíricos y el examen de la distribución (v.g., 75 por ciento de las preguntas respondidas correctamente) en cada país. Puede entonces especificarse la proporción de niños que alcanzan el conocimiento fundacional de la lectura con fluidez y comprensión. La evaluación puede brindar información analítica y diagnóstica sobre el progreso en la lectura en centros educativos/salones particulares (con la técnica de muestreo adecuada), lo cual puede entonces ser informado a los docentes a nivel de los salones de clase. Por ejemplo, el progreso podría quedar limitado por una comprensión deficiente, falta de fluidez, la ignorancia del sonido de las letras, la incapacidad para descodificar, un vocabulario restringido, o una combinación de dichos rasgos. Esta información puede entonces ser transmitida de vuelta a los docentes, directores de centros educativos y supervisores escolares, como un indicio de a dónde podría desplazarse el peso de la enseñanza para así mejorar los resultados del aprendizaje.

Cambiando la enseñanza de la lectura

En segundo lugar—al menos cuando las cosas salen bien—la preocupación parece estar dando a lugar a una toma de conciencia de la necesidad de cambiar la enseñanza en los primeros grados. Hasta ahora, dos países en los cuales se probó EGRA—Gambia y África del Sur—diseñaron espontáneamente materiales bastante extensos para capacitar a los docentes en una mejor enseñanza de la lectura. En Kenia, las actividades de los donantes y la colaboración con la Aga Khan Foundation llevaron a la preparación de un grupo de planes de lecciones con que mejorar la enseñanza de la lectura en los primeros grados. El Anexo E incluye el esbozo de una semana de colaboración entre un experto en lectura, la Aga Khan Foundation y los funcionarios del gobierno local para diseñar planes de lecciones.

Pueden encontrarse en varias fuentes ejemplos de planes de lecciones y estrategias con que enseñar los aspectos fundacionales de la lectura temprana—a saber la conciencia fonémica, el método fonológico, la fluidez, el vocabulario y la comprensión. Entre algunos ejemplos tenemos los manuales del docente, como los de Linan-Thompson y Vaughn para la enseñanza del inglés nativo (2004) y estudiantes de la lengua inglesa (2007), y también en español el manual del docente y del alumno para la prevención de las dificultades específicas de aprendizaje de la lectura Programa de Prevención de las Dificultades Específicas de Aprendizaje, o PREDEA (Jiménez, Artiles, et al., 2009). Estos materiales pueden usarse entonces para dar forma a planes de lecciones diseñados localmente (véase un ejemplo en el Anexo F). Los países pueden entonces diseñar sus propios planes de lecciones sugeridos, usando los resultados de EGRA para identificar aquellas áreas que necesitan mejoras. Pueden encontrarse ejemplos de planes de lecciones (en inglés) de Kenia, Gambia y África del Sur en www.eddataglobal.org, Documents and Data.

Usando los datos para informar de los resultados a los centros educativos

Hasta ahora, las aplicaciones de EGRA han sido usadas fundamentalmente para generar una discusión a nivel nacional, y para estimular a los ministerios a que entren en acción. Esta promoción de una toma de conciencia queda complementada con el informe de los resultados a docentes y centros educativos. Para reiterar un enunciado de una sección anterior: al hacer

un informe a los centros educativos, no debe en ningún caso identificarse a alumnos o docentes individuales, puesto que la medida no busca ser usada como una herramienta de rendición de cuentas de alto impacto. Pero esto al margen, cierta forma de retribuir a los colegios como un medio de agradecerles su participación resulta siempre bienvenida.

Para informar de los resultados a los centros educativos, los analistas deberían crear un resumen simple de los resultados de una página de largo, que incluya un informe por grado y género para cada colegio individual. Pueden también compartirse los resultados promedio de centros educativos de características y medias similares en toda la muestra. Estos informes deben ir acompañados de explicaciones que indiquen cómo es que cada subprueba se relaciona con la enseñanza, y qué pueden hacer los docentes para mejorar los resultados. También deben compartirse ejemplos de planes de lecciones y actividades sugeridas con los centros educativos.

Referencias

- Abadzi, H. (2006). *Efficient learning for the poor*. Washington, DC: Banco Mundial.
- Abadzi, H., Crouch, L., Echegaray, M., Pasco, C. y Sampe, J. (2005). Monitoring basic skills acquisition through rapid learning assessments: A case study from Peru. *Prospects*, 35(2), 137–156.
- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, Massachusetts: MIT Press.
- Adams, M. J., y Collins, A. (1979). A schema-theoretic view of reading. En R. O. Freedle (Ed.), *New directions in discourse processing. Vol. 2: Advances in discourse processes*. Norwood, New Jersey: Ablex Publications.
- Adrián, J. A., Alegría, J., y Morais, J. (1995). Metaphonological abilities of Spanish illiterate adults. *International Journal of Psychology*, 3, 329–353.
- Alameda, J. R., y Cuetos, F. (1995). *Diccionario de frecuencia de las unidades lingüísticas del castellano* (Vols. I y II). Oviedo: Servicio de Publicaciones, Universidad de Oviedo.
- Alegría, J., Carrillo, M., y Sánchez, E. (2005). La enseñanza de la lectura. *Investigación y Ciencia*, 340, 6–15.
- Alegría, J., y Mousty, P. (1996). The development of spelling procedures in French-speaking, normal and reading-disabled children: Effects of frequency and lexicality. *Journal of Experimental Child Psychology*, 63(2), 312–338.
- Ball, E. (1993). Phonological awareness: What's important and to whom? *Reading and Writing: An Interdisciplinary Journal*, 5, 141–159.
- Banco Mundial. (2007). *Por una educación de calidad para el Perú: Estándares, rendición de cuentas, y fortalecimiento de capacidades*. Washington, DC: Autor.
- Banco Mundial: Independent Evaluation Group. (2006). *From schooling access to learning outcomes—An unfinished agenda: An evaluation of World Bank support to primary education*. Washington, DC: Autor.
- Barr, R. (1974). The effect of instruction on pupils' reading strategies. *Reading Research Quarterly*, 10, 555–582.
- Bertelson, P., De Gelder, B., Tfouni, L.V., y Morais, J. (1989). Metaphonological abilities of adult illiterates: New evidence of heterogeneity. *European Journal of Cognitive Psychology*, 1, 239–250.

- Bradley, L., y Bryant, P. (1985). *Children's reading problems*. Oxford/Nueva York: Basil Blackwell.
- Bravo-Valdivieso, L., Villalón, M., y Orellana, E. (2001). Procesos predictivos del aprendizaje inicial de la lectura en primero básico. *Boletín de Investigación Educativa*, 16, 149–160. *Psyche*, 11, 175–182.
- Bravo-Valdivieso, L., Villalón, M., y Orellana, E. (2002). Diferencias de preparación para el aprendizaje de la lectura en niños que ingresan al primer año básico. *Boletín de Investigación Educativa*, 16, 156–171.
- Bravo-Valdivieso, L., Villalón, M., y Orellana, E. (2006). Predictibilidad del rendimiento en la lectura: una investigación de seguimiento entre primer y tercer año. *Revista Latinoamericana de Psicología*, 38, 1–20.
- Bruck, M., Genesee, F., y Caravolas, M. (1997). A cross linguistic study of early literacy acquisition. En B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (págs. 145–162). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Bruck, M., y Waters, G. (1990). An analysis of component spelling and reading skills of good readers-good spellers, good readers-poor spellers and poor readers-poor spellers. En T. H. Carr, y B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (págs. 161–206). San Diego, California: Academic Press/Harcourt Brace Jovanovich.
- Byrne, B., y Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology*, 83, 451–455.
- Carrillo, M. (1994). Development of phonological awareness and reading acquisition. A study in Spanish language. *Reading and Writing: An Interdisciplinary Journal*, 6, 279–298.
- Carrillo, M. S., y Marín, J. (1996). *Desarrollo metafonológico y adquisición de la lectura: Un estudio de entrenamiento*. Madrid: Centro de Investigación y Documentación Educativa, Ministerio de Educación y Ciencia.
- Carver, R. P. (1994). Percentage of unknown vocabulary words in the text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior*, 26, 413–437.
- Carver, R. P. (1998). Predicting reading level in grades 1 to 6 from listening level and decoding level: Testing theory relevant to the simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 10, 121–154.

- Catts, H. W., Hogan, T. P., y Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities*, 36, 151–164.
- Center for Global Development. (2006). *When will we ever learn? Improving lives through impact evaluation*. Consultado en enero de 2007, de www.cgdev.org/files/7973_file_WillWeEverLearn.pdf
- Chabbott, C. (2006). Accelerating early grades reading in high priority EFA countries: A desk review. Consultado en septiembre de 2009, en <http://www.equip123.net/docs/E1-EGRinEFACountriesDeskStudy.pdf>
- Chall, J. (1987). Two vocabularies for reading: Recognition and meaning. En M. G. McKeown y M. E. Curtis (Eds), *The nature of vocabulary acquisition* (págs. 7–17). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Chiappe, P. (2006). *Rapid assessment of beginning reading proficiency: A proposed research design*. Manuscrito inédito. Irvine, California: Department of Education, University of California–Irvine.
- Chiappe, P., Siegel, L., y Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading*, 6(4), 369–400.
- Children's Literacy Initiative. (2000). *Dictation task*. Consultado en marzo de 2007, en http://www.cliontheweb.org/pd_asmntsamp2.html
- Clay, M. M. (1993). *An observation survey of early literacy achievement*. Ortonville, Michigan: Cornucopia Books.
- Clemente, M., y Domínguez, A. (1999). *La enseñanza de la lectura. Enfoque psicolingüístico y sociocultural*. Madrid: Pirámide.
- Crouch, L. (2006). *La fluidez lectora como indicador operacional*. Manuscrito inédito. Washington, DC: RTI International.
- Crouch, L., y Winkler, D. (2007). *Governance, management and financing of Education for All: Basic frameworks and case studies*. Manuscrito inédito. Washington, DC: RTI International.
- Cutting, L. E., y Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10(3), 277–299.

- Davis, C. J., y Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37, 665–671.
- De Gelder, B., Vroomen, J., y Bertelson, P. (1993). The effects of alphabetic-reading competence on language representation in bilingual Chinese subjects. *Psychological Research*, 55, 315–321.
- Defior, S. (2002). Phonological awareness and learning to read: A crosslinguistic perspective. En P. Bryant y T. Nunes (Eds.), *Handbook of children's literacy*. Amsterdam: Kluwer.
- Defior, S. y Herrera, L. (2003). Les habilités de traitement phonologique des enfants prélecteurs espagnols. En M. N. Rondhane, J. E. Gombert y M. Belajonza (Eds.), *L'apprentissage de la lecture. Perspective comparative interlangue* (págs. 161–176). Rennes: Presses Universitaires de Rennes.
- Defior, S., Jiménez-Fernández, G., y Serrano, F. (2007). *Fijó y dujó: ¿Cómo influye el valor léxico en la adquisición de la tilde?* Comunicación presentada al XXV Congreso Internacional de la Asociación Española de Lingüística Aplicada. Murcia, España: Universidad de Murcia.
- Denton, C. A., Ciancio, D. J., y Fletcher, J. M. (2006). Validity, reliability, and utility of the observation survey of early literacy achievement. *Reading Research Quarterly*, 41(1), 8–34.
- Domínguez, A. B. (1996a). Evaluación de los efectos a largo plazo de la enseñanza de habilidades de análisis fonológico en el aprendizaje de la lectura y de la escritura. *Infancia y Aprendizaje*, 76, 83–96.
- Domínguez, A. B. (1996b). Evaluación de los efectos a largo plazo de la enseñanza de análisis fonológico en el aprendizaje de la lectura y escritura. *Infancia y Aprendizaje*, 76, 83–96.
- Educational Testing Service. (2005). *Dictation assessment*. Consultado en marzo de 2007, de <http://www.pathwisestore.com/index.asp?PageAction=VIEWPROD&ProdID=161>
- Ehri, L. (1991). Learning to read and spell words. En L. Rieben y Ch. Perfetti (Eds.), *Learning to read: Basic research and its implications* (págs. 57–73). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Ehri, L. C. (1998). Word reading by sight and by analogy in beginning readers. En C. Hulme y R. M. Joshi (Eds.), *Reading and spelling: Development and disorders* (págs. 87–111). Mahwah, New Jersey: Lawrence Erlbaum Associates.

- Espin, C., y Foegen, A. (1996). Validity of three general outcome measures for predicting secondary students: Performance on content-area tasks. *Exceptional Children*, 62, 497–514.
- Espin, C., y Tindal, G. (1998). Curriculum-based measurement for secondary students. En M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement*. Nueva York: Guilford Press.
- Filmer, D., Hasan, A., y Pritchett, L. (2006). *A millennium learning goal: Measuring real progress in education*. Washington, DC: Banco Mundial.
- Foorman, B. R., Jenkins, L., y Francis, D. J. (1993). Links between segmenting, spelling and reading words in first and second grades. *Reading and Writing: An Interdisciplinary Journal*, 5, 1–15.
- Foorman, B. R., Francis, D. J., Novy, D. M., y Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. *Journal of Educational Psychology*, 83, 456–469.
- Fuchs, L., Fuchs, D., Hosp, M. K., y Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239–256.
- Genard, N., Alegría, J., Leybaert, J., Mousty, P. H., y Defior, S. (2005). La adquisición de la lectura y escritura: comparación translingüística. Comunicación presentada al II Congreso Hispano-Portugués de Psicología, Lisboa, 2004. *IberPsicología*, 10:3:17. Consultado en septiembre de 2009, en http://www.fedap.es/iberPsicologia/iberpsi10/congreso_lisboa/genard/genard.htm
- Gombert, J. E. (1992). *Metalinguistic development*. Great Britain, Harvester Wheatsheaf.
- Gombert, J. E. (1994). How do illiterate adults react to metalinguistic training? *Annals of Dyslexia*, 44, 250–269.
- González, M. J. (1996). Aprendizaje de la lectura y conocimiento fonológico: análisis evolutivo e implicaciones educativas. *Infancia y Aprendizaje*, 76, 97–107.
- Good, R. H., III, Simmons, D. C., y Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review*, 27(1), 45–56.
- Goswami, U., Gombert, J. E., y de Barreira, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French and Spanish. *Applied Psycholinguistics*, 19, 19–52.

- Goswami, U., Ziegler, J. C., Dalton, L., y Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? *Applied Psycholinguistics*, 24, 235–247.
- Gough, P. B. (1996). How children learn to read and why they fail. *Annals of Dyslexia*, 46, 3–20.
- Gough, P. B., y Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6–10.
- Guzmán, R., y Jiménez, J. E. (2001). Estudio normativo sobre parámetros psicolingüísticos en niños de 6 a 8 años: la familiaridad subjetiva. *Cognitiva*, 13, 153–191.
- Hasbrouck, J., y Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, 59(7), 636–644.
- Herrera, L., y Defior, S. (2005). Una aproximación al procesamiento fonológico de los niños prelectores: conciencia fonológica, memoria verbal a corto plazo y denominación. *Psyche*, 14, 81–95.
- Hoover, W. A., y Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127–160.
- International Literacy Institute y UNESCO. (2002). *Towards guidelines for the improvement of literacy assessment in developing countries: Conceptual dimensions based on the LAP [Literacy Assessment Practices] project*. Manuscrito inédito. Filadelfia, Pennsylvania: Autor. Consultado en septiembre de 2009 en <http://www.literacy.org/products/ili/pdf/LAPGuidelines.pdf>
- International Reading Association. (2007). *Teaching reading well: A synthesis of the International Reading Association's research on teacher preparation for reading instruction*. Consultado en enero de 2008, en www.reading.org/resources/issues/status.html
- Jiménez, J. E. (1992). Metaconocimiento fonológico: estudio descriptivo sobre una muestra de niños prelectores en edad preescolar. *Infancia y Aprendizaje*, 57, 49–66.
- Jiménez, J.E., Alvarez, C., Estévez, A., y Hernández-Valle, I. (2000). Onset-rimes units in visual word recognition in Spanish normal readers and reading disabled children. *Learning Disabilities Research & Practice*, 15, 135–141.
- Jiménez, J. E., Artiles, C., Rodríguez, C., Crespo, P., Pérez, F., y González, D. (2009). *Programa de Prevención de las Dificultades Específicas de Aprendizaje (PREDEA)*. Dirección General de Ordenación e Innovación Educativa de la Consejería de Educación, Universidades, Cultura y Deportes del Gobierno de Canarias.

- Jiménez, J. E., García, E., y Venegas, E. (en prensa). Are phonological processes the same or different in low literacy adults and children with or without reading disabilities in a consistent orthography? *Reading and Writing: An Interdisciplinary Journal*.
- Jiménez, J. E., y Guzmán, R. (2003). The influence of code-oriented versus meaning-oriented approaches to reading instruction on word recognition in the Spanish language. *International Journal of Psychology*, 38, 65–78.
- Jiménez, J. E., Guzmán, R., y Artiles, C. (1997). Efectos de la frecuencia silábica posicional en el aprendizaje de la lectura. *Cognitiva*, 1, 3–27.
- Jiménez, J.E., y Haro, C. (1995). Effects of word linguistic properties on phonological awareness in Spanish children. *Journal of Educational Psychology*, 87, 193–201.
- Jiménez, J.E., y Hernández-Valle, I. (2000). Word identification and reading disorders in the Spanish language. *Journal of Learning Disabilities*, 33, 44–60.
- Jiménez, J. E., y Muñetón Ayala, M. A. (2002): *Dificultades de aprendizaje en escritura: aplicaciones de la psicolingüística y de las nuevas tecnologías*. Madrid: Editorial Trotta S.A.
- Jiménez, J. E., y Ortiz, M. (2000). Metalinguistic awareness and reading acquisition in the Spanish language. *The Spanish Journal of Psychology*, 3, 37–46.
- Jiménez, J. E., y O'Shanahan, I. (2008). Enseñanza de la lectura: de la teoría y la investigación a la práctica educativa. *Revista Iberoamericana de Educación*, 45, 1–22.
- Jiménez, J. E., O'Shanahan, I., Pérez, F., Artiles, C., Muñetón, M., Guzmán, R., et al. (2008). Evolución de la escritura de palabras de ortografía arbitraria en lengua castellana. *Psicothema*, 20, 786–794.
- Jiménez, J. E., Pérez, F., O'Shanahan, I., Muñetón, M., y Rojas, E. (2009). ¿Pueden tener problemas con la ortografía los niños que leen bien? *Revista Española de Pedagogía*, 242, 45–60.
- Jiménez, J. E., y Rumeu, M. (1989). Writing disorders and their relationship to reading-writing methods: A longitudinal study, *Journal of Learning Disabilities*, 22, 195–199.
- Jiménez, J. E., Siegel, L. S., O'Shanahan, I., y Ford, L. (2009). The relative roles of IQ and cognitive processes in reading disability. *Educational Psychology*, 29, 27–43.
- Jiménez, J. E., y Venegas, E. (2004). Defining phonological awareness and its relationship to reading skills in illiterate adults. *Journal of Educational Psychology*, 96(4), 798–810.
- Joshi, R. M., y Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, 21, 85–97.

- Justicia, F. (1995). *El desarrollo del vocabulario. Diccionario de frecuencias*. Granada: Servicio de Publicaciones, Universidad de Granada.
- Kame'enui, E. J., Fuchs, L., Francis, D. J., Good, R. H., III, O'Connor, R. E., Simmons, D. C., et al. (2006). The adequacy of tools for assessing reading competence: A framework and review. *Educational Researcher*, 35(4), 3–11.
- Kaminski, R. A., Good, R. H., III, Baker, D., Cummings, K., Dufour-Martel, C., Fleming, K., et al. (2006). *Position paper on use of DIBELS for system-wide accountability decisions*. Consultado en enero de 2007, en www.cde.state.co.us/action/CBLA/Accountability_2006-11-16.pdf
- Kuhn, M. R., y Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, 95, 3–21.
- Leybaert, J., y Content, A. (1995). Reading and spelling acquisition in two different teaching methods: A test of the independence hypothesis. *Reading and Writing: An Interdisciplinary Journal*, 7, 65–88.
- Linan-Thompson, S., y Vaughn, S. (2004). *Research-based methods of reading instruction: Grades K-3*. Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Linan-Thompson, S., y Vaughn, S. (2007). *Research-based methods of reading instruction for English language learners: Grades K-4*. Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Lonigan, C., Wagner, R., Torgesen, J. K., y Rashotte, C. (2002). *Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)*. Tallahassee, Florida: Department of Psychology, Florida State University.
- Lundberg, I., y Høien, T. (1991). Initial enabling knowledge and skills in reading acquisition: Print awareness and phonological segmentation. En D. Sawyer y B. Fox (Eds.), *Phonological awareness in reading: The evolution of current perspective* (págs. 73–95). New York: Springer-Verlag.
- Mason, J. M., y Allen, J. (1986). A review of emergent literacy with implications for research and practice in reading. En E.Z. Rothkopf (Ed.), *Review of research in education* (Vol. 13, págs. 3–47). Washington DC: American Educational Research Association.
- Masonheimer, P. E., Drum, P. A., y Ehri, L. C. (1984). Does environmental print identification lead children into word reading? *Journal of Reading Behavior*, 16, 257–271.
- Morais, J., Bertelson, P., Cary, L., y Alegría, J. (1986). Literacy training and speech segmentation. *Cognition*, 24, 45–64.

- Morais, J., Cary, L., Alegría, J., y Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7, 323–333.
- Mullins, I., Martin, M., Gonzalez, E., y Chrostowski, S. (2004). *TIMSS 2003 international mathematics report*. Boston, Massachusetts: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mullins, I. V. S., Martin, M. O., Kennedy, A. M., y Foy, P. (2007). *IEA's Progress in International Reading Literacy Study in primary school in 40 countries*. Chestnut Hill, Massachusetts: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Muter, V., Hulme, C., y Snowling, M. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology*, 65, 370–396.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups (NIH Publication No. 00-4754)*. Consultado en agosto de 2007, en http://www.nichd.nih.gov/publications/nrp/upload/report_pdf.pdf
- National Literacy Panel. (2004). *National Literacy Panel on Language Minority Children and Youth: Progress report*. Consultado en agosto de 2007, en <http://www.cal.org/natl-lit-panel/reports/progress.pdf>
- Organización para la Cooperación y el Desarrollo Económico (OCDE). (2001). *Knowledge and skills for life: First results from PISA 2000*. París, Francia: Autor.
- OCDE. (2002). *Reading for change: Performance and engagement across countries*. París, Francia: Autor.
- OCDE. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. París, Francia: Autor.
- OCDE. (2004). *Learning for tomorrow's world: First results from PISA 2003*. París, Francia: Autor. Consultado en enero de 2008, en http://www.oecd.org/document/55/0,3343,en_32252351_32236173_33917303_1_1_1_1_00.html
- OCDE. (2005). *PISA 2003 technical report*. París, Francia: Autor.
- OCDE. (2006). *PISA 2006 science competencies for tomorrow's world*. París, Francia: Autor.
- Orr, D. B., y Graham, W. R. (1968). Development of a listening comprehension test to identify educational potential among disadvantaged junior high school students. *American Educational Research Journal*, 5(2), 167–180.

- Ortiz, M.R. (1994). *Conciencia fonológica y conciencia del lenguaje escrito en el aprendizaje de la lectura*. Manuscrito inédito. La Laguna, Tenerife, Islas Canarias: Universidad La Laguna.
- Ortiz, M. R., y Jiménez, J. E. (1993). *Prueba de conocimientos sobre el lenguaje escrito (CLE)*. Madrid: TEA Ediciones.
- Ortiz, M. R., y Jiménez, J. E. (2001). Concepciones tempranas acerca del lenguaje escrito en prelectores. *Infancia y Aprendizaje*, 24, 215–231.
- Paris, S. G., y Paris, A. H. (2006). Chapter 2: Assessments of early reading. En W. Damon y R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development, 6th Edition* (Vol. 4: Child Psychology in Practice). Hoboken, New Jersey: John Wiley and Sons.
- Peereman, R., y Content, A. (1999). LEXOP: A lexical database providing orthography-phonology statistics for French monosyllabic words. *Behavioral Methods, Instruments and Computers* (31), 376–379.
- Pratham. (2005). *Annual Status of Education Report (ASER): Final report*. Consultado en abril de 2006, de <http://www.pratham.org/aserrep.php>
- Read, C. A., Zhang, Nie, H., y Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic reading. *Cognition*, 24, 31–44.
- Riley, J. (1999). The reading debate. En T. Nunes (Ed.), *Learning to read: An integrated view from research and practice* (págs. 217–227). Dordrecht: Kluwer.
- Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J., y Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology*, 96, 119–129.
- Sebastián, N., Martí, M.A., Carreiras, M., y Cuetos, F. (2000). *LEXESP. Léxico informatizado del español*. Barcelona: Edicions Universitat de Barcelona.
- Serrano, F., Defior, S., y Jiménez, G. (2005). Evolución de la relación entre conciencia fonológica y lenguaje escrito en niños españoles de primer curso de educación primaria. *IberPsicología*, 10.3.15. Consultado en septiembre de 2009, de http://www.fedap.es/IberPsicologia/iberpsi10/congreso_lisboa/serrano/serrano.htm
- Seymour, P. H. K., Aro, M., y Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174.
- Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, 72, 95–129.

- Share, D. L. (2008). On the anglocentrism of current reading research and practice: The perils of overreliance on an “outlier orthography.” *Psychological Bulletin*, 134, 584–615.
- Sinclair, A., Jarvella, R., y Levelt, W. (1978). *The child's conception of language*. Berlín: Springer-Verlag.
- Snow, C. E., Burns, M. S., y Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: Committee on Preventing of Reading Difficulties in Young Children y National Academy Press.
- Sowden, P. T., y Stevenson, J. (1994). Beginning reading strategies in children experiencing contrasting teaching methods. *Reading and Writing: An Interdisciplinary Journal*, 6, 109–123.
- Sprenger-Charolles, L., Colé, P., y Serniclaes, W. (2006). *Reading acquisition and developmental dyslexia*. Nueva York: Psychology Press.
- Sprenger-Charolles, L., Siegel, L., Béchenec, D., y Serniclaes, W. (2003). Development of phonological and orthographic processing in reading aloud, in silent reading and in spelling: A four year longitudinal study. *Journal of Experimental Child Psychology*, 84, 194–217.
- Sprenger-Charolles, L., Siegel, L. S., y Bonnet, P. (1998). Phonological mediation and orthographic factors in reading and spelling. *Journal of Experimental Child Psychology*, 68, 134–155.
- Stage S. A., y Wagner, R. K. (1992). Development of young children's phonological and orthographic knowledge as revealed by their spellings. *Developmental Psychology*, 28, 287–296.
- Stahl, S. A., McKenna, M. C., y Pagnucco, J. R. (1994). The effects of whole language instruction: An update and a reappraisal. *Educational Psychologist*, 29, 175–186.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360–406.
- Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *America Educator/American Federation of Teachers*, 22, 32–39.
- Vellutino, F. R., y Scanlon, D. M. (2002). Emergent literacy skills, early instruction and individual differences as determinants of difficulties in learning to read: The case for early intervention. En S. Neuman y D. Dickinson. (Eds.), *Handbook of early literacy research* (págs. 295–321). Londres: The Guilford Press.

- Wagner, D. A. (2003). Smaller, quicker, cheaper: Alternative strategies for literacy assessment in the UN Literacy Decade. *International Journal of Educational Research*, 39, 293–309.
- Wimmer, H., y Hummer, P. (1990). How German speaking first graders read and spell: Doubt on the importance of the logographic stage. *Applied Psycholinguistics*, 11, 349–368.
- Wimmer, H., y Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition*, 51, 91–103.
- Whitehurst, G. J., y Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, 69, 848–872.
- Yovanoff, P., Duesbery, L., Alonzo, J., y Tindall, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement* (Fall), 4–12.
- Zeno, S. M., Ivenz, S. H., Millard, R. T., y Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, New York: Touchstone Applied Science Associates.
- Ziegler, J., y Goswami, U. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 13(1), 3–29.
- Ziegler, J. C., y Goswami, U. (2006). Becoming literate in different languages: Similar problems, different solutions. *Developmental Science*, 9, 429–436.

Anexo A. Normas sobre fluidez de la lectura oral en inglés para Estados Unidos

Grade	Percentile	Fall WCPM	Winter WCPM	Spring WCPM
1	90		81	111
	75		47	82
	50		23	53
	25		12	28
	10		6	15
	SD		32	39
	Count		16,950	19,434
2	90	106	125	142
	75	79	100	117
	50	51	72	89
	25	25	42	61
	10	11	18	31
	SD	37	41	42
	Count	15,896	18,229	20,128
3	90	128	146	162
	75	99	120	137
	50	71	92	107
	25	44	62	78
	10	21	36	48
	SD	40	43	44
	Count	16,988	17,383	18,372
4	90	145	166	180
	75	119	139	152
	50	94	112	123
	25	68	87	98
	10	45	61	72
	SD	40	41	43
	Count	16,523	14,572	16,269
5	90	166	182	194
	75	139	156	168
	50	110	127	139
	25	85	99	109
	10	61	74	83
	SD	45	44	45
	Count	16,212	13,331	15,292
6	90	177	195	204
	75	153	167	177
	50	127	140	150
	25	98	111	122
	10	68	82	93
	SD	42	45	44
	Count	10,520	9,218	11,290
7	90	180	192	202
	75	156	165	177
	50	128	136	150
	25	102	109	123
	10	79	88	98
	SD	40	43	41
	Count	6,482	4,058	5,998
8	90	185	199	199
	75	161	173	177
	50	133	146	151
	25	106	115	124
	10	77	84	97
	SD	43	45	41
	Count	5,546	3,496	5,335

WCPM: Words correct per minute
SD: Standard deviation
Count: Number of student scores

Fuente: Hasbrouck, J., y Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, 59(7), 636–644. Los textos están diseñados para que sean apropiados para el nivel de cada grado.

Anexo B. Consideraciones acerca del tamaño de la muestra en la *Early Grade Reading Assessment*

Introducción

Esta nota presenta consideraciones elementales del tamaño de la muestra, aplicables a las muestras de la Early Grade Reading Assessment (EGRA). Está diseñada para informar al personal del ministerio, los donantes o cualquier otro actor interesado en establecer una EGRA de los requisitos del tamaño de la muestra y los cálculos a llevarse a cabo. La nota asume un conocimiento de estadística elemental, sólo resalta las cuestiones que usualmente no se encuentran en un libro de texto universitario de esta materia, y no define la terminología estadística común. Como se vio ya en la Sección VI, es posible efectuar la mayoría de estos cálculos usando Excel; se prefiere, sin embargo, a Stata o SPSS (para la conservación de los datos y porque los cálculos pueden programarse usando archivos de sintaxis). Este anexo se refiere a cálculos efectuados usando tanto Stata como Excel.

En términos generales, el tamaño de la muestra queda determinado por un grupo de factores que incluyen las características de los datos y una serie supuestos seleccionados por el investigador. En el caso de EGRA, el tamaño de la muestra queda determinado por la variabilidad del desempeño de los estudiantes en anteriores evaluaciones EGRA, el nivel de precisión que el investigador desearía ver en los datos resultantes, y el método de muestreo aplicado.

Cuanto mayor sea la variabilidad en el desempeño de los estudiantes, tanto mayor será el tamaño requerido de la muestra. Si, por ejemplo, uno va a un centro educativo donde todos los educandos están leyendo exactamente con el mismo nivel de fluidez, sólo tendríamos que muestrear a uno de ellos para efectuar un cálculo estimado de la fluidez promedio de la lectura en dicho colegio. Desafortunadamente no podemos saber por adelantado la variabilidad real cuando planeamos el tamaño de la muestra. Sabemos que siempre hay *cierta* variabilidad entre los educandos. Una forma de estimar el tamaño de la muestra para un nuevo caso (nuevo país, nueva región) es examinando otros casos.

Como ya se indicó, además de las características de los datos, la precisión que el investigador desee en los resultados también tendrá un efecto sobre el tamaño de la muestra. Para entender la cuestión elemental en juego aquí, debemos entender que una muestra sólo puede dar estimadores del valor subyacente en la población total que la tiene a ella como base. Por ejemplo, la fluidez de lectura promedio de una muestra sólo es un estimador de la fluidez de lectura de la población. Después de todo, una muestra sólo es eso, y cualquier otra podría arrojar un valor distinto. Luego queremos saber exactamente con cuánta precisión cualquier muestra dada estima el valor en la población. ¿Con qué precisión nuestro estimador basado en una muestra de, digamos, la fluidez de lectura, estima la fluidez en la población? Esta es una cuestión de especial importancia. La ventaja de los muestreos es que ellos pueden reducir los costes en relación con la evaluación de toda la población, pero si los estimadores son demasiado imprecisos, dicha ventaja no tendrá mucho valor. Una mayor precisión generalmente necesita de una muestra más grande, ¿pero cuánto más?

Para comenzar a abordar esta cuestión, es importante advertir que la noción de “precisión” tiene dos aspectos. En primer lugar, ¿cuál es el tamaño del rango dentro del cual puede caer el puntaje del desempeño de los estudiantes? En la estadística es tradicional decir algo así como

que “nuestra muestra estima que los niños leen 50 palabras por minuto, por ende los niños pertenecientes a la población muy probablemente están leyendo 50 palabras por minuto, más o menos 5 palabras por minuto”. Un estimador menos preciso diría: “nuestra muestra estima que los niños leen 50 palabras por minuto, y por ende los de la población muy probablemente leen 50 palabras por minuto, más o menos 20 palabras por minuto”. La noción de “más o menos” se conoce como un intervalo de confianza (IC) y el valor real del “más o menos” es la *amplitud* del mismo. Cuanto menor sea la amplitud, tanto más precisos serán los resultados. A la segunda cuestión ya aludimos anteriormente: podemos decir: “Por ende, es ‘lo más probable’ que los niños en la población lean 50 palabras por minuto, más o menos 10 palabras por minuto”. ¿Pero qué tan probable es “lo más probable”? Este es el segundo aspecto de la precisión: ¿cuánta confianza queremos tener de que hemos captado el valor real del desempeño del educando?: ¿90% de confianza, 95% de confianza o 99% de confianza? Esto se conoce en términos estadísticos como el *nivel de confianza*. Una forma intuitiva de interpretar esto, en el contexto del muestreo, es tomar conciencia de que decir: “estamos 99% confiados en que el promedio para la población es de 50 palabras por minuto, más o menos 5”, equivale más o menos a decir: “hay alrededor de una posibilidad de 1% de que toda muestra dada nos habría dado una muestra promedio de 50, si la media de la población cayera fuera del rango de 45 a 55”. De modo que tenemos un *nivel de confianza* y una *amplitud* del intervalo de confianza. Podemos afirmar con un nivel dado de precisión, si nuestro promedio de la muestra estima el promedio de la población.

El enfoque usado al muestrear o seleccionar educandos para su participación en EGRA también tendrá un impacto sobre el diseño de la muestra. Examinaremos esto con mayor detenimiento a continuación.

Dado que las muestras de mayor tamaño pueden dar cabida a niveles más grandes de variabilidad en el desempeño de los educandos y brindar resultados más precisos, podríamos concluir que siempre debe seleccionarse una muestra grande. Las muestras de este tipo son desafortunadamente muy costosas. El investigador debe elegir una que sea lo suficientemente grande como para proporcionar datos confiables, sin que al mismo tiempo necesite niveles de financiamiento excesivos. Además, toda evaluación con base en los centros educativos interrumpe los procedimientos escolares, resta tiempo a las lecciones y resulta siendo una imposición. Hay, por lo tanto, ciertas consideraciones éticas que abonan en favor de mantener el tamaño de la muestra relativamente pequeño.

Como viéramos en las secciones iniciales de este manual, la finalidad para la cual se diseñó a EGRA fue la de efectuar un diagnóstico a nivel del sistema, para un nivel dado. Teniendo esta finalidad como base, el examen que aparece a continuación se concentra en estrategias con que extraer una muestra nacionalmente representativa. A lo largo de este anexo, la discusión queda encuadrada en función del tamaño mínimo de la muestra necesario para informar de los resultados de un grado. Dicho esto, la mayoría de los ejercicios de EGRA fueron diseñados para evaluar en múltiples grados iniciales (por ejemplo, los grados 1–3 o 2–4). Por lo tanto, los cálculos que aparecen a continuación son aplicables al tamaño mínimo de la muestra necesario para cada grado de interés: si los países están interesados en evaluar habilidades en tres grados, debe extraerse la muestra para cada uno de ellos usando los parámetros que aparecen a continuación.

Afortunadamente se han acumulado suficientes experiencias con EGRA como para permitir determinar razonablemente el tamaño recomendado de las muestras. Las experiencias efectuadas en Perú, Pakistán, Gambia, Senegal y Jamaica, así como dos experiencias en Kenia, pueden usarse ahora para sustentar la discusión del tamaño de la muestra y las

recomendaciones. La tabla B1 contiene información clave de los antecedentes, útil para determinar el tamaño de una muestra, en todos los países mencionados anteriormente. En él se usa la fluidez de la lectura del párrafo como la variable de interés clave. No se dan los nombres de los países, ya que estos datos se presentan para ilustrar patrones de diferencias dentro de ellos entre grados y géneros (para mencionar dos atributos), así como la variabilidad en general. No se busca que estos datos sean usados en comparaciones de la fluidez de la lectura entre países.

Tabla B1. Estimadores claves de la fluidez de lectura a través de países y grados, en diversas aplicaciones de EGRA

País	Grado				Total	Diferencia de género para todos los grados	Peso aproximado por país
	1	2	3	4			
País 1							6
Hombre	2.4	17.8	28.7		16.2		
Mujer	3.3	17.0	35.6		18.6		
Promedio referido al grado a través de los géneros	2.9	17.4	32.4		17.5	-2.4	
Desviación estándar referida al grado	5.9	17.4	23.5		21.0		
Ganancia promedio entre grados		14.8					
País 2							10
Hombre	1.9	4.3	9.9		5.3		
Mujer	2.4	3.6	8.6		4.8		
Promedio referido al grado a través de los géneros	2.2	4.0	9.2		5.1	0.5	
Desviación estándar referida al grado	9.3	12.4	19.9		14.8		
Ganancia promedio entre grados		3.5					
País 3							8
Hombre		59.8	66.8		63.5		
Mujer		58.3	78.7		68.3		
Promedio referido al grado a través de los géneros		59.0	73.1		66.1	-4.9	
Desviación estándar referida al grado		46.8	48.1		47.9		
Ganancia promedio entre grados		14.1					
País 4							10
Hombre	23.2	30.4	50.3	68.3	45.8		
Mujer	28.2	36.2	58.1	90.2	56.1		
Promedio referido al grado a través de los géneros	25.3	33.1	53.9	78.1	50.5	-10.3	
Desviación estándar referida al grado	30.5	34.4	39.2	46.5	43.2		
Ganancia promedio entre grados		17.6					
País 5							6
Promedio referido al grado a través de los géneros	9.2	29.3					

País	Grado				Total	Diferencia de género para todos los grados	Peso aproximado por país
	1	2	3	4			
Desviación estándar referida al grado	16.9	30.7			27.4		
Ganancia promedio entre grados		20.1					
País 6							3
Hombre	6.8	30.0	97.2	100.5	59.6		
Mujer	7.3	31.5	44.4	68.5	37.5		
Promedio referido al grado a través de los géneros	7.0	30.8	69.3	85.0	48.2	22.1	
Desviación estándar referida al grado	15.0	39.0	79.3	68.9	64.0		
Ganancia promedio entre grados		26.0					
País 6 – caso especial							10
Hombre		11.5					
Mujer		11.2				0.4	
Promedio referido al grado a través de los géneros		11.4					
Desviación estándar referida al grado		16.2					
Ganancia promedio entre grados		NA					
Promedios entre países							
Fluidez promedio por grado a través de todos los países	10.5	25.0	43.7	79.7			
Desviación estándar promedio de la fluidez por grado a través de todos los países	16.5	26.5	36.6	51.6			
Desviación estándar promedio a través de todos los grados (usando todos los puntos de datos arriba, no promedios entre los países, como en la fila inmediatamente superior)			29.2				
Ganancia promedio entre grados a través de todos los países		14.1					
Diferencia de género promedio a través de todos los países		-3.4					

NA = No aplicable.

Notas: Todos los promedios están ponderados. El peso aproximado del país tiene como base el juicio que los autores hacen de la importancia total, dados el tamaño de la muestra y el rigor con que ésta fue seleccionada.

Fuentes: Calculado a partir de las bases de datos de EGRA de diversos países.

En lo que sigue explicaremos cómo calcular el tamaño de las muestras dados los datos existentes de EGRA, los supuestos intervalos de confianza, los niveles de confianza y un enfoque de muestreo fijo.

Para recordar la definición ya dada, un *intervalo de confianza* es un rango de valores (por oposición a una cifra) usado para estimar un parámetro de población (como la fluidez de lectura

promedio de la población) en base al estimado de una muestra (el estimado de la fluidez en base a una muestra). Cuanto menor o más pequeño sea la amplitud del intervalo de confianza, tanto más confiables o precisos serán los resultados. El tamaño del intervalo de confianza que el investigador fije, dependerá de las características de la variable bajo estudio. Un enfoque frecuente para sugerir una amplitud apropiada para un intervalo de confianza es examinar los rangos de variación entre atributos claves de los estudiantes, como el grado o el género, y sugerir que los intervalos de confianza sean lo suficientemente angostos como para permitir distinguir el desempeño en estos atributos claves. Por ejemplo, es razonable pedir que los intervalos de confianza de distintos grados no se superpongan entre sí. A partir del último panel de la tabla B1 comentada anteriormente, queda claro que las diferencias promedio entre grados de todos los países es de 14.1, o 14 para tomar una cifra redondeada. Éste, entonces, parece ser un ancho razonable sobre el cual basar la estimación del tamaño de la muestra.

Los *intervalos de confianza* están asociados con *niveles de confianza* específicos. El nivel de confianza nos dice la probabilidad de que el intervalo de confianza contenga el parámetro de población real (la fluidez de lectura media). Cuanto mayor sea el nivel de confianza, tanto mayor será el nivel de precisión. Los investigadores generalmente asumen niveles de confianza de 90%, 95% y 99%, considerándose un 90% como algo marginal.

Método de muestreo

Como ya se dijo, el método del muestreo que se aplique también tendrá un efecto sobre los requisitos del tamaño de la muestra. *Ceteris paribus*, seleccionar los alumnos aleatoriamente en un listado nacional requerirá de una muestra más pequeña, en tanto que las muestras *estratificadas* y *por conglomerados* necesitarán muestras de tamaño relativamente más grande. Aunque podría parecer contradictorio, las muestras puramente aleatorias son relativamente costosas en comparación con otros métodos de muestreo. Si, por ejemplo, intentáramos aplicar una muestra aleatoria pura y simple de 400 niños, podríamos tener que enfrentar una situación en la cual habría que ir a casi 400 centros educativos y evaluar un solo niño en cada escuela, lo que incrementaría enormemente los costes de transporte y mano de obra.²²

Además, para obtener una muestra aleatoria simple de niños necesitaríamos, en principio, una lista de todos los del país y su ubicación. En la mayoría de los países no se cuenta con tales listas. Con el muestreo por conglomerados se selecciona primero los centros educativos y luego a los educandos dentro de los colegios (conglomerados). Escoger primero a los colegios y luego a los niños reduce los costes y el tiempo de viaje, y asimismo elimina la necesidad de depender de un listado nacional de educandos. Dado que gran parte del costo de las encuestas es en primer lugar llegar a los colegios, uno podría muy bien evaluar a tantos niños como sea factible en cada centro educativo en una mañana de visita, como forma de incrementar el tamaño de la muestra a un coste relativamente bajo.

Las anteriores aplicaciones EGRA mostraron que es posible que un examinador entreviste entre 12 y 15 niños en una mañana escolar.²³ Asumiendo—*sólo a modo de ejemplo*—una muestra de 15 niños por centro educativo, una muestra con un tamaño de 400 requeriría que

²² Habría necesidad de ir a casi 400 colegios porque dada la suerte del sorteo, y dependiendo del número total de centros educativos en el país, algunos de ellos tendrían más de un niño escogido. Es muy probable que en un país con digamos sólo 500 colegios, muestrear 400 niños con una muestra aleatoria simple arroje varios casos en donde habría más de un niño por colegio, lo que no sucedería en un país con unos 80,000 centros educativos.

²³ Este número específico de niños que se puede entrevistar depende de la versión del instrumento EGRA que se esté aplicando, el número de lenguas en que se esté llevando a cabo, y de si EGRA forma parte o no de otras investigaciones que se vienen llevando a cabo en el colegio.

se visiten sólo unas 27 escuelas, lo que es una economía considerable con respecto a visitar unas 400. (La muestra real deseada de niños por colegio podría variar, dependiendo de las características del país.) Recomendamos por ello aplicar un método de muestreo por conglomerados.

Sin embargo, aplicar el método de conglomerados tiene como resultado una pérdida de realismo, puesto que los niños típicamente varían menos dentro de los centros educativos de lo que el “niño representativo” de cada colegio varía de los niños en otros colegios. Los niños dentro de un centro educativo tienden a pertenecer a la misma clase social, a tener las mismas ventajas o desventajas en el lenguaje, o a tener una calidad similar de docentes y a quedar expuestos a prácticas administrativas similares, en mayor medida que los niños de distintos colegios. En este sentido, la verdad o la variabilidad de la población entre niños tiende a quedar subestimada si uno usa un método de muestreo por conglomerados: en otras palabras, se consigue la eficiencia en los costes del transporte y la mano de obra a costa de perder información acerca de la variabilidad, con lo cual habrá una pérdida de precisión a menos que se hagan ajustes. Afortunadamente contamos con una medida que nos indica el grado en que los conglomerados pueden hacer que se subestime la variabilidad. Esta medida, a la que se conoce como el *efecto de diseño* (design effect, *DEFF*), puede ser usada para ajustar el tamaño de la muestra para dar cuenta de la pérdida en variabilidad provocada por los conglomerados.

A modo de recapitulación, hemos examinado cuatro puntos que deben incluirse en el cálculo del tamaño de nuestra muestra. Estos son:

1. *La variabilidad* en los puntajes de lectura de los educandos (u otra variable EGRA si así se desea)
2. *La amplitud del intervalo de confianza*, determinada por el investigador
3. El *nivel de confianza*, determinado por el investigador
4. El *DEFF* causado por la aplicación del muestreo por conglomerados

Calculando el tamaño de la muestra para un intervalo de confianza y un nivel de confianza dados

Como fórmula, podríamos representar el tamaño de la muestra necesaria como sigue:

$$n = 4 \left(\frac{CLtvalue \ DEFT \ SD}{Width} \right)^2,$$

donde:

n es el tamaño necesario de la muestra;

CLtvalue es el valor-*t* asociado al nivel de confianza escogido,

DEFT es la raíz cuadrada del DEFF, donde se usa DEFT porque el cuadrado del término nos vuelve a dar el DEFF;

SD es una desviación estándar (*standard deviation*), que es la medida de la variabilidad en nuestra variable elegida;

Width = la *amplitud del intervalo de confianza*, determinada por el investigador; y

el número 4 se deriva de la ecuación básica de un intervalo de confianza.²⁴

Como podemos ver en esta ecuación, los incrementos en el *nivel de confianza*, el *efecto de diseño* y la *variabilidad* (medidos por la SD), operan todos para incrementar el tamaño requerido de la muestra (*n*). A la inversa, todo incremento en la amplitud (*Width*) del intervalo de confianza reduce el requisito del tamaño de la muestra, pero por definición también reduce la precisión.

Para los fines de diseñar recomendaciones para el tamaño de la muestra, se calculan la raíz cuadrada del efecto de diseño (DEFT es la raíz cuadrada de DEFF) y la desviación estándar usando datos de aplicaciones anteriores de EGRA, utilizando los datos de la tabla.

El DEFF se calcula como sigue:

$$EDD = 1 + (\text{clustersize} - 1) ICC ,$$

donde:

clustersize es el tamaño del conglomerado promedio (el número de niños muestreados en cada centro educativo), y

ICC es el coeficiente de correlación intraclases (intraclass correlation coefficient).

Los incrementos en el *clustersize* o en el ICC incrementarán ambos el efecto de diseño. Si el *clustersize* es 1 (un niño por colegio en la muestra), entonces el ICC no importa y el DEFF será 1. En otras palabras, el conglomerado (*clustering*) no afecta la variabilidad estimada si el *clustersize* sólo es 1.

El ICC es una medida de cuánto de la variabilidad yace entre los colegios, y cuánto dentro de ellos. Una forma intuitiva de pensar esto es que ello indica la probabilidad de encontrar dos observaciones que están en el mismo conglomerado, en relación con hallar dos observaciones idénticas seleccionadas aleatoriamente. Por ejemplo, un ICC de 0.41 indicaría que uno tiene 41% más probabilidades de encontrar dos alumnos con la misma fluidez de lectura dentro de un conglomerado (colegio), que dos con los mismos niveles de fluidez y que hayan sido extraídos al azar en dos colegios cualquiera.

En la bibliografía hay diversas formas de entender el ICC. En el presente contexto, éste permite el uso del software Stata y se le calcula como sigue:

$$ICC = \frac{MSE_{between} - MSE_{within}}{MSE_{between} + (\text{clustersize} - 1) MSE_{within}} ,$$

²⁴ Esta ecuación se deriva de la fórmula tradicional de un intervalo de confianza, $\bar{X} \pm CLvalue \frac{SD DEFT}{\sqrt{n}}$,

donde la expresión al lado derecho del signo \pm es la amplitud de una cola. La amplitud total de dos colas es entonces $Width = 2 CLvalue \frac{SD DEFT}{\sqrt{n}}$. La manipulación algebraica lo llevará a uno a la ecuación usada en el

texto principal, y mostrará por qué el 2 se convierte en un 4.

donde:

MSE es el error cuadrático medio, y $clustersize$ es el tamaño promedio de los conglomerados (el número de niños en cada centro educativo seleccionado).

$MSE_{between}$ mide el monto de variación que existe entre colegios (nuestros conglomerados). Aritméticamente, $MSE_{between}$ es la suma de las desviaciones cuadradas existentes entre la media de cada conglomerado (de cada colegio) y la media general, ponderada con el tamaño del conglomerado (el número de niños muestreados en el colegio). MSE_{within} mide el monto de variación existente dentro de los colegios (nuestros conglomerados). En términos aritméticos, MSE_{within} viene a ser la suma de las desviaciones cuadradas existentes entre cada niño y la media del conglomerado (colegio), dividido por el número total de niños menos el número de conglomerados. En símbolos:

$$MSE_{between} = \frac{\sum_{j=1}^{cluster} n_j (\bar{X}_j - \tilde{X})^2}{cluster - 1}$$

y

$$MSE_{within} = \frac{\sum_{j=1}^{cluster} \sum_{i \in j=1}^{n_j} (x_{ij} - \bar{X}_j)^2}{\sum_{j=1}^{cluster} n_j - cluster},$$

donde:

\tilde{X} es la media “general” o global,

j es un índice de los conglomerados,

$i \in j$ es un índice del i -ésimo niño del conglomerado j ,

\bar{X}_j es la media del j -ésimo conglomerado (o colegio),

$cluster$ es el número de conglomerados, o el índice del último conglomerado, y

n_j es el tamaño del j -ésimo conglomerado, o el índice del último miembro del j -ésimo conglomerado.

Se puede usar el procedimiento de análisis de varianza (analysis of variance, ANOVA) en Excel para calcular tanto MSE_{within} como $MSE_{between}$.

La tabla B2 muestra un rango de estimados tanto del ICC como del DEFT para unos cuantos casos particulares, y las implicaciones que estas variables tienen para el número de colegios (conglomerados), así como para el tamaño total resultante de la muestra. Se asume un SD de 29 para todos los casos, se especifica una amplitud total del intervalo de confianza (amplitud de

dos colas) de 10, y se usa un nivel de confianza de 95%. El ICC, DEFT y *clustersize* son valores reales procedentes de los estudios de EGRA hasta ahora efectuados. El SD de 29 es un valor estilizado generalizado a partir de los diversos estudios de EGRA que hasta ahora se han llevado a cabo.

Tabla B2. ICC y DEFT estimados en diversos países y grados, mostrando en cada caso el tamaño promedio de los conglomerados

País	ICC	DEFT	<i>clustersize</i>	<i>n</i>
País A, Grado 3	0.17	1.2	3.75	198
País B, Grado 2	0.22	2.3	20	698
País C, Grado 3	0.25	1.6	7.57	356
País D, Grado 3	0.47	2.3	10.05	708
País E, Grado 2	0.48	1.8	5.35	416

Fuente: Cálculos de los autores, efectuados a partir de diversas encuestas EGRA.

Los DEFT en la tabla B2 se ven afectados tanto por el ICC como por el tamaño de los conglomerados. Ambos les afectan, como podemos ver en la ecuación del DEFT. En el País B, por ejemplo, el DEFT resulta ser ligeramente alto (2.3) a pesar que el ICC es bajo (0.22) porque el tamaño del conglomerado es 20, de modo que uno suprime bastantes variaciones tomando tantos de los niños de colegios específicos. En el País D, lo que impulsa el DEFT alto es un ICC alto. El DEFT es más bajo en el País A porque tanto el tamaño de los conglomerados como el ICC eran bajos. Los impactos sobre el tamaño requerido de la muestra son significativos. En el País A se necesitaría una muestra de apenas 198 niños (pero unos 53 centros educativos), en tanto que en el País D se necesitaría una muestra de 708 niños y unos 70 colegios.

Tamaños recomendados de las muestras para los intervalos de confianza

Al determinar los tamaños reales recomendados para las muestras, un requisito razonable sería que las diferencias entre grados sean de algún modo “significativas”, v.g., que los intervalos de confianza globales deberían ser lo suficientemente angostos, para que los intervalos de confianza de grados contiguos no se superpongan. Usando la tabla B1 vemos que la diferencia promedio entre grados es de 14. De este modo, una *amplitud* de 14 resulta sensata.

Si uno asume una amplitud de 14, un ICC de 0.45, un tamaño de conglomerado de 12 y un SD de 29 (según la tabla B1 superior), el tamaño “correcto” de la muestra es de 409 niños en cada grado de interés.

Dadas las muy pequeñas diferencias existentes entre niños y niñas en los datos de la muestra en la tabla B1 (y/o dado que las diferencias de género son sumamente variables entre países, a diferencia de la progresión gradual por grados), y dada la ecuación del tamaño de la muestra, debería quedar en claro que se requiere de una *amplitud* muy pequeña para detectar las diferencias de género, y por ello una muestra de muy gran tamaño: alrededor de 7,000 personas. Parece sensato aceptar la noción de que no es probable que los tamaños más razonables de las muestras capten diferencias estadísticamente significativas entre niños y

niñas. Esto resalta de paso la importancia que tiene distinguir entre una *diferencia sustantiva* y una *diferencia estadísticamente significativa*. En general, si hay alguna diferencia entre dos estratos cualesquiera de la población, incluso si no resulta sustantivamente interesante, los investigadores pueden “forzarla” a que sea estadísticamente significativa recurriendo a una muestra enorme. En otras palabras, las diferencias pequeñas de interés marginal pueden ser determinadas como estadísticamente significativas con una muestra de gran tamaño. El juicio que aquí se hace es que hasta ahora las diferencias de género parecen ser lo suficientemente pequeñas, de modo que sólo muestras muy grandes podrían detectarlas con significación estadística.

Dado que las SD pueden ser hasta de 60 o más, podría parecer un poco audaz proponer una de 29 al recomendar el tamaño de la muestra. Esto requiere de algo de examen. En primer lugar y tal como vimos en la tabla B1 superior, las SD más altas tienden vigorosamente a ser observadas sólo en los grados finales, y EGRA intenta sobre todo examinar los dos o tres primeros grados. También es el caso que las SD parecen estar correlacionadas de modo bastante positivo con los niveles medios estimados de fluidez, y con las diferencias estimadas entre grados (véanse los gráficos B1 y B2 en la parte inferior).

Gráfico B1. Media y desviación estándar de la fluidez

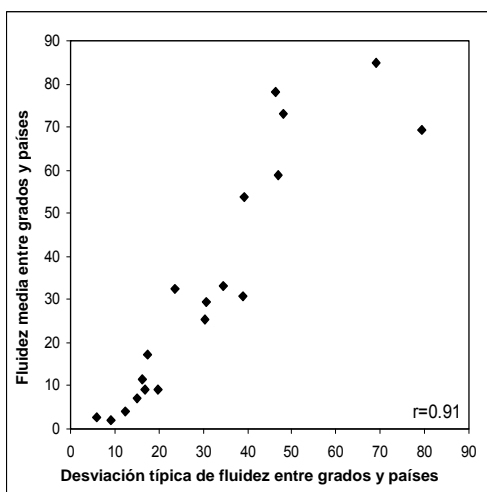
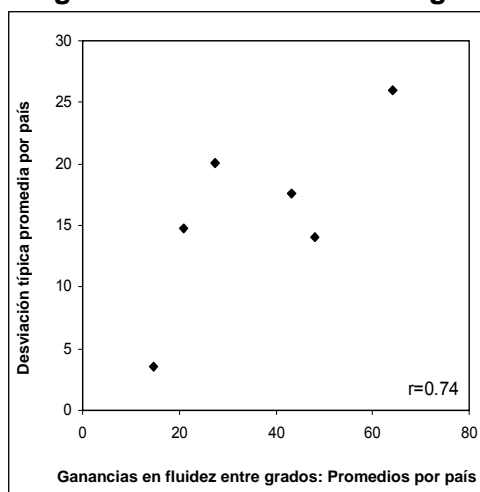


Gráfico B2. Desviación estándar y ganancias en fluidez entre grados



En los casos en donde las SD de la fluidez son altas, también lo son los niveles medios de fluidez, al igual que las diferencias entre grados. Esto quiere decir que las diferencias en general tenderán a ser detectables: cuando las SD son altas, también lo son las diferencias que uno intenta detectar. O, desde la perspectiva del intervalo de confianza, cuando las SD son altas, también lo son los puntos centrales de dichos intervalos. Esto significa que en principio, una *Amplitud [Width]* mayor bastará para establecer intervalos de confianza que no se superpongan, pero siempre se necesitará contar con una muestra de mayor tamaño. Es igualmente cierto que la amplitud absoluta de los intervalos importa menos que la relativa cuando todo lo que se desea es el intervalo de confianza mismo (y no la no superposición entre intervalos de distintos atributos de los educandos).

Por ejemplo, decir que la media de la muestra es 20, con un intervalo de confianza para la media de la población de [10, 30], transmite una sensación inferior de precisión que decir que la media de la muestra es de 45, con un intervalo de confianza para la media de la población de,

digamos, [33, 58], a pesar que este último es más ancho en sentido absoluto puesto que su amplitud es menor con respecto a su punto medio que la amplitud del primero. Pero incluso con estas condiciones se requerirán muestras más grandes si la SD es más alta.

Por ejemplo, con una SD de 50 pero ahora con diferencias entre grados de 20 en lugar de 14 (donde 20 es el valor esperado dada una SD de 50, en base al gráfico 2 de la parte superior), el tamaño de la muestra tendría que subir hasta 597—y no 409—para cada grado de interés.

Si uno deseara ser más cauteloso, tal vez deberían recomendarse tamaños de muestras de 600 educandos por grado de interés. *De otro lado, si con unas SD más altas uno simplemente decidiese contentarse con un intervalo de confianza de 90%, entonces el tamaño requerido de la muestra caerá nuevamente a 416.* (Muchas aplicaciones internacionales que buscan producir resultados rápidos, como el Programa Ampliado de Inmunización de la Organización Mundial de la Salud en el sector salud, usan intervalos de confianza de 90%.)

Parece sensato concluir que un muestreo de entre 410 y 600 educandos por grado de interés (en números redondos) dará cuenta de la mayoría de las contingencias.

Probando las hipótesis versus intervalos de confianza: implicaciones del muestreo

Al decidir el tamaño de las muestras, un factor a tomar en cuenta es si la base de las comparaciones entre grupos (v.g., entre niveles de fluidez en distintos grados) deberían ser intervalos de confianza que no se superponen o pruebas de hipótesis de una cola. Una práctica común es presentar IC para las variables claves y afirmar o implicar que los que no se superponen son una primera selección útil para ver si las diferencias entre grupos son significativas. Esto a menudo se hace porque el investigador no sabe de antemano qué contrastes o pruebas de hipótesis serán de mayor interés. En este sentido, en EGRA, presentar IC para las variables claves parece ser una práctica sensata. Además, en general los lectores que tienen un interés sustantivo por el asunto se preocupan bastante por los parámetros reales que se está estimando (por ejemplo, los niveles medios de fluidez) y su probable rango, y tal vez les interesan menos si las diferencias entre las subpoblaciones de interés son estadísticamente significativas.

Sin embargo, intentar hacer que los IC sean lo suficientemente angostos como para no superponerse y detectar así una diferencia dada entre las medias, requiere de muestras de mayor tamaño. Efectuar pruebas de hipótesis de una cola podría necesitar muestras de menor tamaño. De otro lado, las pruebas de hipótesis son más difíciles de interpretar, atrayendo tal vez demasiado la atención hacia la “significación estadística” y alejándola en algo de los parámetros bajo consideración. Es más, sólo se puede lograr parte de la economía al efectuar pruebas de hipótesis si éstas son de una cola.

Hay cierto debate en la bibliografía referida a las evaluaciones, en lo que toca a las condiciones que justifican una prueba de hipótesis de una cola. Sin embargo, el debate no resultó concluyente, de modo que podría ser útil recordar las cuestiones a mano.

La prueba de una hipótesis por lo general postula una hipótesis “nula” según la cual, digamos (usando la fluidez como ejemplo), el nivel de fluidez de un grado dado es igual al nivel de fluidez de un grado previo, o que el nivel de fluidez después de una intervención es el mismo

que el nivel anterior a ella. Uno postula entonces hipótesis alternativas. Una forma de hipótesis alternativa es que el nivel de fluidez de un grado más alto es simplemente distinto del de un grado previo, o que el nivel de fluidez después de una intervención difiere del nivel anterior a ella. Para probar esta hipótesis, uno lleva a cabo una prueba de hipótesis de dos colas. Esto es común cuando uno se interesa por análisis algo exploratorios, donde cierto tratamiento o variable (nivel de ruralidad, experiencia del docente, etc.) podría tener un impacto positivo o negativo en otra cosa (los puntajes de las pruebas podrían recibir un impacto negativo o positivo del grado de ruralidad, y uno no tiene una sólida razón a priori para probar una hipótesis que va en una dirección particular).

En la mayoría de las aplicaciones de EGRA parece razonable creer que la mayoría de las hipótesis que viene probándose, o la mayoría de los enunciados que uno podría desear hacer, son unidireccionales. Uno podría así estar justificado en postular las pruebas de hipótesis de una cola, para de este modo alcanzar economías en el tamaño de la muestra. Si hay buenas razones para creer que el análisis debe ser de naturaleza más exploratoria y descriptiva, deberá entonces usarse una evaluación de hipótesis de dos colas.

Sea cual fuere el enfoque, siempre es una buena idea presentar intervalos de confianza y no simplemente evaluar hipótesis. La mayoría de los programas de estadística, Stata inclusive, frecuentemente presentan ambos con el mismo comando, de modo que esto no es algo difícil de lograr. Los programas de uso más general, como Excel, no lo incluyen, pero los intervalos de confianza son muy fáciles de generar. La finalidad de presentar los IC es fomentar la concentración en el parámetro en cuestión, como la fluidez oral en un párrafo. Pero se ha señalado que si el tamaño de las muestras es lo bastante grande como para permitir la detección de las diferencias en las pruebas de hipótesis de una cola, la amplitud de los IC tenderá a ser relativamente grande. De este modo el enfoque EGRA debería decidir primero si las pruebas de hipótesis de una cola son aceptables, con la condición de que esto podría significar IC ligeramente más amplios. La discusión siguiente resalta estas cuestiones.

Supongamos que tenemos dos medias de muestra, \bar{X}_1 y \bar{X}_2 . Para mantener las cosas simples, digamos que los errores estándares estimados (standard error, SE) de ambos son iguales, de modo tal que $SE_1 = SE_2 = SE$. También asumimos, sin gran pérdida de la generalización, que esto se debe a iguales SD y a tamaños equivalentes de las muestras.²⁵ Para este examen nos quedaremos con pruebas de 5% o IC de 95%. Se asume que las ordenadas t son los grados apropiados de libertad. Los IC de 95% son

$$\begin{aligned} &\bar{X}_1 \pm t_{.025} SE \\ &\bar{X}_2 \pm t_{.025} SE , \end{aligned}$$

donde $t_{.025}$ es la ordenada t requerida para una prueba de dos colas de 5%, con los grados apropiados de libertad. La condición de que los dos IC de cada media no se superpongan, equivale a requerir que

$$\bar{X}_1 + t_{.025} SE < \bar{X}_2 - t_{.025} SE ,$$

o

²⁵ En realidad la mayoría de las SD y SE diferirán entre sí. En esta exposición se asumen la igualdad del tamaño de la muestra y la de las SD únicamente con fines de claridad.

$$\bar{X}_2 - \bar{X}_1 > t_{.025} SE + t_{.025} SE = 2t_{.025} SE ,$$

si la primera media estimada es menor que la segunda, y de igual modo—pero con signos diferentes—si la segunda es más pequeña, o de modo más general:

$$|\bar{X}_1 - \bar{X}_2| > 2t_{.025} SE ,$$

porque los IC de las medias son simétricos alrededor de la media y tienen la misma amplitud, dado que las SE y los grados de libertad (según les impulsa n) son los mismos.

Pero el requisito de que los IC para la *diferencia* no se superpongan *con 0* equivale a solicitar que

$$|\bar{X}_1 - \bar{X}_2| > 1.41 t_{.025} SE ,$$

dada la ecuación de la desviación estándar de una diferencia entre medias, que es como sigue, dado el supuesto de desviaciones estándar iguales y muestras iguales:

$$SD_{diff} = \sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}} = \sqrt{2 \frac{SD^2}{n}} = 1.41 DT .$$

Adviértase que la razón de 2 a 1.41 es 1.41, puesto que todo número dividido por su raíz cuadrada es igual a su raíz cuadrada. Esto significa que en el primer caso uno necesitaría una SE más pequeña que en el segundo, para así no crear ninguna superposición de los IC—más pequeños 1.41 veces. Dado que $SE = SD / \sqrt{n}$, una SE que es 1.41 veces más pequeña requiere de un tamaño de la muestra 2 veces más grande, puesto que

$$\frac{SE}{1.41} = \frac{SD}{1.41\sqrt{n}} = \frac{SD}{\sqrt{2n}} .$$

Las siguientes pruebas instantáneas de Stata (usando el comando “ttesti”) sirven de ilustración. Ellas usan los valores ya empleados en los ejemplos ya expuestos anteriormente. Para ilustrar el principio elemental referido a las diferencias entre intervalos de confianza y pruebas de hipótesis, nos concentramos en un caso en el cual el DEFF es 1. El *procedimiento* usado es el de varianzas desiguales, aunque en la práctica y para hacer que la exposición fuera más fácil, las desviaciones estándar ingresadas a las ilustraciones son iguales la una a la otra.

En primer lugar tenemos un caso en el cual el intervalo de confianza de la *diferencia* entre dos medias no se superpone al cero pero casi lo hace, como se ve en el área inferior resaltada. Adviértase que Stata presenta los IC para cada variable, el IC de la diferencia entre las variables, y todas las pruebas de hipótesis relevantes para la diferencia entre las variables.

ttesti 34 20 29 34 34 29, desigual

Prueba t de dos muestras con varianzas desiguales

	Obs	Media	Err. Est.	Desv. Est.	[Inter. de Conf. 95%]	
x	34	20	4.973459	29	9.881422	30.11858
y	34	34	4.973459	29	23.88142	44.11858
combinado	68	27	3.593661	29.63409	19.82702	34.17298
dif		-14	7.033533		-28.0429	.042902
dif = media(x) - media(y)				t =	-1.9905	
Ho: dif = 0				Grados de libertad de Satterthwaite =	66	
Ha: dif < 0		Ha: dif != 0		Ha: dif > 0		
Pr(T < t) = 0.0253		Pr(T > t) = 0.0507		Pr(T > t) = 0.9747		

Los IC de ambas medias se superponen considerablemente, como se observa en las dos áreas superiores resaltadas, pero el IC de la *diferencia* no se superpone al cero (aunque casi lo hace por diseño), como lo vemos en el área inferior resaltada. Sin embargo, esta realmente es la forma correcta de interpretar el requisito de detectar una diferencia entre los grupos. Para evitar la superposición en los IC para las mismas medias tendríamos que duplicar el tamaño de las muestras.

La siguiente prueba muestra que con la duplicación del tamaño de la muestra, los IC de las medias individuales apenas si evitan superponerse, como vemos en las áreas superiores resaltadas:

ttesti 69 20 29 69 34 29, desigual

Prueba t de dos muestras con varianzas desiguales

	Obs	Media	Err. Est.	Desv. Est.	[Inter. de Conf. 95%]	
x	69	20	3.49119	29	13.03344	26.96656
y	69	34	3.49119	29	27.03344	40.96656
combinado	138	27	2.531281	29.73582	21.99457	32.00543
dif		-14	4.937288		-23.76379	-4.236213
dif = media(x) - media(y)				t =	-2.8356	
Ho: dif = 0				Grados de libertad de Satterthwaite =	136	
Ha: dif < 0		Ha: dif != 0		Ha: dif > 0		
Pr(T < t) = 0.0026		Pr(T > t) = 0.0053		Pr(T > t) = 0.9974		

Pero una duplicación del tamaño de la muestra es un precio alto (e innecesario) que pagar para tener IC no superpuestos para las medias, en lugar de un IC no superpuesto con cero para la diferencia entre ellas. Podemos ver esto en el hecho que el IC para la diferencia entre las medias está bastante lejos del cero (el resaltado del medio), o en el hecho de que la prueba de una hipótesis de dos colas para la diferencia entre las dos medias arroja un valor de probabilidad muy por debajo del umbral de 5% (el resaltado inferior).

Pero uno tiene incluso algo más de libertad de acción. La mayor parte de la ganancia en eficiencia entre la prueba de las hipótesis en torno a la noción de “intervalos de confianza no superpuestos”, se alcanza simplemente planteando el problema como una prueba de hipótesis. Pero si lo deseamos y justificamos *a priori*, podemos ganar un poquito más de eficiencia suponiendo una prueba de hipótesis de una cola. Advuértase que una prueba de una hipótesis *de una cola* es sumamente fuerte a pesar que el IC de la diferencia casi toca el cero en la primera copia impresa de Stata de la parte superior: “demasiado” fuerte con respecto a una prueba de 5%. Dado que el IC de 95% de la diferencia casi toca cero, el valor de probabilidad de una prueba de hipótesis *de doble cola* efectivamente es de 0.05 (o cercano a ello), como sería de esperar dada la equivalencia existente entre una prueba de hipótesis de doble cola y un IC para una diferencia entre medias que no incluya a cero. Pero el valor de probabilidad de una prueba de hipótesis de una sola cola, en el listado de resultados que mostramos arriba, es de apenas 0.025 (0.0249 en realidad), de modo que tenemos grados de libertad de sobra si todo lo que deseamos es una prueba de 5%. Dado que el valor *t* de una prueba de hipótesis de una cola de 5% es de 1.67 (o por ahí para una gran *n*), en tanto que lo que se necesita para una de dos colas es de alrededor de 1.96, podríamos hacer que la muestra sea más pequeña en una razón de aproximadamente $\sqrt{1.67/1.96} = 0.73$.

En efecto, sólo estamos pidiendo que

$$|\bar{X}_1 - \bar{X}_2| > 1.41 t_{.05} SE$$

para una prueba *t* de una cola, siendo $t \approx 1.67$ con cualquier *n* razonablemente alta.

La siguiente prueba instantánea Stata demuestra que cuando el tamaño de la muestra es reducido del primer grupo de resultados a una razón de 0.73 de 34, o 25, entonces la prueba de la hipótesis de una cola tiene un valor de probabilidad apenas menor a 0.05, según se necesite (resaltado inferior). Los IC ahora se superponen plenamente (resaltados superiores). El IC de 95% para la diferencia incluso se superpone con cero, porque requerir un IC para la diferencia que no lo hiciera sería equivalente a una prueba de hipótesis de dos colas.

```
ttesti 25 20 29 25 34 29, desigual
```

Prueba t de dos muestras con varianzas desiguales

	Obs	Media	Err. Est.	Desv. Est.	[Inter. de Conf. 95%]	
x	25	20	5.8	29	8.029388	31.97061
y	25	34	5.8	29	22.02939	45.97061
combinado	50	27	4.180518	29.56073	18.59893	35.40107
dif		-14	8.202439		-30.49211	2.492108
dif = media(x) - media(y)				t =	-1.7068	
Ho: dif = 0				Grados de libertad de Satterthwaite =	48	
Ha: diff < 0		Ha: dif != 0		Ha: dif > 0		
Pr(T < t) = 0.0472		Pr(T > t) = 0.0943		Pr(T > t) = 0.9528		

Tomando ambos factores juntos, el tamaño de la muestra necesaria para una prueba de hipótesis de una cola es de alrededor de 0.36 de lo que se necesita para crear IC no superpuestos (de dos colas) en ambas medias.

Nótese que si la SD es efectivamente aumentada por un DEFT de 2.44 (el resultado de los mismos supuestos usados al establecer el tamaño de la muestra de 409 para un IC, a saber un ICC de 0.45 y un tamaño de conglomerado de 12), entonces el tamaño de la muestra necesario para una prueba de 5% sube, esencialmente a 2.44^2 veces 25, o 148.

ttesti 148 20 70.7 148 34 70.7, desigual

Prueba t de dos muestras con varianzas desiguales

	Obs	Media	Err. Est.	Desv. Est.	[Inter. de Conf. 95%]	
x	148	20	5.811504	70.7	8.515112	31.48489
y	148	34	5.811504	70.7	22.51511	45.48489
combinado	296	27	4.122578	70.92751	18.88661	35.11339
dif		-14	8.218708		-30.17496	2.174957
dif = media(x) - media(y)					t =	-1.7034
Ho: dif = 0		Grados de libertad de Satterthwaite =		294		
Ha: dif < 0		Ha: dif != 0		Ha: dif > 0		
Pr(T < t) = 0.0448		Pr(T > t) = 0.0895		Pr(T > t) = 0.9552		

Estos factores permiten cierta economía en el tamaño de la muestra con una prueba de hipótesis de una cola, por oposición a intervalos de confianza no superpuestos. Sin embargo, hay una presión opuesta, a saber la necesidad de tomar en cuenta la potencia. Tomándola en cuenta, asumiendo una potencia de 0.8 y una prueba de hipótesis de 5%, e introduciendo la noción de que las SD *podrían* ser distintas, un tamaño de la muestra para una hipótesis de una cola sería:

$$n = \frac{(SD_1^2 DEFF + SD_2^2 DEFF)(0.85 + 1.67)}{DIFF^2},$$

donde:

- 0.85 es el valor t de una cola para una potencia de 0.8,
- 1.67 es el valor t de una cola para una prueba de 5% (ambos con 60 grados de libertad, una cifra apropiadamente baja), y
- DIFF es la diferencia planteada como hipótesis entre, digamos, grados.

Usando los mismos parámetros que para el intervalo de confianza, a saber un DEFF de 5.595 (DEFT de 2.44) (debido a un ICC de 0.45 y un tamaño de conglomerados de 12), y una SD de 29 (lo que significa que en este ejemplo resultan ser las mismas, pero usando la ecuación que permite tener distintas SD) y un DIFF de 14, el tamaño requerido de la muestra es de 324. En el caso más pesimista en que las SD son 50 pero se permite que el DIFF sea 20, el tamaño necesario de la muestra es de 472. En ambos casos son algo más pequeños de lo necesario para un intervalo de confianza de 95%.

Si fuéramos a concluir en base a los tipos de exámenes expuestos anteriormente que las pruebas de dos colas son más apropiadas, entonces la ecuación correcta sería:

$$n = \frac{(SD_1^2 DEFF + SD_2^2 DEFF)(0.85 + 2)}{DIFF^2}$$

En dicho caso y usando los mismos supuestos ya expuestos anteriormente, el tamaño de la muestra con una SD de 29 es de 414, y sería 603 con la SD más pesimista de 50, pero con un DIFF de 20.

Para resumir:

Si se desea establecer intervalos de confianza no superpuestos, entonces con parámetros similares a los que se encuentran en los escenarios casi del peor tipo, pero con una SD de 29 (promedio entre todos los estudiados), bastará con un tamaño de muestra de 409 alumnos por grado.

En los casos en donde se sospeche que la SD llega tan alto como 50, pero donde las diferencias entre grados son correspondientemente más altas, se necesitan tamaños de muestras de hasta 597 alumnos por grado.

Si el objetivo es brindar pruebas de hipótesis de una cola (dos colas), entonces para cada grado podrían ser necesarios tamaños de muestra de 324 (414 para las de dos colas) a 472 (604 para las de dos colas).

En cuanto a las diferencias de género, ningún tamaño de muestra razonable logra detectar diferencias estadísticamente significativas, al menos a juzgar por las diferencias de género hasta ahora observadas. Las diferencias de género sencillamente son demasiado pequeñas.

Resumen del tamaño de las muestras en base a intervalos de confianza y pruebas de hipótesis

La tabla B3 resume una serie de sugerencias sobre el tamaño de las muestras. El cuadro asume una SD de 29, un ICC de 0.45 (que se halla en el extremo elevado de lo que los estudios EGRA han hallado hasta hoy), y un *clustersize* (el número de niños muestreados por centro educativo) de 10. En el caso de las pruebas de hipótesis se asume una potencia de 0.8. En cada caso se deriva el número de centros educativos necesarios redondeando el resultado de dividir el tamaño de la muestra por 10.

Tabla B3. Resumen del tamaños de las muestras según diversas consideraciones

	Tamaño de la muestra	No. de centros educativos
Nivel de confianza de 90%		
Enfoque del intervalo de confianza:		
Amplitud de dos colas del intervalo: 10	475	48
Amplitud de dos colas del intervalo: 15	211	22
Enfoque de prueba de hipótesis – una cola:		
Diferencia mínima detectable: 10	390	39
Diferencia mínima detectable: 15	173	18

	Tamaño de la muestra	No. de centros educativos
Enfoque de prueba de hipótesis – dos-colas:		
Diferencia mínima detectable: 10	539	54
Diferencia mínima detectable: 15	239	24
Nivel de confianza de 95%		
Enfoque del intervalo de confianza:		
Amplitud de dos colas del intervalo: 10	680	68
Amplitud de dos colas del intervalo: 15	303	31
Enfoque de prueba de hipótesis – una cola:		
Diferencia mínima detectable: 10	539	54
Diferencia mínima detectable: 15	239	24
Enfoque de prueba de hipótesis – una cola:		
Diferencia mínima detectable: 10	689	69
Diferencia mínima detectable: 15	306	31

Fuente: Cálculos de los autores.

Pesos del muestreo y otras consideraciones introducidas por la clustering

Resumiendo primero a partir del examen efectuado anteriormente, resulta difícil construir y llevar a cabo un simple método de muestreo aleatorio—como una muestra aleatoria simple de niños—ya que pocos países en vías de desarrollo (ninguno que los autores conozcan) cuenta con listas de nombres de niños a nivel nacional. Incluso de existir tales listas no sería práctico, puesto que sería sumamente costoso evaluar un niño en un centro educativo y luego viajar una gran distancia para evaluar otro niño en otro colegio. En lugar de ello se puede primero muestrear centros educativos y luego muestrear a los alumnos dentro de ellos.

Además y como ya se describió, en lugar de muestrear a 400 alumnos de modo completamente aleatorio, lo que en un país grande podría significar ir a 400 o casi 400 colegios, uno primero muestrea digamos 40 centros educativos y luego a 10 educandos por colegio en el(los) grado(s) de interés. Elegir un número fijo de educandos por centro educativo podría mantener las cosas simples (y en algunas metodologías podría ser necesario: véase en una sección posterior el Muestreo de Aceptación de Lotes (*Lot Quality Assurance Sampling*, LQAS), pero dado que los colegios varían en tamaño, calcular un promedio simple podría resultar engañoso en el caso que a los más grandes o más pequeños de ellos les vaya sistemáticamente mejor o peor.

Un ejemplo artificioso para aclarar este punto: supóngase que la mitad de los centros educativos son grandes (50 alumnos por grado de interés) y la mitad pequeños (25 alumnos por grado de interés). En los primeros, los educandos tienen una fluidez de lectura de 20, pero en los pequeños es de 50. Un promedio *simple* arrojará una fluidez de 37.5. Un promedio *ponderado* dará una fluidez de 30. Esto se debe a que si bien los centros educativos más pequeños tienen una fluidez más alta, pesan menos; la media cae hacia la media de los colegios grandes. Ninguna de estas medias es inherentemente “correcta”. La de 37.5 caracteriza centros educativos y podría usarse para focalizar aquellos en donde se deben efectuar intervenciones; la media de 30 caracteriza en cambio *niños*.

Si el número de niños seleccionados por centro educativo varía según el tamaño del colegio (digamos, si uno escoge cada 10° niño en el grado de interés), entonces no habrá necesidad alguna de ponderar los resultados puesto que los colegios grandes quedarán representados en

una precisa proporción inversa a su tamaño, y la media simple calculada representará a los alumnos. En este caso, si los investigadores desean tener un promedio de los niveles de lectura de los centros educativos tendrán que calcular el porcentaje de lectura en cada colegio, y luego calcular la media simple de dichos porcentajes.²⁶

Una cuestión final a la que hay que volver es que como ya se indicó, si el proceso de muestreo usa conglomerados, y si al calcularse los intervalos de confianza se ignora el hecho que se ha usado el *clustering*, entonces la variabilidad total quedará subestimada y los intervalos de confianza resultarán más angostos de lo que realmente deberían ser.

Diversos paquetes estadísticos pueden tomar en cuenta esta cuestión. Para hacerlo, se les debe indicar cuál es el criterio de *clustering* (digamos, centros educativos) y luego decirsele que use un supuesto de *clustering* al calcular los intervalos de confianza; finalmente se le debe ordenar que calcule usando este supuesto. Por ejemplo, en Stata uno declara que el criterio de *clustering* sea el número de colegios dando el comando “svyset: school”, y solicitando luego un intervalo de confianza para, digamos, la fluidez vía el comando “svy: mean fluency”. Para dar ejemplos usando países en donde sí se han efectuado estudios de EGRA, las desviaciones estándar de la fluidez media—asumiendo que no hay *clustering* y que sí la hay, respectivamente—son de 3.6 y 4.4 en un caso, y de 0.57 y 1.31 en el otro.

Técnicas no paramétricas de monitoreo del progreso a nivel del centro educativo: el enfoque LQAS

La mayoría de las aplicaciones de EGRA hasta ahora han estado dirigidas a producir una conciencia a *nivel nacional* de que los niños podrían no están leyendo tan bien como deberían, y que la fase fundacional en los sistemas educativos es excesivamente endeble, lo que hace que toda la empresa resulte débil. Estos esfuerzos buscan asimismo motivar a los responsables e implementadores de políticas para que presten atención a esta cuestión y establezcan procesos de recuperación.

Si los países u organizaciones no gubernamentales (ONG) hacen un seguimiento e intentan efectuar una recuperación y mejoras, como algunos ya lo están haciendo, entonces saldrán a primer plano consideraciones del monitoreo del progreso y no sólo de conciencia de las políticas. Para una amplia motivación y análisis, las muestras a nivel nacional de los tamaños examinados en este documento son tanto suficientes como posibles, y permiten contar con intervalos de confianza razonablemente estrechos.

De otro lado, el monitoreo a lo largo del tiempo, usualmente con mediciones repetidas, requiere de un enfoque diferente a nivel escolar.²⁷ Hasta ahora las aplicaciones de EGRA han usado un rango de 5 a 20 alumnos por grado por centro educativo. Las muestras a nivel de colegios que tienen más de unos 20 alumnos por grado son muy costosas, en particular si uno va a monitorear muchos centros educativos. Sin embargo, 20 alumnos por grado no bastan para la mayoría de los estimadores estadísticos tradicionales de los parámetros, como la fluidez de lectura promedio a nivel del colegio.

Hay dos soluciones a este problema. En primer lugar, en la modalidad de monitoreo, EGRA podría ser aplicado por los mismos docentes. En otras palabras, a nivel de los centros

²⁶ La tabla B7 al final de este anexo muestra un ejemplo de estas cuestiones.

²⁷ Ello naturalmente requiere de un enfoque distinto en lo que toca a los instrumentos empleados, y asume que se ha diseñado una intervención. Estas cuestiones son importantes pero no guardan relación con el muestreo.

educativos, cuando los profesores llevan a cabo el monitoreo evaluarán a todos los niños, de modo tal que la cuestión del muestreo no estará en discusión. Hacer que los docentes efectúen el monitoreo reduce los costes económicos (aunque no necesariamente el coste de oportunidad o social) del monitoreo de los niños.

En segundo lugar, si se desea monitorear a los niños mediante una agencia externa, entonces sí surgirá la cuestión del muestreo, puesto que ese enfoque tiene costes en efectivo y no sólo de oportunidad. Como ya se indicó, muestrear a más de unos 20 niños por grado por colegio, mediante personas de fuera, sobre la base de un monitoreo, es algo que se hace extremadamente costoso. Pero como también se indicó ya, 20 niños por grado no es una muestra lo suficientemente grande como para estar muy cómodo cuando se está usando la muestra para determinar los niveles promedio de lectura en un centro educativo dado. Sin embargo, si uno está dispuesto a pensar de modo no paramétrico, un tamaño de muestra de 20 alumnos o no mucho más, es suficientemente bueno para monitorear cualquier unidad administrativa (centro educativo o distrito).

En primer lugar podría ser bueno explicar la terminología. Por *enfoque paramétrico* queremos decir un enfoque que tiene como base el intento de estimar, digamos, las palabras correctas promedio por minuto en el colegio o distrito (un parámetro) mediante un promedio de muestra (un estimador).

Hay varios problemas relacionados con el enfoque paramétrico cuando se le aplica a nivel de los colegios. Un problema importante con este enfoque es la falta de potencia con muestras de tamaño pequeño. Usando los datos que este documento ha venido manipulando, advertimos que el “promedio de promedios” entre países para las palabras leídas correctamente por minuto en párrafos de 2º grado y en los países bajo consideración, es 25 y la SD 29. El DEFT no resulta relevante puesto que a nivel escolar uno no haría conglomerados. Supóngase que con un promedio de 25, uno tiene la meta de llegar a 39 (para llevar adelante la diferencia entre grados promedio de 14) palabras leídas correctamente por minuto, esto es ganar un grado, y uno necesita detectar centros educativos con un desempeño inferior a este nivel, para saber cuáles deberían recibir ayuda.

Incluso con una SD tan alta como 29, la hipótesis de que “Un centro educativo con un tamaño de muestra de 20 alumnos cae por debajo de la meta” tiene un alto nivel de significación tradicional (*alfa*), con un valor de probabilidad menor a 0.02. Sin embargo, esta prueba tiene una potencia de 0.60, o una *beta* de sólo alrededor de 0.31.²⁸ Esto significa que si bien hay una baja probabilidad—de menos del 5%—de ayudar a colegios que no lo necesitan (esto es, una baja probabilidad de aceptar la hipótesis alternativa cuando la hipótesis nula es correcta), sí hay una alta probabilidad de no ayudar a los que sí lo necesitan (esto es, de aceptar la hipótesis nula cuando la alternativa es correcta).

Dejando de lado estos factores, podría ser algo más difícil trabajar con los resultados paramétricos desde la perspectiva del monitoreo. Para un equipo EGRA interesado en monitorear, podría resultarle más fácil trabajar con una simple regla no paramétrica como: “muestree 20 niños por grado por centro educativo [podría ser para sólo un grado: el grado de interés en un proceso de monitoreo], y si más de X pueden leer, el colegio no necesita ayuda o no requiere el nivel más alto de ayuda”. El enunciado se presenta en cifras absolutas y no hay promedios, desviaciones estándar o porcentajes que calcular: esta es una propuesta de monitoreo más típica y simple.

²⁸ Comando de Stata “`sampsi 25 39, sd(29) n(20) onesample onesided`”.

Por ende, una alternativa a la estadística paramétrica es alguna forma de enfoque no paramétrico, donde a uno no le interesan los promedios a nivel de los centros educativos, sino únicamente el número de niños que leen en o por encima de cierto nivel en una muestra dada de tamaño fijo (o, para el caso, puesto que alguien podría también querer monitorear procesos, el número de niños que quedan expuestos a ciertas técnicas de enseñanza). Estos enfoques típicamente utilizan la distribución binomial para fijar tamaños de muestras y “reglas de decisión” (como cuántos no-lectores pueden aceptarse antes de decidir que un centro educativo no está cumpliendo las metas), para así mantener las probabilidades alfa y beta bajas. La técnica más común de éstas, más usada hasta ahora en el sector salud de los países de bajos ingresos, es el LQAS.

Para usar LQAS, uno tiene que convertir una variable de lectura dada en una variable binaria. La variable toma el valor 1 si los educandos leen por encima de cierto nivel y el valor 0 si lo hacen debajo del nivel objetivo de palabras leídas correctamente por minuto. LQAS se usa a menudo para monitorear procesos. Alternativamente se podría asignar un valor de 1 si los educandos están siendo expuestos a cierta práctica y un valor de 0 en caso contrario. Esto último no requiere convertir una variable continua en otra binaria.

El enfoque más común de LQAS es el de pensar en términos de dos riesgos:

1. el riesgo (coste) de ayudar a un centro educativo que no lo requiere, denominado comúnmente “riesgo de gobierno” o “riesgo de proveedor”, puesto que la intervención innecesaria tiene un costo, similar al error de tipo I en estadística, con el nivel alfa asociado, y
2. el riesgo de no intervenir en un colegio que sí lo necesita, comúnmente conocido como el “riesgo social” o riesgo del “consumidor”, similar a un error de tipo II en estadística, con un nivel beta asociado.

Para minimizar el riesgo al proveedor, uno querrá encontrar un alto nivel de no-lectores en un colegio antes de ayudarlo. De ese modo se puede estar muy seguro de que éste lo necesita. Pero si uno va demasiado lejos en esta dirección, también se corre el riesgo de *no* ayudar a un colegio que sí lo requiere.

Para cuidarse de ambos riesgos, usualmente se fija un umbral *superior* (por ejemplo al 50%) y el número de no-lectores permitido en una muestra se fija de modo tal que reduzca las probabilidades de clasificar como no-lector a un colegio que en realidad sí está leyendo. Esto argumenta a favor de tolerar más—y no menos—no-lectores en la muestra antes de declarar al colegio como que no cumple la meta. Desde el extremo opuesto se establece un umbral *inferior* (por ejemplo a 20%) y se fija el número de no-lectores permitidos en una muestra, de modo tal que se reduce la probabilidad de que uno clasifique al colegio como no necesitado de ayuda si sólo el 20% o más de los niños no están leyendo. Esto argumenta a favor de tolerar menos—y no más—no-lectores en la muestra. En suma, uno tiene presión en ambas direcciones. Dadas estas presiones opuestas, el paso final es determinar el tamaño total de la muestra y el número de no-lectores a tolerarse antes de que el centro educativo reciba ayuda, usando dichos umbrales, para así contener el riesgo *total* (alfa más beta) a un nivel razonable, como menos de 15%. En el sector salud por lo general se considera que este nivel es apropiado para los fines del monitoreo.

Para repetir, la técnica puede usarse con variables continuas (digamos, la fluidez en la lectura de un párrafo) que han sido convertidas en binarias (todo niño leyendo a más de X palabras leídas correctamente por minuto recibe 1, otros reciben 0), o con variables de monitoreo que son naturalmente binarias (por ejemplo, si un niño participa en lectura con sus iguales todos los días).

La tabla B4 ilustra los principios involucrados para umbrales de 50% y 20%. La bibliografía del sector salud suele emplear ejemplos de 80% y 50% en, por ejemplo, el monitoreo del empleo de vacunas. Pero si va a usarse esta técnica para monitorear la lectura, resulta sensato fijar umbrales bastante bajos dado cuán débil la situación actual parece ser en la mayoría de los países estudiados.

Tabla B4. Ejemplo de un cuadro de decisión de LQAS

Tamaño de la muestra	No. de fracasos, f	No de éxitos, s	Probabilidad acumulativa de, hasta e incluyendo $s-1$ éxitos	Probabilidad acumulativa de, hasta e incluyendo f fracasos	Riesgo total
			Equivalente de la probabilidad o riesgo de aceptar que un colegio necesita ayuda cuando no es así	Equivalente de la probabilidad o riesgo de no ayudar a un colegio que necesita ayuda	
			Umbral superior: 0.50 (50%)	Umbral inferior: 0.20 (20%)	
20	15	5	0.006	0.370	0.376
20	14	6	0.021	0.196	0.216
20	13	7	0.058	0.087	0.144
20	12	8	0.132	0.032	0.164
20	11	9	0.252	0.010	0.262
20	10	10	0.412	0.003	0.414
20	9	11	0.588	0.001	0.589
20	8	12	0.748	0.000	0.748
20	7	13	0.868	0.000	0.868
20	6	14	0.942	0.000	0.942
20	5	15	0.979	0.000	0.979
20	4	16	0.994	0.000	0.994
20	3	17	0.999	0.000	0.999
20	2	18	1.000	0.000	1.000
20	1	19	1.000	0.000	1.000
20	0	20	1.000	0.000	1.000

Fuente: Calculado por los autores a partir de valores asumidos usando la distribución binomial.

Leyendo ***hacia arriba*** a lo largo de la columna del umbral de 50%, advertimos que para reducir el riesgo gubernamental queremos aceptar como colegios que no necesitan una mayor ayuda sólo a aquellos que estamos seguros tienen más del 50% de lectores. Uno podría estar así tentado de aceptar como colegios que no necesitan ayuda sólo a aquellos donde hay al menos 10 lectores (puesto que $10/20=50\%$). Pero incluso aquellos centros educativos donde hasta el 50% de los niños leen tienen un probabilidad tan alta como 13.2% de producir muestras en donde sólo hasta e inclusive 8 niños leen, puesto que estamos hablando únicamente de una

muestra. Por dicha razón una regla basada en el promedio esperado estará errada la mayor parte del tiempo.

Uno podría estar tentado de reducir este riesgo aún más y realmente sólo aceptar como necesitados de ayuda a los centros educativos con, digamos, 6 lectores. Esto reduciría el riesgo del gobierno a 2.1%. De este modo, para minimizar el riesgo de aceptar demasiados colegios como necesitados de ayuda (ahorrando así el riesgo del gobierno), podrían aceptarse montones de no-lectores antes de reconocerse que el centro educativo necesita asistencia. Cuantos más educandos que no leen uno esté dispuesto a tolerar antes de decidir que el colegio requiere ayuda, tanto menor será la probabilidad de aceptar que éste la necesita aun si no la requiere; en otras palabras, tanto más seguro estará uno, en base a la muestra, de que hay suficientes no-lectores en la población escolar como para volcar al colegio por debajo del umbral de 50%. Podemos ver esto leyendo hacia arriba por la columna del umbral del 50%. Esto reduce el riesgo del gobierno, o el riesgo de ayudar a centros educativos que no lo necesitan.

Pero entonces (leyendo a lo largo del panel a la columna 5, la del umbral de 20%), descender hasta un punto tan bajo como 6 lectores sería incurrir en una probabilidad de 19.6% de *no* intervenir en colegios donde menos del 20% de los alumnos están alcanzando la meta. Esto es porque hay una probabilidad de 19.6%, incluso si sólo 20% de los educandos son lectores, de observar hasta 6 lectores o más de 20. Este es un nivel demasiado alto de “riesgo del consumidor”. Corremos el riesgo de no aceptar colegios que lo necesitan.

El truco, entonces, es sumar los dos riesgos, crear un concepto de riesgo total y examinar el patrón de este último. Así, con un tamaño de muestra de 20, el número de umbral de lectores—esto es, el número de lectores que minimiza el riesgo total—será 7. La “regla de decisión” es: aceptar un centro educativo como que no necesita ayuda si hay 7 o más lectores, o aceptar que la necesita si hay menos de 7 lectores, lo que crea un riesgo total de 0.144. Recuérdese que en un enfoque paramétrico, sólo el segundo tipo de riesgo era tan alto como 30%. En este caso, tanto los riesgos alfa como beta son mantenidos por debajo de 10% y la suma de ambos por debajo de 15%. Si uno deseara llevar este riesgo por debajo de 10%, para un conjunto de umbrales de 50%–20%, el tamaño de la muestra tendría que subir hasta 26, un precio bastante alto que pagar. La mayoría de las aplicaciones de salud usan 19.

Todo lo que se ha dicho de los niños dentro de los centros educativos es por ejemplo aplicable a los colegios dentro de los distritos. De igual modo, todo lo que se ha dicho sobre los niños leyendo podría ser aplicado a alguna otra práctica de nivel escolar, por ejemplo si el colegio tiene un programa de lectura de algún tipo. Se aplican los mismos tipos de números.

Sería difícil explicar estas reglas y asegurarse de que los trabajadores a nivel de distrito en países de bajos ingresos realmente puedan entenderlas, pero son extraordinariamente simples de aplicar sobre una base de reglas simples, una vez que se han establecido las normas básicas. Es simplemente cuestión de decir: “Tome muestras de 20 niños. Si 7 o más están leyendo a más de X palabras leídas correctamente por minuto, el colegio no recibe ayuda intensiva. Si menos de 7 lo están haciendo, el colegio necesita más ayuda, o ser advertido”.

El cálculo de las probabilidades se puede programar muy fácilmente. La fórmula necesaria en Excel para el primer riesgo (umbral de 50%) es

$=1-BINOMDIST(f,n,1-hithresh, TRUE),$

donde:

n es el tamaño de la muestra,
 f es el número de fracasos,
 $hithresh$ es el umbral más alto que protege del riesgo del gobierno, y
 $TRUE$ quiere decir que es el binomio acumulativo deseado.

La fórmula del segundo riesgo es

$$=BINOMDIST(f,n,1-lothresh,TRUE).$$

Así, por ejemplo, para producir el 2.1% ligado al riesgo de gobierno (columna de 50%) arriba, cuando se tolera más de 6 no-lectores (segunda hilera desde arriba) antes que se ayude a un colegio, los investigadores usarían $1-BINOMDIST(14,20,1-0.5,TRUE)$.²⁹

Como ya se indicó, una de las virtudes de LQAS es que las cosas pueden enunciarse en términos de reglas de decisión muy simples y fijas. En la tabla B5 hemos calculado un ejemplo que muestra el tamaño óptimo de la muestra y la regla de decisión (v.g., aceptar como que no necesita ayuda a un centro educativo con 7 o más lectores). Esto puede ser usado directamente por los monitores sin preocuparse en absoluto por la teoría y el cuadro más complicado mostrado anteriormente (ver la tabla B4), como lo hacen los proyectos de salud.

Tabla B5. Reglas de decisión preprogramadas en base a la metodología LQAS

Cumplimiento o umbrales de desempeño		Para un riesgo total < 0.10		Para un riesgo total < 0.15	
		Tamaño óptimo de la muestra	Regla de decisión	Tamaño óptimo de la muestra	Regla de decisión
Superior	Inferior				
0.95	0.65	15	13	13	11
0.90	0.60	20	16	15	12
0.85	0.55	23	17	18	13
0.80	0.50	25	17	20	14
0.75	0.45	27	17	21	13
0.70	0.40	28	16	22	13
0.65	0.35	29	15	23	12
0.60	0.30	29	13	24	12
0.55	0.25	27	11	21	9
0.50	0.20	26	9	20	7
0.45	0.15	23	7	18	6
0.40	0.10	20	5	15	4
0.35	0.05	15	3	13	3

Fuente: Calculado por los autores a partir de valores asumidos usando la distribución binomial.

²⁹ En realidad uno usaría las referencias extraídas de celdas de Excel en lugar de cifras exactas.

Combinando y comparando LQAS y el estimado de parámetros

La bibliografía sugiere que hay una forma inteligente de combinar LQAS y la estimación de parámetros. Para hacerlo se reúnen datos sobre la lectura a nivel de los colegios. Pero sabemos que un tamaño de muestra de hasta 20 niños por grado (o por todo el colegio, cuando uno está monitoreando sólo un grado, y anotando que esta cifra en todo caso es más alta de lo óptimo en un enfoque de conglomerados) no permite ni evaluar el porcentaje de los lectores como parámetro, ni tampoco un estimado de las palabras leídas correctamente por minuto, también como parámetro, sin mucha confianza.

Como ya se indicó, el intervalo de confianza en nuestro “caso [más] típico”, con un promedio de palabras leídas correctamente por minuto de 25, una SD de 29 y un tamaño de muestra tan alto como 20 a nivel del centro educativo individual, arroja un intervalo de confianza de 95% tan amplio como [11.4, 38.6] a nivel del colegio, demasiado amplio como para trabajar con facilidad. Aceptar un nivel de confianza inferior de 90% no ayuda mucho pues produce un IC de [13.8, 35.2] a nivel del centro educativo individual. Incluso el estimado del porcentaje de lectores que leen por encima de cierto nivel produce un IC muy amplio. En los países que hemos estudiado, una media de 25 corresponde a alrededor del 20% de los estudiantes que leen por encima de las 40 palabras leídas correctamente por minuto. Un IC de 905 para el porcentaje de la población es [0.071, 0.401] a nivel del centro educativo.

Sin embargo, la bibliografía sugiere que tomar muestras de 20 alumnos por colegio, y luego tal vez 20 a 40 centros educativos, haría que el tamaño de la muestra sea lo suficientemente grande como para permitir la estimación de parámetros a niveles más altos que el colegio. Esto corresponde a los cálculos del tamaño de la muestra por encima de lo necesario para la estimación *paramétrica* de ya sea los intervalos de confianza o las pruebas de hipótesis, *por encima* del nivel del centro educativo.

Esto puede ser bastante útil. Uno puede efectuar una encuesta de tipo LQAS para reunir los datos de parámetros de línea de base a un nivel más alto que el centro educativo. El tamaño de la muestra o de conglomerado a nivel del colegio (digamos 20) para una encuesta inicial, puede basarse en lo que ya sabemos sobre la lectura en el país, y en evidencias procedentes de alrededor del mundo. Los datos pueden entonces usarse para tomar decisiones no paramétricas de “sí-no” acerca de *cada* colegio, para establecer cuáles de ellos necesitan más ayuda en el primer año de una intervención, y para trazar una línea de base para mejorar al centro educativo en los parámetros mismos. Medidas repetidas pueden hacer la misma cosa, excepto que uno puede ser capaz de optimizar ligeramente el tamaño de la muestra interna al colegio, una vez que hay una línea de base a partir de la cual establecer los umbrales superior e inferior. Evidentemente, si la línea de base del tamaño de la muestra del colegio se fija en 20, una consideración de DEFT impulsará entonces el número total de centros educativos que se debe permitir.

Las recomendaciones de un ejemplo de línea de base de LQAS aparecen en la tabla B6 que presentamos debajo. A diferencia del caso de los intervalos de confianza y pruebas de hipótesis estándares, la muestra interna al colegio o tamaño del conglomerado es fijada en 20. Esto es porque éste es el mínimo que puede aceptarse para hacer juicios no paramétricos acerca de cada colegio (el centro educativo “aprueba” la inspección o no). Pero entonces uno puede usar la información a nivel del colegio y aglomerarla para establecer el nivel promedio (paramétrico) de, digamos, la fluidez. El supuesto que aparece debajo es que el tamaño de la muestra interna al colegio, o el tamaño del conglomerado, es fijado en 20.

Tabla B6. Número de centros educativos necesarios para brindar un estimado de parámetros por encima de los colegios, al usar un tamaño de conglomerado LQAS de 20 a nivel de los centros educativos

Amplitud de confianza de dos colas	10	10	15	15
Nivel de confianza	90%	95%	90%	95%
Número de centros educativos necesarios	45	65	20	28

Fuente: Cálculos de los autores.

Al aglomerar hacia arriba los datos de LQAS para estimar los parámetros, es importante ponderar los centros educativos elegidos, como era el caso anterior con los métodos estándares de estimación de parámetros. Supongamos por ejemplo que los colegios de un distrito fueron elegidos en forma íntegramente aleatoria, usando una técnica simple de selección al azar. Pero ellos difieren en su tamaño. De este modo los 20 alumnos elegidos dentro de cada colegio representan totales distintos. En el ejemplo artificial en la tabla B7, se muestra un promedio simple y también otro ponderado, asumiendo una muestra pequeña de 20 centros educativos y 20 alumnos por colegio.

Tabla B7. Usando datos obtenidos con LQAS para calcular promedios

Colegio no.	Alumnos elegidos	Número de lectores	Total de alumnos en grado evaluado	Lectores ponderados
1	20	12	50	600
2	20	18	49	882
3	20	7	23	161
4	20	8	25	200
5	20	13	46	598
6	20	11	47	517
7	20	18	48	864
8	20	20	63	1260
9	20	5	17	85
10	20	3	15	45
11	20	6	23	138
12	20	7	19	133
13	20	20	46	920
14	20	19	39	741
15	20	17	48	816
16	20	3	18	54
17	20	4	17	68
18	20	2	26	52
19	20	5	23	115
20	20	8	35	280
Total	400	206	677	8529
Media simple		10.3		
Media ponderada				12.6

Fuente: Cálculos de los autores.

En esta tabla se calcula la media simple tomando el número total de lectores (206) y dividiéndola por el número total de colegios (20), para obtener un número promedio de lectores por centro educativo de 10.3 (o un porcentaje de lectura de $10.3 \cdot 100 / 20 = 51.5\%$). Sin embargo, podemos advertir un patrón en los datos: en cada conglomerado de tamaño fijo o muestra-colegio de 20, los colegios más grandes tienen más lectores que los pequeños. Por lo tanto, una media simple subestimaré la proporción de niños que pueden leer. Si uno pondera los lectores de cada centro educativo según el número total de alumnos en el grado evaluado, la proporción de lectura será mucho mayor. Esto se puede hacer multiplicando el número de lectores en cada colegio por el número total de alumnos en el grado evaluado de cada centro educativo, para llegar así a la última columna. Uno divide entonces el total de la última columna para todos los colegios entre el total en el grado para todos los centros educativos, para obtener así el promedio ponderado ($8529/677=12.6$), o un porcentaje de lectura de 63%.

La diferencia entre 51.5% y 63% es importante, pero este es un caso bastante exagerado. La diferencia no será tan grande como esta en la mayoría de los ejemplos de la vida real, porque no es probable que haya una asociación tan fuerte entre el tamaño de los centros educativos y la habilidad lectora (0.87 en este ejemplo artificial). Adviértase, además, que el porcentaje de 51.5% no necesariamente está "errado" y que el de 63% no necesariamente es "correcto". El primero representa el porcentaje de alumnos que pueden leer en un centro educativo típico, y esta es una idea válida si la unidad de análisis es el colegio: representa el resultado del centro educativo promedio, en términos de cuánto el colegio promedio "produce" lectura. El segundo representa el porcentaje de alumnos que pueden leer en la población como un todo, y esta también es una idea válida si la unidad de análisis es el alumno.

Anexo C. Evaluando la calidad técnica del instrumento EGRA

Es importante evaluar la calidad técnica de todo instrumento usado para medir el desempeño de los educandos. El instrumento EGRA no es la excepción. Los procedimientos usados para llevar a cabo estas revisiones provienen del campo de la psicometría. Estos procedimientos tradicionalmente se concentraron en dos conceptos claves: confiabilidad y validez.

Se recomienda vigorosamente que los equipos que dirigen el EGRA incluyan a una persona familiarizada con la psicometría y que esté en condiciones de efectuar las pruebas necesarias. El examen que aparece a continuación sólo busca brindarle al lector una breve introducción al tema y resaltar algunas de las cuestiones involucradas. No quiere ser un examen exhaustivo, ni tampoco da instrucciones paso a paso con que efectuar dichas pruebas.

Confiabilidad

Podemos definir la confiabilidad como el grado en que los puntajes de un grupo de educandos son consistentes con la aplicación repetida de una prueba. Una analogía de la vida cotidiana sería una balanza. Si colocamos en ella una bolsa de arroz cinco veces y en cada ocasión dice “20”, entonces ella arroja resultados confiables. Sin embargo, si la balanza arroja un número distinto (v.g., 19, 20, 18, 22, 16) cada vez que se pesa la bolsa, entonces muestra no ser confiable.

La medida más usada para calcular la confiabilidad o consistencia interna de una prueba es el **alfa de Cronbach**, una medida de la consistencia interna de una prueba (los paquetes estadísticos como SPSS y Stata pueden calcular este coeficiente rápidamente). Ésta, sin embargo, podría no ser la medida más apropiada de la confiabilidad de los puntajes EGRA, fundamentalmente debido a que algunas partes del instrumento EGRA están cronometradas. Las medidas cronometradas o limitadas a un lapso determinado afectan el cálculo del coeficiente alfa de modo tal que le convierten en un estimado inflado de la confiabilidad del puntaje de la prueba; sin embargo, se ignora el grado en que los puntajes quedan inflados.

El **método de *test-retest*** es más idóneo para calcular la confiabilidad de los puntajes obtenidos con el instrumento EGRA. Esta prueba, que puede efectuarse como parte del piloto del instrumento EGRA, esencialmente involucra la aplicación del instrumento al mismo grupo de educandos en dos momentos distintos (v.g., separadas aproximadamente por una semana). Los alumnos seleccionados deberían ser representativos de la población objetivo en áreas claves tales como el género y la edad, el estatus/antecedentes socioeconómicos del hogar, las habilidades cognitivas y así por el estilo. El método *test-retest* para el cálculo de la confiabilidad representa la correlación existente entre el puntaje de los educandos en dos aplicaciones de la prueba. Cuanto más alta sea la correlación (por lo general se considera aceptable un valor de 0.7 o más), tanto menos susceptibles serán los puntajes EGRA a cambios aleatorios cotidianos en la condición de los que toman la prueba o del entorno de la misma.

Una variación en esto es llevar a cabo la *test-retest* usando dos formularios similares del instrumento EGRA. En este caso el procedimiento consiste en aplicar el Formulario 1 de la prueba, esperar alrededor de una hora y luego aplicar el Formulario 2. De ser posible, es deseable que el orden de aplicación de los formularios se invierta para la mitad del grupo. La correlación (nuevamente se considera aceptable un valor de 0.7 o más) entre los dos grupos de

puntajes brinda una medida del grado de estabilidad de los puntajes EGRA a lo largo de aplicaciones repetidas, así como el grado de equivalencia de los puntajes producidos por los dos formularios del instrumento.

Otra cuestión relacionada con la confiabilidad de los puntajes de EGRA es **la consistencia y exactitud del desempeño del examinador**. Si dos examinadores están escuchando al mismo niño leer una lista de palabras de la prueba EGRA, ¿es probable que registren el mismo número de palabras como correctamente leídas? ¿Alguno de ellos estará en lo correcto? Dado que usualmente sólo un examinador escucha y anota las respuestas de cada niño en la aplicación principal de EGRA, el mejor momento para ocuparse de esta cuestión de la consistencia (y exactitud) es durante la formación de los examinadores. Se debe pedir a todos ellos que escuchen la misma grabación de un niño que toma la prueba EGRA y anoten individualmente la corrección de las respuestas del educando en términos del número de palabras leídas correctamente, etc. Debería recibir formación adicional todo examinador cuyo registro de respuestas incluya errores significativos (v.g., en términos de la discrepancia entre el número de palabras leídas correctamente por minuto registradas por él, y el número de palabras leídas correctamente por minuto por el educando). De no mejorar, se les debería retirar del grupo de examinadores para así no afectar negativamente a la calidad de los datos recogidos durante el estudio principal.

Validez

La validez se refiere a la idoneidad o la corrección de las inferencias o decisiones que tienen como base los resultados de la prueba. Para volver una vez más al ejemplo de la balanza, si colocamos cinco veces en ella una bolsa de arroz que pesa 30 Kg. y cada vez dice “30”, entonces la balanza arroja resultados que no sólo son confiables, sino también válidos. Si la balanza consistentemente dice “20” cada vez que se coloca en ella la bolsa de 30 Kg., entonces da resultados que no son válidos (pero que siguen siendo confiables, ¡porque si bien la medida está errada, es en cambio muy consistente!).

Nada viene a ser una prueba válida general. La validez de una prueba se debe establecer en relación con una inferencia o uso particulares que habrán de basarse en sus resultados. La validación es el proceso mediante el cual el diseñador o el usuario de una prueba reúnen evidencias para apoyar la inferencia/uso deseados. Se describen a continuación las evidencias de validez que son relevantes para el instrumento EGRA.

Las evidencias relacionadas con el contenido de la prueba se refieren al grado en que los ítems de la prueba EGRA son representativos del constructo que viene siendo medido (i.e., la habilidad temprana de lectura en un país particular). El taller a nivel de país celebrado al inicio del proceso de diseño de la prueba EGRA brinda una oportunidad para que los países incorporen la validez del contenido al instrumento haciendo que los funcionarios del ministerio, los expertos en currículo y otros grupos relevantes examinen la plantilla de la prueba y hagan juicios sobre la idoneidad de cada tipo de ítem, para medir las habilidades de lectura temprana de sus educandos. Tras este examen, estas personas adaptan el instrumento EGRA según sea necesario y preparan ítems apropiados al país para cada sección del mismo.

Las evidencias relacionadas con criterios se refieren a la fortaleza de la relación (correlación) existente entre los puntajes en la prueba EGRA y aquellos de otras medidas externas a ella. Esto en general involucrará el examen de la relación existente entre los puntajes EGRA y aquellos en las medidas de algunos criterios que se espera la prueba prediga (v.g., los puntajes de comprensión de lectura en grados posteriores), así como las relaciones

con otras pruebas planteadas como hipótesis para medir el mismo constructo u otros afines (v.g., el puntaje de los educandos en otras pruebas de habilidades de lectura temprana). Los datos de estas otras medidas pueden ser reunidos al mismo tiempo que los datos de EGRA, o bien en un momento posterior (pero se les debe reunir con los mismos educandos). Este tipo de evidencias de validez será difícil de reunir en países con pocas medidas estandarizadas de resultados de los alumnos. Sin embargo, vale la pena tener en mente que las extensas investigaciones efectuadas en otros países demostraron que los instrumentos de tipo EGRA muestran vigorosas relaciones (0.7 y más) con los tipos de medidas externas dados como ejemplo en este párrafo.

Algunos diseñadores de pruebas recomiendan que se reúna un tipo adicional de evidencias como parte de la validación de la prueba, a saber las **evidencias de las consecuencias que el uso del puntaje de la prueba** tiene sobre los diseñadores de la misma y otras partes interesadas. Esto involucra la recolección de datos para establecer si se están alcanzando los efectos beneficiosos deseados de la prueba (v.g., en el caso de EGRA, ellos incluyen el dar a los responsables de política educativa resultados a nivel del sistema de los niveles de habilidad en la lectura temprana, de tal modo que puedan focalizar los recursos y la formación con mayor efectividad). También involucra la recolección de evidencias de toda consecuencia negativa no buscada del uso de la prueba (v.g., castigar a los centros educativos con un desempeño deficiente en EGRA retirándoles recursos) y tomar medidas para impedir que dichos resultados adversos vuelvan a darse.

Anexo D. Carta abierta del subdirector general, África del Sur, a los directores de centros educativos

AN OPEN LETTER TO ALL PRIMARY SCHOOL PRINCIPALS

Dear Principal

Let's teach our children to read

Reading is a foundational skill that all our children need if they are to succeed in life. Sadly, all our assessments of how well our children read reveal that a shockingly high number cannot read at the appropriate grade and age level. Many simply cannot read at all. We cannot allow this to continue. We are therefore challenging all primary schools to improve the reading skills of all their learners. Even learners who are perceived to be reading at the appropriate level for the grade they are in should be encouraged to move to the next level.

Since the introduction of the National Curriculum Statement, many teachers believe they do not have to teach reading anymore. Nothing could be further from the truth. Reading is probably the single most essential skill a child needs and it should be acquired as early as possible.

There are five areas that are critical for reading:-

- o Phonemic awareness (understanding the sounds in spoken words);
- o Phonemes (linking sounds to the alphabet and combining these to form words);
- o Vocabulary (learning and using new words);
- o Fluency (reading with speed, accuracy and understanding); and
- o Comprehension (understanding the meaning of what they read).

The Curriculum Guidelines allocate sufficient time for the Language Learning Area and the above should be taught within that time allocation. The Department will provide additional support and guidelines to all schools so principals know how to support teachers.

In addition to the formal teaching of reading, we want all schools to set aside at least 30 minutes a day for the entire school to read (including principal and staff) in any language. This period can be handled in different ways: children who are able to read independently should be encouraged to do so, older learners can read to the younger ones, or teachers to their classes or to groups of children. The Department of Education will help you put storybooks into all classrooms, starting with schools in the poorest communities, so that children can also read for pleasure.

At the beginning of 2006 we put packs of 100 storybooks each, in the different South African languages, into over 5 628 foundation phase classes. Next year we will put similar packs into another 6 000 schools. All we ask is that you help learners read the books and enjoy them.

Later this year we will be testing the reading levels of Grade 3 learners all over the country. A short test to assess whether we are succeeding in our efforts to teach reading will from now on be a regular feature in our monitoring of learner achievement in schools. Take up the challenge and help your children read better!

Palesa T Tyobeka
Deputy Director-General: General Education and Training
Department of Education



education

Department:
Education
REPUBLIC OF SOUTH AFRICA

Anexo D (traducción)

UNA CARTA ABIERTA A TODOS LOS DIRECTORES DE ESCUELAS PRIMARIAS

Estimado/a Director(a)

Enseñemos a nuestros niños a leer

Leer es una habilidad básica que todos nuestros niños necesitan para tener éxito en la vida. Lamentablemente, todas nuestras evaluaciones de su nivel de lectura indican que una cantidad sorprendentemente alta no puede leer al nivel apropiado para su grado y edad, y muchos no pueden siquiera leer. No puede continuar así. Por eso, retamos a todas las escuelas primarias para que mejoren la capacidad de lectura de todos sus educandos. Aún para aquellos que leen al nivel apropiado para su grado, se les debe animar a pasar al siguiente nivel.

Desde la introducción de la Declaración Nacional del Plan de Estudios, muchos docentes creen que ya no tienen que enseñar a leer. Nada podría estar más lejos de la verdad. La lectura es la habilidad más esencial para los niños, y se debe adquirir esta capacidad lo más pronto posible.

Hay cinco áreas críticas con respecto a la lectura:

- Conciencia fonémica (entender los sonidos en las palabras habladas en voz alta);
- Fonemas (vincular los sonidos al alfabeto y combinar éstas para crear palabras);
- Vocabulario (aprender y usar nuevas palabras);
- Fluidez (leer con rapidez, exactitud, y conocimiento); y
- Comprensión (comprender el significado de lo leído).

Las Guías de Currículo asignan tiempo suficiente para dedicarle al Área de Aprendizaje de Lenguaje, y se deben enseñar los elementos arriba mencionadas en el tiempo asignado para ello. El Departamento proveerá apoyo adicional e instrucciones a todas las escuelas para que los directores entiendan cómo ayudar a los docentes.

Además de la enseñanza formal de la lectura, queremos que todas las escuelas aparten al menos 30 minutos diarios para que toda la escuela lea (inclusivo el/la director(a) y los docentes), en cualquier idioma. Este período de tiempo se puede manejar de diversas maneras: para los educandos que sí pueden leer independientemente, se debe animarles a hacerlo; los educandos mayores pueden leer a los menores; o docentes pueden leer a sus clases o a grupos de educandos. El Departamento de Educación les ayudará a poner libros de cuentos en todas las aulas, empezando con las escuelas en las comunidades más pobres, para que los niños puedan leer por placer.

A principios de 2006, pusimos paquetes de 100 libros de cuentos, y en varios idiomas sudafricanos, en más de 5628 aulas de educación primaria. El año que viene, pondremos paquetes semejantes en otras 6000 escuelas. Todo lo que pedimos es que ayude a sus educandos a leer y gozar los libros.

A finales de este año, se pondrá a prueba la capacidad de leer de los educandos en todo el país del tercer grado. De ahora en adelante, para evaluar si nuestras esfuerzos para enseñar la lectura están teniendo éxito, una breve prueba de lectura será una característica consistente de

nuestro monitoreo del rendimiento de los estudiantes en nuestras escuelas. ¡Acepte el reto y ayude a sus niños a leer mejor!

Palesa T. Tyobeka
Subdirector General: Educación Básica y Formación
Departamento de Educación



educación

Departamento:
Educación
REPUBLICA DE SUDAFRICA

Anexo E. Agenda de taller de recuperación en lectura en Kenia, en base a los resultados de EGRA

DÍA 1: Lun 8:30-9:00	ENFOQUE Presentaciones y fijar las metas del taller <ul style="list-style-type: none">▪ Presentación de los participantes (10 min.)▪ Examen global de los objetivos del taller (20 min.) (refiérase a lo anterior)<ul style="list-style-type: none">▪ Revisión de las cuestiones básicas en torno a la fluidez y el pensamiento actual acerca de la lectura, en términos más generales▪ Examen global del proyecto Early Grade Reading (EGR): diseño y evaluación de línea de base▪ Resultados de la evaluación de línea de base: examen global e implicaciones para las intervenciones de recuperación▪ Examen de las intervenciones de recuperación: Education of Marginalized Children in Kenya (EMACK, Educación de Niños Marginados en Kenia) y otras experiencias fuera de las de dicho país▪ Diseño de intervenciones de recuperación para preunidad, grado 1 y grado 2▪ Diseño de metodologías de evaluación del progreso de los educandos▪ Prueba de las intervenciones de recuperación diseñadas, y mejoras▪ Diseño de la estrategia de implementación
9:00-10:30	Cuestiones fundamentales de la fluidez Por qué EGR <ul style="list-style-type: none">▪ Conciencia fonémica, método fonológico, fluidez, vocabulario, comprensión
11:00-12:00	Proyecto EGR <ul style="list-style-type: none">▪ Proceso de diseño de la evaluación, taller en abril:<ul style="list-style-type: none">○ Revise brevemente las metas de este taller, los participantes y sus logros. (30 min.)▪ Evaluación de línea de base:<ul style="list-style-type: none">○ Examen global de la implementación: recolección, ingreso y análisis de datos (30 min.)
12:00-13:00	Finalidad de las intervenciones de recuperación (sólo una diapositiva o dos; más sobre este tema más adelante) Presentación de los resultados de la evaluación de línea de base, Parte I <ul style="list-style-type: none">▪ Desempeño estudiantil en todas las tareas, en ambas lenguas
14:00-16:00	Implicaciones de la evaluación de línea de base para las intervenciones de recuperación <ul style="list-style-type: none">▪ Perspectivas de los docentes sobre algunos de los problemas encontrados en las aulas. Sugerimos que la Aga Khan Foundation (AKF) invite a unos cuantos profesores para que presenten algunos de los desafíos que ellos enfrentan cotidianamente. Si la AKF elige formadores de docentes y no profesores, éstos deben haber sido docentes en los últimos 3–4 años.
16:30-17:30	[prosigue] Implicaciones de la evaluación de línea de base para las intervenciones de recuperación <ul style="list-style-type: none">▪ Cómo superar algunos de estos obstáculos: involucrar a los padres, brindar materiales de enseñanza, etc.
DÍA 2: Mar 8:30-10:30	ENFOQUE Examen general de las intervenciones de recuperación <ul style="list-style-type: none">▪ Finalidad de las intervenciones de recuperación▪ Experiencia Pratham, Perú, África del Sur, Guatemala, etc. Examen general del programa de formación docente de EMACK

- Relevancia de los programas de EMACK para la EGR y posible adopción de alguna práctica
- Examen de los mecanismos de suministro de formación de docentes e implicaciones para la EGR

11:00-12:00	[prosigue] Examen global de las intervenciones de recuperación
12:00-13:00	Preunidad: Diseño de las intervenciones de recuperación (Sylvia Linan-Thompson) <ul style="list-style-type: none"> ▪ Análisis del currículo actual ▪ Identificación de componentes y ámbito y secuencia
14:00-16:00	[prosigue] Preunidad: diseño de las intervenciones de recuperación
16:30-17:30	Grado 1: Diseño de las intervenciones de recuperación <ul style="list-style-type: none"> ▪ Análisis del currículo actual ▪ Identificación de componentes y ámbito y secuencia

DÍA 3: Mie ENFOQUE

8:30- 10:30	[prosigue] 1º grado: diseño de las intervenciones de recuperación
11:00-12:00	[prosigue] 1º grado: diseño de las intervenciones de recuperación
12:00-13:00	Grado 2: intervenciones de recuperación <ul style="list-style-type: none"> ▪ Análisis del currículo actual ▪ Identificación de componentes y ámbito y secuencia
14:00-16:00	[prosigue] 2º grado: intervenciones de recuperación
16:30-17:30	[prosigue] 2º grado: intervenciones de recuperación

DÍA 4: Jue ENFOQUE

8:30-10:30	Diseño de metodologías de evaluación del progreso de los educandos <ul style="list-style-type: none"> ▪ Herramientas del docente para la autoevaluación de cada grado: preunidad, 1º y 2º ▪ Herramientas para el personal de EMACK: preunidad, 1º y 2º
11:00-12:00	[prosigue] Diseño de metodologías de evaluación del progreso de los educandos
12:00-13:00	[prosigue] Diseño de metodologías de evaluación del progreso de los educandos
14:00-16:00	Estrategia de implementación
16:30-17:30	[prosigue] Estrategia de implementación <ul style="list-style-type: none"> ▪ Centros educativos de tratamiento: <ul style="list-style-type: none"> ○ Aclare con AKF qué centros educativos recibirán tratamiento. Si sólo se focalizaran en los colegios de tratamiento, debemos entonces asegurarnos que focalicen en los centros educativos de reemplazo y no en los que fueron elegidos originalmente. Si el tratamiento se focalizara en todos los colegios del distrito no habrá problema alguno, pero entonces para la evaluación post-tratamiento deberemos seleccionar centros educativos de control en algún otro lugar. ○ Emisión de recursos: por ejemplo, si sólo hay 1 libro por cada 3 niños, deben asegurarse de que haya más materiales. Esta podría ser una razón para focalizarse sólo en apenas 20 centros educativos: al menos podemos asegurar materiales para ESTOS colegios. ○ Comience ESTE AÑO con la preunidad y el 1º grado, al año siguiente agregue el 2º grado. Al año siguiente se evaluará 2º grado. ▪ Cómo formar a docentes ▪ Cómo brindarles los materiales necesarios ▪ Cómo brindarles respaldo ▪ Cómo organizar el monitoreo

DÍA 5: Vie	ENFOQUE
8:30-10:30	[prosigue] Estrategia de implementación
11:00-12:00	[prosigue] Estrategia de implementación
12:00-13:00	Cierre del taller

Anexo F. Ejemplos de planes de lecciones para la recuperación de lectura, en base a los resultados de EGRA

Grupo __1__ Lección __1__ Elementos a, s, m, l

<p>Conciencia fonológica (5 minutos)</p>	<p>Objetivo: Dados una palabra, los estudiantes identificarán sus sonidos iniciales</p> <p>Recursos: lista de palabras</p>
--	--

Lista de palabras: mamá, mas, sal, sala, las, lama, ama, al

Actividad:

Diga a los estudiantes que identificarán el primer sonido de la palabra que usted diga.

Ejemplifique la tarea. **La palabra es “ama”. El primer sonido es /a/.**

Ahora intentémoslo juntos. El primer sonido de “ama” es ... /a/.

Ahora inténtelo ustedes. El primer sonido de “al” es Los alumnos responden.

Si los estudiantes responden correctamente, pase a las siguientes palabras.

<p>Método fonológico (5 minutos)</p>	<p>Objetivo: dados una letra, nómbrarla</p> <p>Recursos: tarjetas con letras: a, s, m, l en mayúsculas y minúsculas</p>
--	---

Letras/elementos de letras: a, s, m, l

Palabras:

Palabras de alta frecuencia:

Actividad:

Diga a los estudiantes que van a aprender los nombres y los sonidos de algunas letras. Dígalos que después de aprender estas letras comenzarán a aprender a leer palabras.

Muestre cada letra y diga su nombre y sonido a los estudiantes. Pídales que repitan el nombre y sonido. Después de haber presentado todas las letras, muéstrelas en orden aleatorio y pregunte su nombre y sonido a los estudiantes.

<p>Fluidez (5 minutos)</p>	<p>Objetivo: muestre una tarjeta con letras; el estudiante la nombrará y dirá el sonido dentro de un periodo de tiempo de 3 segundos.</p> <p>Recursos: tarjetas con letras de a, m, s, l en mayúsculas y minúsculas</p>
--------------------------------	---

Eje: Letras

Actividad: muestre a los estudiantes cada letra y pídale que la nombren y dicen el sonido. Si no lo hacen en 3 segundos, dígalos el nombre y sonido y pídale que lo repitan. Continúe con otras letras.

<p>Vocabulario y comprensión (15 minutos)</p>	<p>Objetivo: Escuchan a una oración e identifican de quién trata.</p> <p>Recursos: Oración escrita en una pizarra o gráfico.</p>
---	--

Actividad:

Diga a los estudiantes que leerá una oración.

Mamá me ama.

¿Quién ama al niño? Mamá.