



CONSTRUCCIÓN Y VALIDACIÓN DE UNA PRUEBA DE COMPRESIÓN DE LECTURA MEDIANTE EL MODELO DE RASCH

CONSTRUCTION AND VALIDATION OF A READING COMPREHENSION TEST
THOUGH THE RASCH MODEL

Volumen 11, Número 2
pp. 1-27

Este número se publicó el 30 de agosto de 2011

Julieta Solórzano Salas
Eiliana Montero Rojas

La revista está indexada en los directorios:

[LATINDEX](#), [REDALYC](#), [IRESIE](#), [CLASE](#), [DIALNET](#), [DOAJ](#), [E-REVIST@S](#),

La revista está incluida en los sitios:

[REDIE](#), [RINACE](#), [OEI](#), [MAESTROTECA](#), [PREAL](#), [HUASCARAN](#), [CLASCO](#)

Los contenidos de este artículo están bajo una licencia [Creative Commons](#)



CONSTRUCCIÓN Y VALIDACIÓN DE UNA PRUEBA DE COMPREENSIÓN DE LECTURA MEDIANTE EL MODELO DE RASCH CONSTRUCTION AND VALIDATION OF A READING COMPREHENSION TEST THOUGH THE RASCH MODEL

Julieta Solórzano Salas¹
Eiliana Montero Rojas²

Resumen: El presente artículo expone los resultados del diseño y la construcción de una prueba diagnóstica de comprensión de lectura literal e inferencial, para estudiantes de escuelas públicas de III grado de la Educación General Básica, correspondientes a la Dirección Regional de San José, Costa Rica, realizada en el año 2009. Se trabajó con una muestra aleatoria de 406 estudiantes. La prueba está compuesta por 30 ítems, divididos en tres niveles de complejidad (fácil, intermedio y avanzado), según el criterio de personas juezas y los resultados obtenidos en el análisis. Se utilizaron los supuestos teóricos del Modelo de Rasch para el análisis de los resultados sobre confiabilidad y validez de los ítems, así como de la habilidad de las personas. Los resultados evidenciaron un desajuste constante en el 50% del grupo de personas examinadas, así como veinticuatro ítems con una alta confiabilidad (0.98). La media de dificultad de la prueba se ubicó por debajo de la media de habilidad de las personas examinadas. Un 98% del grupo de participantes se ubicó en los niveles de habilidades intermedias y altas, en la resolución de ítems que miden la comprensión de lectura literal e inferencial.

Palabras clave: VALIDACIÓN DE PRUEBA, COMPREENSIÓN DE LECTURA, MODELO DE RASCH, COSTA RICA

Abstract: This paper presents the results of the design and validation of a diagnostic test of literal and inferential reading comprehension was made on 2009. The test was designed for III grade General Basic Education students, Regional Direction of San José, Costa Rica. The study was done with a random sample of 406 students. The test consists of 30 items, divided into three levels of complexity (easy, intermediate and advanced), according to the criteria of expert judges and the results obtained in the analysis. The theoretical assumptions of the Rasch Model were used for the analysis of the results on the reliability and validity of the items, as well as on the ability of the individuals. The results show a continuing gap in 50% of persons examined and twenty-four items with high reliability (0.98). The mean difficulty of the test was below the average ability of participants. 98% of the group of students were located in intermediate and high skills levels in the resolution of items that measure literal and inferential reading comprehension.

Keywords: TEST VALIDATION, READING COMPREHENSION, RASCH MODEL, COSTA RICA

¹ Máster en Evaluación Educativa, Universidad de Costa Rica, Coordinadora y Docente de la Sección de Educación Especial, Investigadora del Instituto de Investigación Educativa (INIE), Universidad de Costa Rica. Dirección electrónica: julieta.solorzano@ucr.ac.cr

² Doctora en Investigación Educativa con énfasis en Medición y Evaluación, Instituto de Investigaciones Psicológicas y Escuela de Estadística de la Florida State University, Estados Unidos, Bachillerato en Estadística de la Universidad de Costa Rica. Actualmente es docente e Investigadora de la Universidad de Costa Rica. Dirección electrónica: eiliana.montero@ucr.ac.cr

Artículo recibido: 16 de marzo, 2011

Aprobado: 28 de julio, 2011

1. Introducción

La Política Educativa Hacia el Siglo XXI (1994, p. 6) en Costa Rica, señala la necesidad eminente de formar personas que sean capaces

(...) de comunicarse con el mundo de manera inteligente de tal manera que, a partir de la valoración de lo que lo identifica como costarricense, tome las decisiones que lo relacionen con otras culturas desde un punto de vista de pensador independiente, flexible y crítico, teniendo por guía los derechos y los deberes humanos.

Lo anterior significa que la lectura y su comprensión vienen a ser una de las vías para garantizar la comunicación entre las personas y su realidad inmediata, nacional y universal. La lectura y su comprensión se constituyen en las herramientas para que las personas puedan interactuar de manera responsable, comprometida y crítica con el medio que los circunda.

A partir de la importancia que tiene la comprensión de lectura para el desarrollo personal y social del estudiantado, es que las investigadoras se interesaron en construir y aplicar un instrumento para valorar en estudiantes de III grado de la Educación General Básica las habilidades en comprensión de lectura, en el nivel literal e inferencial. Para lograrlo, se utilizaron los supuestos teóricos y matemáticos del Modelo de Rasch.

El Programa de Español para I Ciclo de la Educación General Básica en Costa Rica (2005, p. 23), menciona que a lo largo de este ciclo se: "*Adquieren las destrezas y los hábitos para la lectura y para la escritura, lo mismo que las habilidades para expresarse oralmente de forma coherente y lógica*". De ahí la importancia de promover la comprensión de la lectura desde los niveles iniciales de enseñanza.

Ambas autoras consideran ideal el logro de una lectura madura desde los primeros años de escolaridad, es decir, aquella que le permite a la persona interactuar con el texto de manera independiente y alcanzar procesos complejos de pensamiento. Applegate, Applegate, McGeehan, Pinto y Kong (2009), indican que se entiende por *lectura madura* el proceso dinámico del pensamiento en la persona que lee, por el cual se ajustan los contenidos con las experiencias y los valores previos. Por lo tanto, hay una interacción dinámica entre la persona y el texto que construye significados. Este proceso es sumamente individual y particular.

La persona que tiene un nivel de lectura madura recurre a su bagaje intelectual de experiencias y de construcciones significativas para apropiarse del texto; de ahí que la lectura es

una habilidad que va más allá de la decodificación de grafemas establecidos en una unidad coherente de significado.

Medir la comprensión lectora para la toma de decisiones pedagógicas es un tema de interés frecuente en el profesorado y ha sido explorado en múltiples contextos y experiencias; sin embargo, en el contexto costarricense no se han utilizado las posibilidades que ofrece el Modelo de Rasch, tanto en el diseño de pruebas como en el análisis de los resultados obtenidos por la muestra de participantes.

El diagnóstico y la definición de acciones pedagógicas que favorezcan el desarrollo de una comprensión de lectura madura en el estudiantado inciden en un mayor desempeño académico y permiten establecer apoyos para responder a las necesidades de aprendizaje presentes en el aula.

Según Lyon (2001 citado por Katz, Carlisle, Lyman y Zeng 2008), si las necesidades específicas de aprendizaje son detectadas de manera temprana y se provee una intervención educativa sistemática, explícita e intensiva que implique el empleo de las herramientas necesarias para favorecer el aprendizaje de la lectura, el conjunto de estudiantes en riesgo puede alcanzar rendimientos promedios o superiores en la habilidad de comprensión lectora.

Para el caso de Costa Rica, Atorresi, Centanino, Bengochea, Jurado y Martínez (2009), indican que el 18.2 % del grupo de estudiantes de tercer grado que se examinaron en el Segundo Estudio Regional Comparativo y Explicativo (SERCE) alcanzaron el nivel de mayor complejidad (Nivel IV). De dicho informe se puede extraer que la población estudiantil en tercer grado necesita reforzar la habilidad para enfrentarse a textos explicativos de carácter científico. A pesar del resultado anterior, cabe rescatar que el conjunto de estudiantes que participó en las pruebas supera el porcentaje promedio general en todas las clases de textos.

2. Referentes teóricos: El concepto de la comprensión de lectura

Para los efectos de la presente investigación, la comprensión de lectura se entiende como la interpretación que realiza la persona que lee un texto en particular. Esta interpretación se encuentra enmarcada por las experiencias previas y los valores que posee la persona dentro de un contexto socio cultural determinado, por lo tanto, se puede considerar que la interpretación realizada por la persona va a ser compartida por el grupo que integra el contexto. Este consenso posibilita la definición de líneas de significados a partir de la lectura.

Además, es importante considerar otras conceptualizaciones que corresponden a diferentes posicionamientos teóricos de la comprensión lectora, las cuales no distan mayormente de la expuesta. Carrasco (1999, p. 93, citado en Méndez, 2000) expresa que

Existen diversas posiciones teóricas al respecto. La primera de ellas se centra en el contenido de texto y considera el acto lector como la habilidad de "recuperar" lo que ha escrito el autor. La segunda corriente enfatiza en el lector y concibe el acto de la lectura como la actualización de los esquemas de este al entrar en contacto con el texto. La tercera postura, denominada "transaccional", afirman que intervienen tanto el texto como los esquemas del individuo, de manera tal que cada uno condiciona y es condicionado por el otro.

Lo citado anteriormente, sin importar la posición teórica desde la que se asume la comprensión lectora, considera a la persona que realiza el acto de la lectura como un ente activo, que recupera y actualiza la información previamente adquirida con el aporte del texto, por lo que se produce una construcción personal. Al darse esta interacción entre el texto y la persona se evidencia un proceso comunicativo.

Mientras tanto, para Rosemberg, Borzone de Manrique y Diuk (2003, p. 128): "*La comprensión de un texto depende del interjuego entre la información propia del texto y los conocimientos del lector*". Lo ideal, por lo tanto, es que haya una gran cercanía entre el texto y los esquemas o categorías propias de la persona que lee.

Entre mayor conocimiento tenga la persona sobre el medio y mayores experiencias haya acumulado, mayor será la comprensión que puede tener del texto que lee. Sánchez (1995, p.139) explica que: "*Un relato basado en sus esquemas culturales y lingüísticos será procesado más fácilmente, en relación con otro cuyos elementos estructurales y de contenido representan una cultura ajena a la del lector*".

Por su parte, Pinzás (1997) también rescata, dentro de los ejercicios propios de la metacognición en la comprensión lectora, la importancia de los esquemas de la persona para enfrentarse al significado del texto, así como las habilidades de autocorrección generadas ante la incompreensión del texto. Esta última es una de las habilidades superiores de la comprensión de lectura madura.

A la hora de realizar una evaluación diagnóstica sobre la comprensión lectora, se debe tener presente las experiencias que poseen las personas examinadas y ajustarse a esa realidad.

La comprensión de lectura como constructo

La prueba de comprensión de lectura que se diseñó en esta investigación mide el constructo "comprensión lectora", explicado por los sub constructos "comprensión literal e inferencial". Ambos se definen del siguiente modo.

- *La comprensión de lectura literal.* Es el reconocimiento o localización de la información que contiene un determinado texto. Para Villafuerte (1998, p. 22) en la comprensión de lectura la persona: "(...) *no sale del texto ni del mundo creado, se mueve en su interior, lo recorre, lo reconoce, lo entiende, es capaz de desarmarlo y armarlo de nuevo*".

- *La comprensión de lectura inferencial.* Es la habilidad que posee la persona para establecer conjeturas o hipótesis a partir de la información que le provee una determinada lectura. Catalá, Catalá, Molina y Monclús (2001, p. 47) indican que el nivel inferencial "*Exige que el raciocinio y la imaginación vayan más allá de la página impresa*".

Por su parte, Gómez y Peronard (2000, p. 41) consideran que la comprensión inferencial es un proceso superior no aprendido, propio de las capacidades mentales humanas, que permite a la persona trascender de lo percibido a través de la lectura, para obtener: "(...) *información no explicitada a partir de otra que sí lo está*".

Pinzás (1997, pp.28-29) menciona que

(...) la comprensión de lectura suele entenderse considerando dos niveles: el literal y el inferencial. La comprensión literal se basa en la información que ofrece el texto a través de sus diversas proposiciones; la comprensión inferencial, más bien, aporta al texto la elaboración de las relaciones implícitas en él.

Ambos niveles son relevantes y permiten comprender el tipo de acercamiento que experimenta la persona que lee. En un primer nivel se encuentra aquella habilidad relacionada con una lectura que reconoce detalles; en un segundo nivel se encuentra la capacidad de producir o generar nueva información que no está contenida en el texto.

Características de la prueba de comprensión de lectura

La prueba que se desarrolló se enmarca dentro del propósito que expone Ravela (2006, p. 37); como

(...) una evaluación estandarizada de carácter diagnóstico que no tendrá consecuencias para los estudiantes, es legítimo evaluar aspectos relevantes, aun cuando no hayan sido

enseñados, justamente para entregar al cuerpo docente una señal en el sentido de que dichos aspectos deberían estar siendo enseñados.

La prueba de comprensión de lectura de la presente investigación se desarrolló bajo los postulados de las pruebas referidas a criterios, las cuales, al igual que la evaluación diagnóstica, pretenden identificar las habilidades definidas que evidencia cada estudiante, según el referente teórico que sustenta la prueba

La investigación realizada generó una prueba que se construye y se analiza con la guía de uno de los Modelos de Teoría de Respuesta al ítem: el *Modelo de Rasch*. En Costa Rica, tradicionalmente se han utilizado pruebas estandarizadas que se fundamentan en los supuestos de la Teoría Clásica de los Test, como es el caso de las pruebas nacionales en Bachillerato. Además, la investigación hace aportes en los protocolos recomendados para la construcción y la aplicación de una prueba diagnóstica estandarizada.

El Modelo de Rasch

El Modelo de Rasch, propuesto por el danés Georg Rasch en 1960, corresponde a un modelo matemático que utiliza un parámetro (b), que define a (c) con un valor de 0 y donde (a) es una constante. Este modelo permite, a su vez, medir con una misma escala las puntuaciones obtenidas por las personas y los resultados de los ítems; en la investigación realizada se desarrolló una estimación conjunta de los parámetros θ_s y b_i , según Prieto y Delgado (2003).

Una de las propiedades del Modelo de Rasch es la relación de la habilidad de la persona *versus* la dificultad del ítem. Considerando la dificultad del ítem y la habilidad de la persona en la variable latente y teniendo la posibilidad de azar en las respuestas como 0, se pueden presentar las siguientes situaciones que explican Bond y Fox (2001):

- Si la habilidad de la persona es igual a la dificultad del ítem, la persona tiene un 50% de probabilidad de acertar el ítem.
- Si la habilidad de la persona es superior a la dificultad del ítem, la persona tiene un 75% de probabilidad de acertar el ítem.
- Si la habilidad de la persona es menor a la dificultad del ítem, la persona tiene un 25% de probabilidad de acertar el ítem.

En ese sentido, Gonzáles (2008) indica que aquellos casos que no se ajusten a lo detallado anteriormente se deben a anomalías en los patrones de respuesta a los ítems.

Según Prieto y Dias (2003) la anterior propiedad favorece la construcción de pruebas para la selección de un grupo de personas, así como para la evaluación académica, la certificación profesional o los cambios producidos por una intervención educativa o psicológica. Indican, además, que enriquece la construcción de bancos de ítems y la equiparación de pruebas, haciendo un "escalamiento en la misma métrica" o "medición conjunta".

La escala Logit

Es la escala métrica más utilizada para ubicar tanto la habilidad de la persona como la dificultad del ítem. Esta escala tiene propiedades de intervalo. La escala Logit es el logaritmo natural de $\frac{P_{is}}{1 - P_{is}}$, es decir, $\theta_s - b_i$, que muestra la probabilidad de respuesta al ítem. El valor esperado de este algoritmo es 1.³

Al respecto, Tristán (2002, citado por Gonzáles, 2008, p.11) explica que: "Las unidades de medidas en logits se aplican para caracterizar tanto los grados de dificultad de los reactivos como los grados de habilidad de los sustentantes (...)."

Prieto y Delgado (2003) explican que, de manera arbitraria, se suele colocar como el punto cero (0) la dificultad media de los ítems. Esta medida, que es la más usual, puede adquirir valores infinitos, pero tradicionalmente se usa un rango de ± 5 .

Del mismo modo, Prieto y Delgado (2003) mencionan que se puede trabajar ubicando en el cero la dificultad media de las personas, en cuyo caso se usa un rango de ± 3 .

En suma, utilizar la escala Logit según la dificultad media de los ítems o de las personas depende del evaluador que utiliza el modelo. Ambos procedimientos son igualmente aceptados.

Estimación de los parámetros

Para iniciar este apartado es necesario tener claridad de los símbolos θ_s y b_i , representados por letras del alfabeto griego, zeta y beta. Estos ya se han venido presentando en este informe y son usados frecuentemente en las fórmulas matemáticas que componen al Modelo de Rasch; representan los parámetros de uso constante y corresponde a los sujetos e ítems respectivamente.

³ Muñiz (1997, p.106) explica que "Un logit $(\theta - b) = 1$, o, lo que es lo mismo, $\ln[P(\theta)/Q(\theta)] = 1$, indicará que $P(\theta)/Q(\theta) = 2.72$, ya que $\ln(2.72) = 1$ ".

Para poder definir la $P(\theta)$ se requiere conocer la diferencia entre la habilidad del sujeto y la dificultad del ítem, por lo tanto, se aplica la siguiente fórmula $P(\theta) = \theta_s - b_i$.

González (2008, p. 17) explica que: "(...) *la estimación de los parámetros de dificultad y habilidad se denomina calibración de propiedades de reactivos y habilidades de sustentantes*".

Según Prieto y Delgado (2003), luego de aplicar la prueba y tener disponibles los datos, se pretenden estimar los anteriores parámetros, en cuyo caso se pueden dar las siguientes situaciones:

- a. Estimación condicional: se pretende conocer el parámetro θ_s cuando ya se conocen los parámetros de b_i . Esto ocurre con ítems que ya se han estudiado en otras aplicaciones de la prueba, como en el caso de los bancos de ítems. Se utiliza el estimador de máxima verosimilitud donde se le asigna a cada persona el valor θ más probable para su patrón de respuestas. Martínez, Hernández y Hernández (2006) explican que la verosimilitud: "(...) *es el término dado a la probabilidad calculada con datos que han surgido en la realidad*" (p.156).
- b. Estimación conjunta: se establecen ambos parámetros θ_s y b_i . En el caso de la presente investigación se trabajó con esta estimación.

Ventajas del Modelo de Rasch

En cuanto al Modelo Rasch, Prieto y Delgado (2003) y Prieto y Dias (2003) explican que se pueden distinguir fundamentalmente cuatro ventajas:

- a. Medición conjunta: los parámetros de las personas y de los ítems se ubican en un mismo continuo de medición y con la misma escala.
Esta propiedad permite considerar la varianza natural de los ítems, porque no se puede suponer que cada ítem mide del mismo modo al constructo. Esta posición es sumamente diferente a la propuesta de la Teoría Clásica. Permite, a su vez, identificar la relación de cada persona con los ítems de la prueba y hace posible identificar cuáles habilidades se presentan. Utilizando los referentes de las pruebas referidas a criterios, se ofrecen amplias posibilidades en la evaluación diagnóstica de las personas.
- b. Estadísticos suficientes: la estimación de los parámetros de las personas va a estar condicionada por las puntuaciones en el test, es decir, con el número de aciertos se

puede hacer la estimación de la habilidad de la persona. Mientras que en los modelos de 2 y 3 parámetros se requieren de otras condiciones para lograr ser estimables.

- c. Objetividad específica: los parámetros de las personas no van a estar condicionados por las características del ítem y los parámetros de cada ítem no van a depender de las personas que lo responden.

La objetividad específica permite establecer las evidencias de validez y generalizar los resultados de las pruebas.

- d. Propiedades de intervalo: se cumple que distancias iguales en la escala representan distancias iguales en el constructo subyacente. La medida intercalar permite tener rigor en los análisis de los parámetros y garantiza la invarianza de las puntuaciones diferenciales a lo largo del continuo.

- e. Especificidad en el error típico de medida: el Modelo de Rasch permite:

1. Cuantificar la cantidad de información con la que se mide en cada punto de la dimensión.
2. Seleccionar los ítems que permiten disminuir el error en regiones del atributo previamente especificadas.

Esta ventaja del modelo es relevante para los tests referidos a criterios y favorecen la toma de decisiones en los puntos de corte.

Ajuste de los datos al modelo

Para poder aprovechar las ventajas que se plantean en el apartado anterior se debe cumplir una condición muy importante: que los datos empíricos se ajusten a las características que requiere el modelo.

La probabilidad de las respuestas dadas en un test va a depender de las habilidades de las personas y de lo que mide el ítem. Se debe recordar aquí que si el ítem es muy difícil, va a condicionar y a disminuir la probabilidad de acierto de una persona con niveles de habilidad bajos. Cuando se presenta la condición de que personas con bajas habilidades logran ítems con altos niveles de complejidad, o en caso contrario, que personas con altas habilidades no logran ítems fáciles, se puede asegurar que los datos empíricos no se ajustan al modelo.

La persona que realiza la evaluación debe observar esas condiciones particulares e identificar cuáles son los motivos que están generando la falta de ajuste. Karabatsos (2000 citado por Prieto y Delgado, 2003, p. 96) menciona que dicho desajuste de los datos se puede

dar por varios motivos tales como: "(...) *multidimensionalidad o sesgo de los ítems, falta de precisión en el enunciado o en las opciones, respuestas al azar, falta de motivación o cooperación, errores al anotar la respuesta, copiado de la solución correcta, etc.*".

Estadísticos utilizados

Para identificar las dificultades de ajuste al modelo se utilizan los estadísticos basados en residuos, es decir, la diferencia entre las respuestas observadas y las esperadas.

Al respecto, Martínez, Hernández y Hernández (2006) explican que los procedimientos matemáticos que utilizan los paquetes computacionales para realizar el análisis de los datos, según el Modelo de Rasch, son muy complejos y prácticamente imposibles de realizar con una calculadora de mano.

Para esta investigación interesa conocer los valores o rangos de valores que deben asumir los estadísticos a considerar, para la toma de decisiones sobre la prueba de comprensión lectora.

Uno de los estadísticos utilizados es el *infit* y por convención se encuentra en los siguientes valores, según el tamaño de la muestra con la que se trabaje:

Cuadro No. 1
Rango de los valores en el *infit*
que indican ajuste en el modelo según el tamaño de la muestra.
2009.

Tamaño de la muestra	Valores.
Menores de 500 casos	Entre 0,7 y 1,3
Entre 500 y 1000 casos	Entre 0,7 y 1,2
Más de 1000 casos	Entre 0,7 y 1,1

Fuente: Smith, Schumaker y Busch (1995 citados por Prieto y Delgado, 2003, Prieto y Dias 2003)

En la presente investigación se trabajó con un *infit* entre 0.7 y 1,3, porque el tamaño de la muestra es menor a 500.

Otro estadístico utilizado es el *outfit* que determina la discriminación de cada ítem. Los valores que indican un buen ajuste del ítem se encuentran entre 0,5 y 1,5.

A modo de resumen, se presentan los valores de los estadísticos utilizados en el análisis de los resultados de esta investigación. Escala Logit: es decir $\theta_s - b_i$, que indica la probabilidad de respuesta al ítem. El valor esperado de este algoritmo es 1. Los estadísticos *Infit* entre 0,7 y 1,3 y el *Outfit*: entre 0,5 y 1,5.

3. Marco Metodológico

Tipo de estudio

Se desarrolla bajo un enfoque cuantitativo. Hernández, Fernández y Baptista (2006, p. 5) indican que el enfoque cuantitativo: "*Usa la recolección de datos (...), con base en la medición numérica y el análisis estadístico, para establecer patrones de comportamiento y probar teorías*".

Universo

El universo del cual se extrae la muestra es el grupo de alumnos y alumnas de tercer grado de I Ciclo de la Educación General Básica correspondiente a la Dirección General de San José. La investigación se concentró en las escuelas públicas que abarca la Dirección Regional de San José, previo al acuerdo publicado en La Gaceta N°187, del viernes 25 de setiembre del 2009, Año CXXXI.

Muestra

El procedimiento para realizar el cálculo de la muestra partió de la lista oficial de las escuelas públicas que comprende la Dirección Regional de San José, su ubicación según circuito escolar, la cantidad de grupos de tercer grado por institución y el total de la población estudiantil matriculada en el año 2008. Los datos anteriores permitieron determinar una media de 28 estudiantes por grupo escolar de tercer grado.

Para definir n se aplicó la siguiente fórmula:

$$n = \frac{(Z_{\alpha/2} \cdot 5)^2}{d} = n = \frac{(1.96 \cdot 5)^2}{0.05} = 384$$

La variabilidad con una proporción de 0,05. La confiabilidad de 5%. Sin embargo, la muestra aumentó a 406 personas examinadas y el error en la fórmula se reduce a 0.02.

La selección de las 14 escuelas se realizó en forma aleatoria. Se trabajó con las tablas de números aleatorios (Mora, 2006) y el marco muestral se estableció numerando las escuelas de 1 hasta 159. Se decidió el escogimiento aleatorio, ya que, según Gómez (1999, p. 13), permite: "(...) *garantizar una probabilidad determinada a cada elemento de la población de ser incluido en la muestra*".

La negociación de entrada en los centros escolares se realizó mediante comunicación telefónica con la persona encargada de la dirección escolar.

Metodología en la elaboración del instrumento

Se partió de una definición del constructo, la comprensión lectora y los subconstructos por medir, que explican de manera independiente al constructo de interés, a partir del referente teórico. En este sentido, Prieto y Días (2003, p. 17) indican que

(...) la generación de ítems parte de un diseño experimental en el que, por un lado se seleccionan las características de los estímulos que inducen el uso de los procesos relevantes para medir el constructo, y por otro, permiten controlar las características asociadas a los procesos irrelevantes.

Para Kerlinger y Lee (2001) la definición sistemática y consciente del constructo permite establecer relaciones con otros constructos y, a su vez, permite realizar la observación y medición del mismo.

Instrumento

El instrumento que se utilizó fue generado por las investigadoras. Se siguieron los requisitos que establece Lindeman (1971) para la formulación del cuadernillo de examen, se detallan a continuación:

- Identificar claramente las instrucciones para responder a las preguntas y explicar cómo se hará la asignación del puntaje en el momento de la revisión.
- Ordenar los ítems según el grado de dificultad, a efecto de que la persona examinada pueda responder todos aquellos que domina y no se vea afectado por el tiempo.
- Ordenar la prueba según los tipos de ítems de modo que se agrupen según su estructura formal, así los que son de marcar con equis como los de pareo, etc.
- Cada ítem se debe ubicar en la misma página para facilitar su lectura.

- Elaborar un manual de guía para la persona aplicadora lo suficientemente detallado para evitar cualquier duda en la aplicación.

Los textos que componen la prueba se seleccionaron de los libros *Hacia el Siglo XXI* del Ministerio de Educación Pública (MEP), con el propósito de tener lecturas pertinentes al contexto costarricense y porque las lecturas están graduadas por dificultad de primero a cuarto grado. Las lecturas de cuarto grado se consideraron apropiadas para medir el nivel de dificultad de III grado avanzado.

Para la confección de la prueba se acordó tener una muestra de treinta ítems, de modo tal que cada nivel de lectura (literal e inferencial) está representado por quince ítems, éstos a su vez, miden tres niveles de complejidad o dificultad, es decir, primer grado avanzado, segundo grado avanzado y tercer grado avanzado, distribuidos según como se detalla en el siguiente cuadro. Se utilizó una única prueba para cada examinada o examinado.

Cuadro No. 2
Cantidad de ítems, nivel de complejidad y nivel de comprensión de lectura de la prueba de comprensión de lectura literal e inferencial. 2009.

Nivel de complejidad	Nivel de comprensión de lectura	
	Literal	Inferencial
I grado avanzado	5 ítems	5 ítems
II grado avanzado	5 ítems	5 ítems
III grado avanzado	5 ítems	5 ítems
<i>Total por nivel de comprensión de lectura</i>	15 ítems	15 ítems
<i>Total de ítems que compone la prueba</i>	30 ítems	

Fuente: Elaboración propia.

El instrumento contiene, además, un apartado de datos sociodemográficos de las personas: edad, sexo y nacionalidad.

Opciones de respuesta

En relación con la cantidad de opciones de respuesta, Prieto y Delgado (1998) explican que no hay evidencia estadística que defienda el uso de tres o cuatro opciones de respuesta. A pesar de lo anterior, recomiendan el uso de tres opciones de respuesta, indicando aspectos

pragmáticos tales como la economía en el tiempo de ejecución de la prueba, facilitación de la tarea de redactar opciones y disposición de mayor tiempo para realizar la prueba.

Por su parte, el MEP (2008) indica que los ítems de selección única para I Ciclo, deben presentar tres opciones de respuesta; dos jueces en el proceso de validación realizaron esta recomendación. Ante dichos antecedentes, la prueba que se utilizó presenta tres opciones de respuesta; una es la clave, dos son distractores.

Se procuró que las opciones de respuesta cumplieran con las siguientes características: homogeneidad en las opciones de respuesta, ausencia de errores en la digitación, usos de las mayúsculas y las minúsculas y de puntuación; vocabulario acorde al contexto de las personas examinadas, ausencia de dobles negativos, utilización del lenguaje inclusivo, evitar que la respuesta se dé sin leer el texto de cada ítem. Todos los enunciados utilizan la frase: *Según el texto anterior*; seguidamente se plantea la pregunta o frase a completar.

En el análisis de los datos los ítems se convierten en dicotómicos, de manera que la persona examinada muestra o no muestra la habilidad de que se trate.

Validación

Para obtener evidencias de contenido y de constructo se recurrió al criterio de personas expertas, se utilizó un total de nueve juezas y un juez, quienes determinaron el nivel de comprensión de lectura que medía cada ítem y la dificultad. Para el análisis de los criterios emitidos por cada persona jueza se realizó una matriz de recopilación de datos. La prueba se compone de aquellos ítems que alcanzaron un 75% de acuerdo entre las personas juezas.

En octubre del 2009 se hizo una aplicación con un grupo de 28 estudiantes de tercer grado, en una escuela pública del circuito 02 de la Dirección Regional de San José, que no participó en la muestra y se incorporaron las recomendaciones aportadas por el estudiantado.

A continuación, se expone la tabla de especificaciones de la prueba que se aplicó:

Cuadro No. 3
Tabla de especificaciones de la prueba de comprensión de lectura literal e inferencial.
2009.

Nivel de complejidad	Descripción	Detalle de preguntas	
		Literal Localizar elementos del texto.	Inferencial Usar las ideas e informaciones explícitamente planteadas en el trozo, su intuición y su experiencia como base para conjeturas e hipótesis.
FÁCIL I grado avanzado	Los textos utilizados en este nivel son descriptivos de 2 a 3 oraciones con un tema en común y con sujeto, verbo en presente o pasado y predicado. Las palabras utilizadas están compuestas por sílabas directas, mixtas e inversas. Monosílabas, bisílabas y trisílabas. Uso de palabras de uso cotidiano y relacionadas con la escuela, familia, animales.	1, 2, 3, 4, 5	16, 17, 18, 19, 20
INTERMEDIO II grado avanzado	Los textos utilizados en este nivel son narraciones de 3 a 4 oraciones por párrafo que describen una situación en común. Las oraciones están compuestas por sujeto, verbo en presente, pasado o futuro y todo tipo de complementos. Se procuró que las palabras utilizadas en los textos fueran desde monosílabas hasta plurisílabas compuestas por sílabas directas, inversas, mixtas y consonánticas	6, 7, 8, 9, 10	21,22, 23, 24, 25
AVANZADO III grado avanzado	Los textos utilizados en este nivel corresponden a narraciones, cuentos breves o relatos que utilizan figuras literarias como la hipérbole o la personificación. También se utilizaron temas informativos, con extensiones de 4 párrafos o más. Las oraciones están compuestas por sujeto, verbos conjugados en pasado, presente o futuro y todo tipo de complementos. Las palabras utilizadas en este nivel están formadas por sílabas directas, inversas, mixtas y consonánticas. Desde monosílabas hasta plurisílabas.	11, 12, 13, 14, 15	26 ,27, 28, 29, 30

Fuente: Elaboración propia.

Análisis estadísticos y generalización de resultados

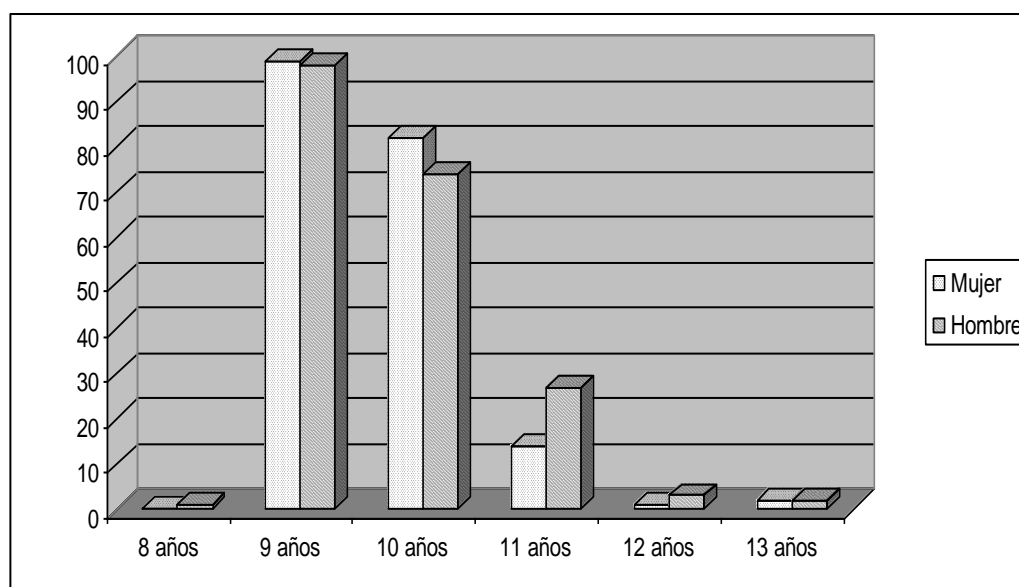
Para el análisis de los datos en crudo se utilizó el programa Winstep (Linacre, 2006), para obtener los criterios de la Bondad de Ajuste al Modelo de Rasch.

Los resultados obtenidos en la validación de la prueba de comprensión de lectura corresponden únicamente a los centros educativos públicos abarcados en la muestra.

4. Análisis y discusión de los resultados

La prueba se aplicó a una muestra de 406 estudiantes provenientes de 14 escuelas de la Dirección Regional de San José. La muestra se compone de 198 mujeres y 208 hombres. Del total del grupo de participantes la mayoría son costarricenses y apenas un 7% indica otras nacionalidades, entre las que la nicaragüense es la más frecuente (2% mujeres y un 3,4% de hombres).

Gráfico No. 1
Distribución de la población estudiantil examinada por sexo y edad.
Estudiantes de III grado de la Dirección Regional de San José.
2009.



Fuente: Resultados de la sección de datos sociodemográficos de estudiantes participantes de la prueba de comprensión de lectura.

Por su parte, en el Gráfico No. 1 se muestra el sexo y la edad de las personas examinadas. La edad mínima es de 8 años y la máxima de 13 años. La mayoría del grupo de estudiantes tiene 9 años de edad, lo cual corresponde a la edad esperada para el nivel de III grado. A partir del presente Gráfico se puede deducir que en la muestra de participantes, un grupo minoritario de estudiantes presenta sobreedad para el grado que cursa. Tanto las mujeres como los hombres presentan una distribución similar por cada año de edad.

Análisis inicial de ajuste de la totalidad de ítems según el Modelo de Rasch

Con el objetivo de depurar la prueba con los ítems que mejor miden el constructo de la comprensión lectora, se hace un primer análisis de la totalidad de la prueba con el grupo completo de personas examinadas. La confiabilidad de los treinta ítems y de la totalidad de personas examinadas se detalla en el Cuadro No. 2.

Cuadro No. 4
Confiabilidad de la prueba de comprensión de lectura y de estudiantes participantes.
Estudiantes de III grado de la Dirección Regional de San José, Costa Rica.
2009.

Parámetros	Cantidad	Valores de la confiabilidad
Personas	406	.69
Ítems	30	.99

Fuente: Elaboración propia según salidas del programa Winsteps con los resultados de la prueba de comprensión de lectura.

La confiabilidad de la prueba es de 0,99 y la confiabilidad de las personas es de 0,69. Este último valor es evidentemente bajo. Una primera explicación de este hecho puede deberse al desajuste de un 30,2% de las personas examinadas.

Cuadro No. 5
Ajuste de los 30 ítems según el Modelo de Rasch.
Prueba de comprensión de lectura literal e inferencia.
2009.

Número de ítem recodificado	Posición de ingreso del ítem	Cantidad de estudiantes que logran el ítem	Dificultad del ítem en logits	Error estándar	Estadístico <i>infitmnsq</i>	Estadístico <i>outfitmnsq</i>	Correlación del ítem con la prueba
P07RECOD	29	3	6.34	.59	1.09	9.90	-.20
P22RECOD	20	43	3.37	.17	1.18	2.41	.03
P28RECOD	26	67	2.80	.14	1.18	1.64	.10
P29RECOD	27	109	2.08	.12	1.12	1.37	.21
P27RECOD	25	142	1.63	.11	1.07	1.19	.30
P18RECOD	16	151	1.52	.11	1.06	1.12	.32
P26RECOD	24	159	1.42	.11	1.03	1.06	.35
P30RECOD	28	221	.69	.11	1.05	1.07	.34
P20RECOD	18	231	.57	.11	.96	.95	.43
P12RECOD	10	246	.39	.11	.95	.91	.44
P23RECOD	21	251	.33	.11	.93	.87	.46
P14RECOD	12	271	.07	.11	1.03	.97	.36
P09RECOD	7	282	-.07	.12	1.05	1.03	.32
P19RECOD	17	302	-.36	.13	.97	.90	.39
P03RECOD	3	311	-.50	.13	1.00	.88	.37
P15RECOD	13	317	-.60	.14	.95	.87	.39
P04RECOD	4	335	-.92	.14	1.01	.91	.33
P16RECOD	14	336	-.94	.14	1.04	1.00	.29
P05RECOD	5	339	-1.00	.14	.93	.77	.40
P06RECOD	30	339	-1.00	.14	.89	.71	.44
P10RECOD	8	340	-1.02	.14	.91	.83	.40
P25RECOD	23	342	-1.06	.14	.96	.90	.36
P02RECOD	2	349	-1.21	.15	.97	.82	.35
P21RECOD	19	365	-1.63	.17	.88	.63	.40
P24RECOD	22	365	-1.63	.17	1.00	.80	.29
P13RECOD	11	366	-1.66	.17	.92	.86	.33
P11RECOD	9	367	-1.69	.18	.94	.83	.32
P01RECOD	1	371	-1.82	.18	1.00	1.07	.23
P08RECOD	6	374	-1.92	.19	.90	.69	.35
P17RECOD	15	381	-2.21	.21	.93	.89	.29

Fuente: Elaboración propia según salidas del programa Winsteps con los resultados de la prueba de comprensión de lectura.

Según los valores del *infitmnsq*, los 30 ítems ajustaron, pues se encuentran entre los valores 0,7 y 1,3. Ítems como el 03, 24 y 01, logran un ajuste perfecto entre los datos y el modelo, porque logran un valor de 1.

Sin embargo, en el estadístico *outfitmnsq* los ítems 07 (9,90), 22 (2,41), 28 (1,64) evidenciaron falta de ajuste por alcanzar valores superiores a 1,5, por lo que se eliminaron de la muestra de ítems, para realizar un nuevo análisis de la confiabilidad. Estos ítems también muestran poca correlación con el total de la prueba y el ítem 07 (0,59) alcanza el mayor error

estándar. El ítem 16 evidenció un ajuste casi perfecto en los valores $infitmnsq$ (1,04) y $outfitmnsq$ (1,00).

Del mismo modo, los ítems 29 (0,21), 24 (0,29) y 17 (0,29) indican una baja correlación con el total de la prueba; lo deseable es que los valores se acerquen a 1. Debido a las dificultades de ajuste de los ítems analizados hasta este momento, se consideró pertinente eliminar los siguientes ítems para el análisis posterior: 07, 22, 28, 29, 24 y 17.

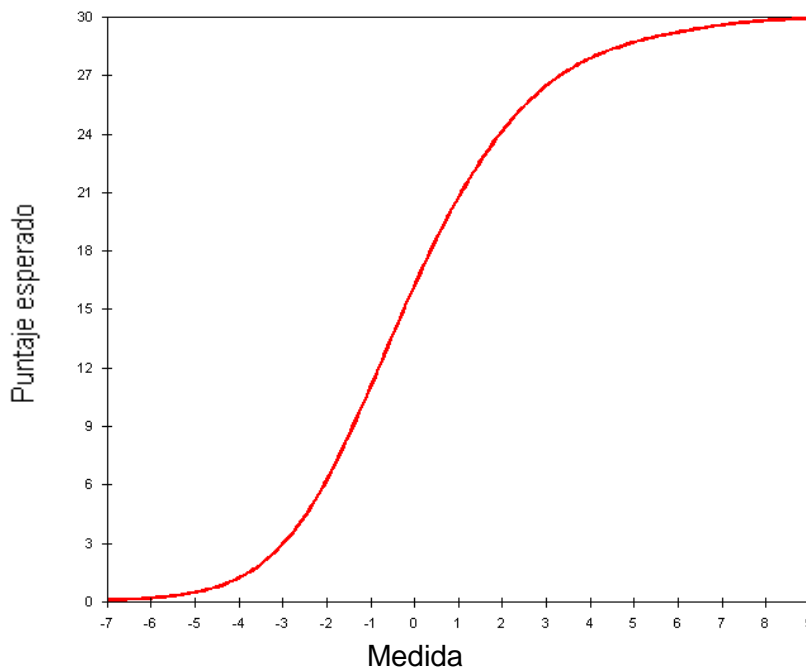
En general, se puede concluir que la mayoría de los ítems resultaron fáciles para las personas examinadas puesto que un 85% en promedio logró con total acierto 18 ítems de la prueba.

Los niveles de habilidad calibrados para las personas examinadas en la prueba de comprensión de lectura van de 6,34 logits a -2,21. El ítem que representa mayor complejidad es el ítem 07, el cual sólo fue acertado por 3 personas examinadas. Con 3 logits de diferencia continúa el ítem 22; los ítems 28 y 29 alcanzan valores altos en el logit 2,80 y 2,08 respectivamente, lo que indica que este grupo de ítems fue el más complejo para las personas examinadas. Es interesante resaltar que al menos los ítems 28 y 29 corresponden a la comprensión de lectura inferencial del nivel de III grado avanzado, según el criterio de las personas juezas. De hecho, los cinco ítems indicados en este nivel se mantienen dentro del grupo de ítems de mayor complejidad confirmando el criterio de las personas juezas.

La dificultad media de los ítems está por debajo de la habilidad de las personas examinadas. Esto reafirma el hecho de que la prueba sea fácil para la mayoría del grupo de estudiantes participantes. Lo anterior es ideal debido a que, al momento de aplicar la prueba, el grupo de personas examinadas se ubicaban en el nivel curricular más alto, es decir, en III grado avanzado.

Gráfico No. 3

Curva característica de la prueba de comprensión de lectura literal e inferencial. 2009.



Fuente: Elaboración propia según salidas del programa Winsteps con los resultados de la prueba de comprensión de lectura.

La Curva Característica del Test evidencia que, la prueba ofrece abundante información en habilidades bajas y poca en las habilidades presentes en estudiantes con alto desempeño en la comprensión de lectura.

Ajuste de los 24 ítems con el Modelo de Rasch

Con este nuevo ajuste se obtiene que los niveles de habilidad calibrada para las personas evaluadas en la prueba de comprensión de lectura van de 4,19 a -2,22 *logits*. El ítem que presenta mayor complejidad es el 25. Con 2 *logits* más abajo se encuentra el 19. El ítem más simple de resolver para la muestra evaluada es el 08 y el 06 se encuentra entre los más difíciles: apenas 86 de las 251 personas que componen este segundo análisis lo lograron. Según el criterio de las personas juezas este es un ítem con un nivel de dificultad intermedio, por lo que se presenta una inconsistencia respecto las personas expertas. Este tipo de análisis invita a una nueva revisión del nivel de dificultad de los ítems que componen la prueba.

Definición de habilidades a partir de la evidencia empírica

Una vez identificados los 24 ítems que ajustaron según las expectativas que tiene el modelo de Rasch, es necesario hacer un análisis general sobre las posibilidades de información que posee cada ítem en el constructo de interés.

La evidencia empírica demuestra que algunos ítems se comportaron diferente a lo previsto según el criterio establecido por las personas juezas. Así, por ejemplo, se tiene que el ítem 21 resultó ser fácil para las personas examinadas y el ítem 18 se ubica entre los de mayor dificultad.

Por lo anterior, se hace necesario establecer, de manera tentativa, niveles de interpretación de las posibles habilidades presentes en el grupo de personas examinadas, a partir del conjunto de ítems que representa cada nivel.

Para lograrlo, se establecieron tres puntos de corte en el mapa que provee el programa Winsteps, que ubica en un mismo continuo los 24 ítems y a las 406 personas. Haciendo un análisis de los ítems que representa cada nivel, las autoras de la presente investigación hacen la siguiente propuesta:

Cuadro No. 6
Niveles de habilidad que se pueden ubicar en la prueba de comprensión de lectura según la evidencia empírica. Prueba de comprensión de lectura literal e inferencial.
Dirección Regional de San José.
Año 2009

Nivel	Rango en logits de los ítems	en de los ítems	Ítems ubicados en este nivel	Interpretación de las habilidades presentes en cada nivel	Porcentaje de población examinada ubicada en este nivel
B A J O	-1.51 a 0.78	-	01,02,08,13,21	Estudiantes que pueden leer textos compuestos por hasta 4 párrafos y logran los ítems que miden principalmente la comprensión de lectura de orden literal.	2%
I N T E R M E D I O	-77 a 1.05		03,04,05,06,09,10,12,14,15,16,19,20,23,25,	Estudiantes que pueden leer textos compuestos por hasta 4 párrafos; muestran alta habilidad para resolver ítems que miden la comprensión de lectura literal y pueden resolver ítems que evalúan la comprensión de lectura inferencial.	42%
A L T O	1.06 a 2.16		18,26,27,30	Estudiantes que pueden leer textos de hasta 4 párrafos; muestran un alto desempeño en ítems que miden la comprensión de lectura literal e inferencial.	56%

Fuente: Elaboración propia con las salidas del Programa Winsteps y la tabla de especificaciones de la prueba.

La propuesta se constituye en un referente de las posibilidades diagnósticas e informativas de la prueba de comprensión de lectura a partir del análisis con el Modelo de Rasch.

El aporte de este ejercicio permite ubicar a los estudiantes en un nivel determinado para poder brindar información sobre las habilidades presentes, según la habilidad demostrada en la resolución de la prueba de comprensión.

Es necesario indicar que el nivel con mayor muestra de ítems es el intermedio y en el nivel alto, debido a la poca cantidad de ítems que mide este nivel, se puede afirmar que existen dificultades para medirlo de manera representativa. Esta distribución confirma que la prueba

ofrece poca información en habilidades altas de comprensión de lectura, puesto que 4 ítems son un escaso referente teórico.

Se puede concluir, entonces, que el grupo de personas examinadas se distribuye entre habilidades intermedias y altas para resolver ítems de comprensión de lectura e inferencial.

5. Conclusiones

La comprensión de lectura es una de las habilidades necesarias para el desarrollo pleno de la persona en una sociedad universal, rodeada de múltiples redes de comunicación. Por lo tanto, debe ser interés del sistema educativo y del profesorado de una nación, perfeccionar las técnicas e instrumentos que permitan valorar las habilidades que presenta el estudiantado. También procede desarrollar acciones pedagógicas para fortalecer y estimular la comprensión de lectura en todo el estudiantado.

Esta investigación procura sistematizar la aplicación del Modelo de Rasch en temas de interés propios del área educativa, como es el caso de la comprensión de lectura. Evidentemente, la implementación del Modelo demanda rigurosidad, tanto en la construcción teórica que explica cada constructo a medir, como en la metodología para establecer la validez de los instrumentos a utilizar, con el respaldo del criterio de personas expertas

Ahora bien, de acuerdo con lo señalado por Messick (1989), es importante recalcar en todo momento la función social que tiene la prueba que propone esta investigación. La prueba es únicamente de diagnóstico y permite identificar las habilidades del estudiantado examinado, según el comportamiento de los ítems en la muestra. No se puede interpretar que el logro de un reactivo determinado indica la habilidad en ese nivel de lectura o de complejidad en el texto. La prueba no debe ser utilizada para la selección de personas.

En el análisis de los resultados de la prueba de comprensión de lectura con el Modelo de Rasch, considerando tanto el ajuste de las personas como de los ítems, se evidencia que:

- La prueba tiene una alta confiabilidad referida a las estimaciones de los parámetros de los 24 ítems de la prueba final (0,98), es decir, que si se administrara el mismo conjunto de reactivos a otra muestra de estudiantes seleccionada al azar, en el mismo universo y si se correlacionan las estimaciones de b para cada ítem esta correlación sería igual a 0,98.

- Es importante hacer una nueva revisión de la validez de contenido de los 24 ítems que ajustaron, con el fin de valorar nuevamente su pertinencia a un nivel determinado de comprensión de lectura según la tabla de especificaciones, Lo anterior en vista de que algunos ítems que, según el criterio de las personas juezas eran fáciles, resultaron complejos según el valor *logit* alcanzado. Por ejemplo, el ítem número tres.
- La media de dificultad de la prueba de comprensión de lectura utilizada en esta investigación resultó por debajo de la media de habilidad de las personas examinadas, lo que evidencia empíricamente que los reactivos que componen la prueba resultan ser fáciles para el grupo de participantes.
- Un 98% de las personas examinadas de III grado se ubican en una habilidad intermedia o alta. Lo anterior se puede interpretar como un logro en el rendimiento de las personas examinadas, porque se ajusta a lo que pretende la propuesta curricular del MEP.

Como último punto, es necesario rescatar que la prueba de comprensión de lectura de la presente investigación puede constituirse en un referente para futuras evaluaciones diagnósticas de la comprensión de lectura literal e inferencial utilizando los aportes teóricos del Modelo de Rasch.

Referencias

- Applegate, Anthony, Applegate, Mary, McGeehan, Catherine, Pinto, Catherine y Kong, Ailing. (2009). The Assessment of Thoughtful Literacy in NAEP: Why the States Aren't Measuring Up. **The readers teachers**, **62** (5), 372 – 381.
- Atorresi, Ana, Centanino, Ivannia, Bengochea, Rodrigo, Jurado, Fabio, Martínez, Ricardo, Pardo, Carlos. (2009). **Aportes para la enseñanza de la Lectura** [Versión electrónica]. Santiago, Chile: Impreso por Salesianos Impresores S.A.
- Bond, Trevor y Fox, Christine. (2001). **Appling the Rasch Model Fundamental Measurement in the human sciences**. New Jersey: Lawrence Erlbaum Associates, Inc.
- Catalá, Gloria, Catalá, Mireia, Molina, Encarna y Monclús, Rosa. (2001). **Evaluación de la comprensión lectora pruebas ACL** (1.º - 6.º de primaria). Barcelona: Graó.
- Consejo Superior de Educación, Costa Rica. (1994). **Política Educativa Hacia el Siglo XXI**. Recuperado de <http://www.oei.es/quipu/costarica/politicaeducativasigloXXI.pdf>.
- Gómez, Miguel. (1999). **Elementos de estadística descriptiva** (5a reimp. de la 3a ed.) San José: EUNED.
- Gómez, Luis y Peronard, Marianne. (2000). La comprensión de textos escritos. En Magdalena Viramonte (Comp.) **Comprensión lectora dificultades estratégicas en resolución de preguntas inferenciales** (pp.15- 52). Argentina: Ediciones Colihue.
- González, Manuel. (2008). **El análisis de reactivos con el modelo Rasch, manual técnico A**. México: Instituto Nacional para la evaluación de la Educación, Universidad de Sonora.
- Hernández, Roberto, Fernández, Carlos y Baptista Pilar. (1991/2006). **Metodología de la investigación** (3a ed.). México: Mc Graw Hill.
- Katz, Lauren, Carlisle, Joanne, Lyman, Douglas y Zeng, Ji. (2008) Initial progress of children identified with disabilities in Michigan's reading first schools. **Exceptional children** **74** (2), 235-256.
- Kerlinger, Frederick y Lee, Howard. (2001) **Investigación del comportamiento: Métodos de investigación en ciencias sociales** (4º ed.). México: Mc Graw Hill.
- Lindeman, Richard. (1971). **Tratado de medición educacional**. México: Paidós.
- Martínez, Rosario, Hernández, María y Hernández, María. (2006). **Psicometría**. Madrid: Alianza Editorial, S.A.
- Messick, Samuel. (1989). Meaning and values in test validation: the science and ethics of assesment. **Educational Researcher**, **18** (2), 5 – 11.

- Méndez, Silvia. (2000). La mediación docente en la comprensión de lectura de textos literarios. **Revista Educación, 24** (1), 89 – 105.
- Ministerio de Educación Pública (MEP). (2008). **La prueba escrita**. San José: Editorama.
- Ministerio de Educación Pública (MEP), Costa Rica. (2005). **Programa de español para I Ciclo de la Educación General Básica**. San José: MEP.
- Mora, Marjorie. (2006). **Fórmulas y tablas estadísticas**. San José: Editorial UCR.
- Pinzás, Juana. (1997). **Metacognición y lectura**. Perú: Fondo Editorial de la Pontificia Universidad Católica del Perú.
- Prieto, Gerardo. y Delgado, Ana. (2003). Análisis de un test mediante el modelo de Rasch [Versión electrónica]. **Revista Psicothema, 15** (1), 94-100.
- Prieto, Gerardo y Dias, Angela. (2003). Uso del modelo de Rasch para poner en la misma escala las puntuaciones de distintos test. **Actividades en Psicología, 19** (106), 5 – 23.
- Ravela, Pedro. (2006). **Para comprender las evaluaciones educativas Fichas didácticas** [versión electrónica]. PREAL: Editorial San Marino.
- Rosemberg, Celia, Borzone de Manrique, Ana y Diuk, Beatriz. (2003). La lectura de textos expositivos: estrategias de interacción para acercar los textos a los niños. **Revista Interdisciplinaria, 20** (2), 121-145.
- Sánchez, Marta. (1995). La teoría de los esquemas y la comprensión de lectura: Algunas consideraciones generales. **Káñina, Revista Artes y letras, Universidad de Costa Rica, XIX** (1), 137 – 142.
- Villafuerte, Oscar. (1998). **Fortalecimiento de la comprensión inferencial y las operaciones fundamentales en los niños de cuarto y quinto grado de la comunidad de San Rafael, Liberia**. Memoria del curso FD 0129 El maestro investigador en el aula, Escuela de Formación docente, Universidad de Costa Rica, Guanacaste, Costa Rica.