

**EVALUACIÓN OBJETIVA DE LOS PROCESOS COGNITIVOS INVOLUCRADOS  
EN LA COMPRENSIÓN DE LECTURA**

**ÁNGELA MARÍA ESPINOSA GARZÓN**

**UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS HUMANAS  
MAESTRÍA EN PSICOLOGÍA  
2013**

**EVALUACIÓN OBJETIVA DE LOS PROCESOS COGNITIVOS INVOLUCRADOS  
EN LA COMPRENSIÓN DE LECTURA**

**ÁNGELA MARÍA ESPINOSA GARZÓN**

**Tesis para optar el título de Magister en Psicología**

**Director: Aura Nidia Herrera Rojas, Ph. D.**

**UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS HUMANAS  
MAESTRÍA EN PSICOLOGÍA  
2013**

*Agradezco a Dios por iluminar mi camino y  
Darme la sabiduría para hacer la mejor elección.*

*A mis padres, mi hermana y mi sobrino que me han apoyado en todo momento  
Y porque gracias a su amor y sus consejos he logrado ser la persona que soy.*

*A la profesora Aura Nidia Herrera, a quien admiro como maestra y como ser humano,  
Por sus enseñanzas.*

*A mi amiga Pilar Soler por su gran apoyo para la realización de esta investigación.*

*A los jurados de tesis Gerardo Prieto y Erika Margarita Arias.*

*A todos mis compañeros y amigos, especialmente a los que pertenecen al  
Laboratorio de Psicometría.*

## Resumen

El presente estudio pretende identificar los elementos o características de las preguntas de las pruebas SABER 11, que pueden hacer diferencia en los procesos involucrados en la evaluación de la comprensión de lectura en población con y sin limitación visual, con el fin de diseñar y construir una propuesta de banco de ítems estructurados para evaluar comprensión de lectura en personas sin limitación visual, siendo la base para construcción de medidas equivalentes para las dos poblaciones. Esta investigación hace parte del Proyecto "Procedimiento para establecer equivalencia en las puntuaciones de pruebas de aplicación masiva en personas con y sin limitación visual" financiado por Colciencias e ICFES en el marco de la Convocatoria Nacional Colciencias. Se utilizó la base de datos de la subprueba de Lenguaje de la prueba SABER 11 de 2008 segundo cohorte y se realizaron los análisis estadísticos bajo el Modelo de Rasch para identificar las posibles diferencias en la evaluación de personas con y sin limitación visual, se propone el diseño y construcción de ítems para evaluar la comprensión de lectura, tomando como base los procesos psicológicos involucrados y teniendo en cuenta las posibles diferencias entre personas con y sin limitación visual. Los resultados muestran que existe una diferencia entre las personas con y sin limitación visual, se detectaron cuatro ítems con DIF y se realizó un análisis cualitativo de los mismos, además, se construyeron 30 ítems que evalúan comprensión de lectura con base en los procesos cognitivos involucrados. Finalmente, se discuten algunas implicaciones prácticas y la pertinencia de construir ítems equivalentes para las poblaciones con y sin limitación visual.

**Palabras Clave:** Comprensión de lectura, Procesos Cognitivos, Modelo de Rasch, Construcción de Ítems

## Abstract

This study aims to identify the elements or characteristics of the test questions SABER 11 which can make a difference in the processes involved in assessing reading comprehension in people with and without visual impairment, in order to design and build a proposed bank structured to assess reading comprehension in visually impaired people, being the basis for construction of equivalent measures for the two populations. This research is part of the "Procedure for establishing equivalence test scores of mass application in people with and without visual impairment" and ICFES Colciencias aimed under the National Call Colciencias. The database language subtest was used SABER 11 2008 test, second cohort and statistical analysis under the Rasch Model were performed to identify possible differences in the assessment of people with and without visual impairment, we propose the design and construction of items to assess reading comprehension, based on the psychological processes involved and taking into account possible differences between people with and without visual impairment. The results show that there is a difference between people with and without visual impairment; four items with DIF were detected and a qualitative analysis of the data was performed also 30 items that assess reading comprehension based on the cognitive processes involved were built. Finally, some practical implications and relevance of building equivalent to the populations with and without visual impairment items were discussed.

**Keywords: Reading Comprehension, Cognitive Processes, Rasch Model, Construction items.**

## TABLA DE CONTENIDO

|  |                                      |
|--|--------------------------------------|
| Resumen .....  | 4                                    |
| Abstract .....   | 5                                    |
| TABLA DE CONTENIDO.....  | 6                                    |
| LISTA DE TABLAS .....  | 8                                    |
| LISTA DE FIGURAS.....  | 9                                    |
| INTRODUCCIÓN .....   | 10                                   |
| REVISIÓN TEÓRICA.....  | <b>¡Error! Marcador no definido.</b> |
| Comprensión de lectura.....  | 16                                   |
| Modelos de comprensión lectora.....                                    | 16                                   |
| Procesos psicológicos involucrados en la comprensión de lectura .....  | 23                                   |
| Evaluación de la comprensión de lectura.....                           | 27                                   |
| Diseño y construcción de bancos de ítems .....                         | 31                                   |
| Diseño y planeación de un banco de ítems .....                         | 31                                   |
| Definición del atributo .....  | 33                                   |
| Componentes y competencias a evaluar .....                             | 34                                   |
| Estructura del banco de ítems.....                                     | 34                                   |
| Escogencia del grupo de Autores de preguntas.....                      | 35                                   |
| Tipos de pregunta .....  | 35                                   |
| Consideraciones para la construcción de ítems .....                    | 37                                   |
| Consideraciones generales .....  | 38                                   |
| Consideraciones para la elaboración de los enunciados .....            | 38                                   |
| Consideraciones para la elaboración de las opciones de respuesta ..... | 39                                   |
| Características de las preguntas.....                                  | 39                                   |
| Análisis estadístico con el Modelo de Rasch .....                      | 40                                   |
| Definición del Modelo de Rasch .....                                   | 41                                   |
| Estimación de parámetros .....   | 44                                   |
| Utilidades del modelo .....  | 46                                   |

|   |                                |
|---|--------------------------------|
| METODO .....                                | ¡Error! Marcador no definido.1 |
| Fase 1.....                                 | 511                            |
| Instrumento.....                            | 511                            |
| Procedimiento.....                          | 544                            |
| Fase 2.....                                 | 555                            |
| Diseño y planeación del Banco de ítems..... | 566                            |
| Desarrollo del Banco de ítems.....          | 588                            |
| RESULTADOS .....                            | ¡Error! Marcador no definido.1 |
| Fase 1.....                                 | 611                            |
| Fase 2.....                                 | 700                            |
| DISCUSIÓN Y CONCLUSIONES .....              | ¡Error! Marcador no definido.3 |
| REFERENCIAS.....                            | 766                            |
| ANEXOS .....                                | 822                            |

**LISTA DE TABLAS**

|  |    |
|--|----|
| Tabla 1. Componentes evaluados en las pruebas SABER 11 del año 2008 .....                          | 15 |
| Tabla 2 Ejemplo de una estructura de prueba .....  | 34 |
| Tabla 3 Tipos de preguntas utilizadas en la evaluación objetiva (ICFES, 2009) .....                | 36 |
| Tabla 4. Rango de valores en el Infit que indican ajuste al modelo.....                            | 46 |
| Tabla 5 Estructura de prueba subprueba de Lenguaje Saber 11 2008-II.....                           | 52 |
| Tabla 6 Estructura de prueba para evaluar comprensión de lectura.....                              | 57 |
| Tabla 7 Estadísticos descriptivos de ítems y sujetos para la población general y subpoblaciones    | 61 |
| Tabla 8 Estadísticos descriptivos de ítems y sujetos para las subpoblaciones Sin LV y Con LV       | 62 |
| Tabla 9 Ajuste de los ítems al modelo en cada subpoblación.....                                    | 63 |
| Tabla 10 Estimación de la diferencia de la dificultad en población con y sin limitación visual ... | 65 |
| Tabla 12 Estimaciones de la dificultad de ítems identificados con DIF .....                        | 66 |

**LISTA DE FIGURAS**

Figura 1 Curva característica de los ítems según su nivel de dificultad .....43

## INTRODUCCIÓN

En nuestro país el derecho a la educación está contemplado en el art. 67 de la Constitución Política de Colombia (Gaceta C, 1991), donde reza que "La educación es un derecho de la persona y un servicio público que tiene una función social; con ella se busca el acceso al conocimiento, a la ciencia, a la técnica, y a los demás bienes y valores de la cultura". Además, en el Art. 68° contempla que "La erradicación del analfabetismo y la educación de personas con limitaciones físicas o mentales, o con capacidades excepcionales, son obligaciones especiales del Estado."

Aunque la inclusión educativa está estipulada en las leyes Colombianas, hasta hace poco tiempo las instituciones encargadas de educación superior iniciaron estrategias conducentes a facilitar el acceso de estudiantes con algún tipo de limitación realizando adaptaciones en el entorno de aplicación; en el caso de personas con limitación visual se ha buscado la implementación de *software* que permite apoyo auditivo, así como las adaptaciones del entorno gráfico y el acompañamiento con lectores entrenados.

Estas adaptaciones han constituido un gran avance para la inclusión al sistema educativo de las personas que presentan algún tipo de limitación visual, sin embargo en la evaluación educativa, es precaria la investigación, el diseño y construcción de pruebas objetivas que consideren las diferencias entre poblaciones con y sin limitación visual y que se tengan en cuenta para la construcción de instrumentos de evaluación; esto impide afirmar que las adaptaciones utilizadas sean suficientes para garantizar que los exámenes masivos, y particularmente los ítems que los conforman, estén diseñados de acuerdo a las necesidades que presentan las personas con esta limitación, ya que no se tiene en cuenta la modalidad sensorial que es utilizada para acceder

a la información. De este modo, no puede afirmarse que las puntuaciones obtenidas en este tipo de exámenes masivos sean equivalentes entre poblaciones con y sin limitación visual.

Autores como Crowder (1985) y Hoover & Gough (1990) afirma que, según la modalidad sensorial utilizada para la entrada de la información, se presentan diferencias en el procesamiento de un texto, bien sea de manera visual o auditiva; por ejemplo, cuando un texto se procesa por medio de la audición las exigencias de memoria son mayores que en la comprensión del lenguaje por medio de la lectura de textos escritos, esto puede deberse a que no se puede volver atrás en el texto para aclarar algo cuando es otra persona quien lo lee, pero si se puede lograr cuando se lee el texto directamente. También existen diferencias en lo complejo de la sintaxis y esto puede explicarse porque en el lenguaje hablado, normalmente se utiliza un amplio conjunto de indicaciones sobre la estructura de la frase, también llamadas claves prosódicas, entre las que se encuentran el tono, la entonación, el acento y el ritmo, que proporcionan considerable información sobre la estructura subyacente de las frases.

En este sentido, en el proceso de comprensión existen diferencias en la forma como se procesa el texto, bien sea de manera visual o auditiva. Cuando un texto se procesa por medio de la audición las exigencias de memoria son mayores que en la comprensión del lenguaje por medio de la lectura visual del texto; también debe tenerse en cuenta lo complejo de la sintaxis y las claves prosódicas entre las que se encuentran el tono, la entonación, el acento y el ritmo, que proporcionan información sobre la estructura subyacente de las frases. Además de lo mencionado anteriormente, en el proceso de la comprensión de un texto, cobra un papel importante el individuo, debido a que el lector aporta sus conocimientos previos que a su vez interactúan con las ideas que plantea el autor del texto y de esta manera se construye el significado de lo que se lee (Crowder, 1985; Alonso et al., 1985).

De acuerdo con lo expuesto anteriormente, surge el interés particular por evaluar la comprensión de lectura en personas con y sin limitación visual, como un proceso básico que debe realizar cualquier persona que presente una prueba o examen de aplicación masiva, como las pruebas Saber realizadas por el Instituto Colombiano para la Evaluación de la Educación-ICFES. En este contexto, la comprensión de lectura puede concebirse como una competencia transversal e independiente del área de conocimiento que se quiera evaluar, y más aun, si se tiene en cuenta que dichas pruebas se aplican de manera escrita, en el caso de las personas sin limitación visual, o auditiva, en el caso de las personas con limitación visual.

Es pertinente entonces preguntarse si se está evaluando el mismo constructo, en este caso comprensión de lectura, en las dos poblaciones e identificar si existe alguna diferencia en el procesamiento, almacenamiento y acceso a la información. Si se tienen en cuenta estas posibles diferencias de orden psicológico en el momento de la construcción de los ítems que conforman las pruebas de evaluación de la comprensión de lectura, podría considerarse que los puntajes obtenidos son equivalentes entre las dos poblaciones y se minimiza el riesgo de beneficiar en la evaluación a alguno de los dos grupos.

Esta investigación pretende realizar una propuesta para el diseño y construcción de un banco de ítems para evaluar comprensión lectora identificando las posibles diferencias en los procesos involucrados en las poblaciones, de manera que se contribuya a que las puntuaciones obtenidas con estos ítems sean equivalentes para las poblaciones con y sin limitación visual. Este banco de ítems podría utilizarse como insumo para desarrollar a futuro un Test Adaptativos Informatizados (TAI's) que le permitan a las personas con limitación visual, inicialmente, responder los exámenes de aplicación masiva de manera autónoma y con un nivel de precisión mayor, empleando un número reducido de ítems.

El presente trabajo hace parte del Proyecto "Procedimiento para establecer equivalencia en las puntuaciones de pruebas de aplicación masiva en personas con y sin limitación visual" financiado por Colciencias e ICFES en el marco de la Convocatoria Nacional Colciencias - ICFES para la conformación del banco de proyectos de investigación en el área de la calidad de la educación año 2010, mediante contrato CONV COLC/UNAL 255-2011.

### **Objetivo General**

Identificar los elementos o características de las preguntas de las pruebas SABER 11, que pueden hacer diferencia en los procesos involucrados en la evaluación de la comprensión de lectura en población con y sin limitación visual

Diseñar y construir una propuesta de banco de ítems estructurados para evaluar comprensión de lectura en personas sin limitación visual, siendo la base para construcción de medidas equivalentes para las dos poblaciones.

### **Objetivos Específicos**

1. Identificar las posibles diferencias en la evaluación de personas con y sin limitación visual, mediante los análisis estadísticos de la subprueba de lenguaje de la Prueba Saber 11.
2. Proponer un diseño y construcción de un banco de ítems para evaluar la comprensión de lectura, tomando como base los procesos involucrados y teniendo en cuenta las posibles diferencias entre personas con y sin limitación visual.
3. Brindar algunas directrices o recomendaciones para la construcción de ítems que pretendan evaluar comprensión de textos

## REVISIÓN TEÓRICA

Durante la última década se ha llevado a cabo un cambio en el modelo educativo en Latinoamérica basándose en la educación por competencias, y de esta manera se convirtió en una alternativa para superar las deficiencias del modelo tradicional que se enfocaba solamente en la transmisión de información, la memorización y la repetición de contenidos dejando de lado procesos complejos como el razonamiento, la solución de problemas y el pensamiento crítico (Guzmán, 2010). Colombia también ha implementado el modelo de formación basado en competencias académicas que le permite al estudiante adquirir conocimientos y habilidades necesarias para desempeñarse en su vida laboral y académica (Castillo, M., 2004).

El Instituto Colombiano para la Evaluación de la Educación - ICFES, en los últimos años, ha venido desarrollando la evaluación de competencias educativas definidas como un *saber hacer flexible* donde se evidencia la capacidad de usar los conocimientos adquiridos en situaciones distintas de aquellas en las que se aprendieron, además, son competencias transversales a todas las áreas del conocimiento. Esta evaluación se ha llevado a cabo a nivel nacional y ha teniendo en cuenta los diferentes niveles educativos como lo son la educación básica, media y profesional (ICFES, 2010).

En la educación básica se han desarrollado las pruebas SABER 5° y 9° (ICFES, 2009) con el fin de valorar las competencias básicas en lenguaje, matemáticas y ciencias, permitiendo utilizar esos resultados para identificar las fortalezas y debilidades en las competencias evaluadas e implementar planes de mejora. En la educación media se han desarrollado las pruebas SABER 11° que evalúan seis áreas de núcleo común y dos componentes flexibles que se presentan en la Tabla 1; en cada área de núcleo común se evalúan competencias específicas, por ejemplo, en el área de lenguaje se evalúan las tres competencias comunicativas, la primera es la acción

interpretativa que enfocada en la comprensión del sentido del texto, la segunda, es la acción argumentativa que busca explicar las ideas que articulan y dan sentido a un texto a partir de la interpretación del mismo, por último, la acción propositiva se caracteriza por ser una actuación crítica basada en la interpretación, que exige al lector acudir a sus saberes previos. Los resultados obtenidos en la aplicación de estas pruebas se utilizan, además de evaluar la calidad de la educación, como un requisito indispensable para el ingreso a la educación superior y/o para otorgar beneficios como becas o premios.

Tabla 1. Componentes evaluados en las pruebas SABER 11 del año 2008

| <b>SABER 11° (EXAMEN DE ESTADO)</b> |  |  |
|-------------------------------------|--|--|
| <b>Núcleo Común</b>                 | Lenguaje, Matemáticas, Biología, Química, Física, Ciencias Sociales, Filosofía, Inglés |  |
| <b>Componente Flexible</b>          | Profundización   | Biología, Ciencias sociales, Matemáticas, Lenguaje |
|                                     | Interdisciplinar   | Violencia y Sociedad, Medio Ambiente               |

En la educación superior, se han implementado las pruebas SABER PRO (ICFES, 2010) que tiene como objetivo general evaluar la calidad de la educación superior en Colombia y de las instituciones que brindan este tipo de formación. Además, esta evaluación busca comprobar el grado de desarrollo de las competencias de los estudiantes próximos a obtener un título de formación profesional; para cada una de las profesiones o disciplinas se evalúan contenido y competencias relacionadas con sus temáticas propias de su formación,

A nivel internacional se ha desarrollado el Programa Internacional de Evaluación de Estudiantes PISA (OCDE, 2006) enfocado en la evaluación de las competencias de comprensión de lectura, matemáticas y ciencias. La evaluación de la comprensión de lectura se realiza por medio de dos tipos de textos; el primer grupo está compuesto por los textos continuos, que se componen de oraciones y párrafos como los textos narrativos, expositivos, descriptivos y

argumentativos, y el segundo grupo, conformado por los textos discontinuos entre los que se encuentran las Tablas, graficas, cuadros, diagramas, mapas, entre otros. Los textos continuos representan las dos terceras partes de los textos utilizados en la evaluación, siendo los textos expositivos los más utilizados.

### **Comprensión de lectura**

Históricamente se ha evidenciado un interés por el estudio de la comprensión de lectura desde el inicio de la alfabetización, dado que la lectura no se adquiere de manera espontánea como el lenguaje oral, ella necesita de un aprendizaje, generalmente, provocado por el entorno (Golder & Gaonac'h, 2002).

A través del tiempo el concepto de lectura ha tenido modificaciones importantes. Inicialmente se concebía la lectura como un reconocimiento visual de códigos y/o palabras que se articulaban con una comprensión del lenguaje oral; luego, se pensaba que la lectura y la comprensión no eran procesos que se daban de manera independiente sino que existía una interacción entre ellos. Al tomar la comprensión de lectura como un proceso interactivo, es decir, un diálogo entre el lector, el autor y el texto, se planteó la importancia del lector al momento de leer un texto, en esta instancia se otorgó a los conocimientos previos del lector un papel fundamental en la interpretación del texto leído, dando paso a la relación antes mencionada interacción lector - autor.

### **Modelos de comprensión lectora**

Debido a la complejidad del proceso lector se han desarrollado modelos explicativos, y por supuesto, derivar algunas líneas de trabajo, que permitan a su vez, abordarlos para favorecer los procesos lectores en los sujetos. Estos modelos, con sus diferencias, coinciden en que la comprensión lectora es un proceso multinivel, que va desde el reconocimiento de los grafemas,

hasta concebir el texto como un todo, ligado incluso, a las prácticas sociales y culturales de contextos determinados. Sin embargo, aun no hay acuerdo sobre cómo estos niveles se relacionan entre sí, y los modos en que afectan los procesos de lectura. La caracterización de los mismos reside en nombres como ascendente, descendente e interactivo (Alonso et al, 1985).

El **modelo de procesamiento ascendente** está basado en el procesamiento de la lectura a partir de los componentes lingüísticos más básicos como lo son las letras, las palabras, las frases en un proceso que se da de manera ascendente, permitiéndole al lector comprender el texto; de esta manera, se exige al lector una adecuada competencia descodificadora donde haya consolidado de una manera adecuada las reglas de correspondencia grafema-fonema para que los recursos de memoria sean utilizados en el proceso de comprensión (Vallés, 2005). Este modelo propone la existencia de distintos niveles que conservan una dependencia unidireccional donde cada nivel actúa como prerrequisito para el siguiente nivel; de esta forma se concibe que el proceso de comprensión inicia con el reconocimiento visual de las palabras hasta el procesamiento semántico del texto como un todo, pero la relación inversa puede no ser necesaria. De este modo, puede afirmarse que quien comprende un contenido es porque tiene la capacidad de decodificar el mismo.

El principal exponente de este enfoque es el Modelo Proposicional de Kintsch (Crowder, 1985; Duro, 1992) quien fundamenta su modelo de procesamiento sobre estructuras proposicionales existentes en los textos. En este sentido el proceso de lectura se da en cuatro fases (Tapia, 2005). En primer lugar, ocurre la identificación de los patrones gráficos, se fija la vista en pequeños saltos que dan como resultado el reconocimiento y el significado de las palabras; en la segunda fase, se identifica la familiaridad que la persona tiene con los patrones gráficos identificados en la fase anterior, lo que lleva a considerar que las palabras o grupos de

letras asociados con mayor frecuencia se leen de manera más rápida. En tercer lugar, se lleva a cabo el reconocimiento del léxico donde juegan variables como la cantidad de vocabulario que maneja el lector, la rapidez para acceder al mismo y el contexto en el que aparece cada palabra, para permitir con ello, realizar una inferencia del significado de la misma. En último lugar, se encuentra el acceso del léxico, éste se relaciona estrechamente con la cantidad de vocabulario que conoce el lector, la familiaridad con las palabras que lee y el contexto en el que aparecen las mismas

El **modelo de procesamiento descendente** se sustenta en que el procesamiento cognitivo que se realiza en la lectura se da desde el aporte previo de los conocimientos del lector sobre la lectura y el reconocimiento global de las palabras (Vallés, 2005); es así como los "buenos" lectores, al interpretar el significado del texto, se sustentan más en sus conocimientos sintácticos y semánticos adquiridos previamente que en los detalles gráficos que brinda el texto.

Sandford y Garrod (Crowder, 1985; Duro, 1992) apoyados en este planteamiento, propusieron el Modelo de Memoria o Marco de Procesamiento donde planteaban que el proceso de comprensión se fundamenta en las estructuras de conocimiento almacenadas en la memoria del lector; en la memoria se almacenan escenarios que son estructuras de conocimiento con un contenido representacional organizado y su función es proporcionar información que sirva de base para la comprensión, en este sentido, la comprensión se concibe como la asimilación del contenido del texto a los escenarios presentes en el lector.

En estos dos modelos, procesamiento ascendente y descendente, es claro que el énfasis de la comprensión de lectura se hace en el texto y en el individuo, respectivamente. En un estudio realizado por Duro (1992) donde se contrastaban estos dos modelos en tareas relacionadas con el procesamiento de los textos como un indicador de la comprensión de lectura, se evidenció que

los procesos cognitivos subyacentes a la comprensión de textos son procesos que recuperan y actúan sobre estructuras de conocimiento almacenadas en la memoria tal como lo plantea el Modelo de Memoria de Sandford y Garrod.

Finalmente, se han propuesto **modelos interactivos** que afirman que la comprensión está dirigida de manera simultánea por los datos explícitos brindados por el texto y por los conocimientos previos del lector, de esta manera, los conocimientos previos facilitan la asimilación de la información de orden inferior consistente con las expectativas del lector y los datos del texto aseguran que el lector esté atento a la información nueva o que no encaje con sus hipótesis previas; en este modelo el significado está dado por el lector, quien lo construye de manera activa recurriendo a sus conocimientos y a su motivación al momento de interactuar con el texto. De esta forma, la comprensión es el resultado de la interacción entre el texto, los esquemas de conocimientos del lector y el contexto, bien sea lingüístico, actitudinal o situacional. Desde una perspectiva cognitiva (Vallés, 2005), la comprensión ha sido considerada como un producto y un proceso; como producto se entiende la interacción entre lector y el texto, almacenándose en la memoria a largo plazo que después evocara al formular las preguntas sobre el texto que ha sido leído; de este modo, la memoria a largo plazo y las rutinas de acceso a la información son vitales y determinan el éxito que puede tener el lector en la comprensión de un texto.

Desde este modelo interactivo, autores como Goodman (1967; 1982) sostienen que toda lectura es una interpretación y que lo que el lector es capaz de comprender depende fuertemente de sus conocimientos y creencias; de esta manera, es posible que dos personas que lean un mismo texto puedan diferir en lo que comprenden del mismo, dadas las contribuciones que cada uno dé al significado del texto se puede interpretar sobre la base de lo que se conoce.

Adicionalmente, Goodman afirma que aprender a leer involucra un desarrollo de estrategias para adquirir el sentido del texto y el desarrollo de esquemas acerca de la información que es representada en el texto, esto ocurre solamente cuando los lectores principiantes responden a textos significativos que son interesantes y tienen sentido para ellos; así como se necesita saber a dónde ir cuando se inicia un viaje, el lector inicial necesita conocer y utilizar los elementos claves de las pautas de una oración como los nexos y sufijos gramaticales, así como la puntuación para predecir las pautas sintácticas.

Este autor ha sostenido que la lectura es una conducta inteligente y se da como un proceso cíclico que inicia con un ciclo óptico, luego pasa por los ciclos perceptual y gramatical, terminando con el ciclo de búsqueda del significado. Este último ciclo es el más importante en el proceso de lectura porque el significado se construye mientras se lee, pero también es reconstruido, ya que la nueva información debe ser resignificada constantemente. En este sentido, se hace evidente que la lectura es un proceso esencialmente interactivo donde es el lector quien da sentido y significado al texto, y donde éste otorga lo suyo -el contenido- como creación necesaria para esa vivencia que se origina en el acto de leer.

En este proceso de comprensión suelen distinguirse dos niveles, el primero es un nivel elemental que se relaciona con la comprensión de las proposiciones del texto, el segundo nivel se refiere a la integración de la información proporcionada por el texto a través de cómo el lector va relacionando las diversas proposiciones del nivel elemental dando pleno sentido a lo que se lee; de este modo, del lector depende el sentido de un texto y es por esto que la comprensión es una interpretación de lo que se lee; esto sucede porque, a diferencia del lenguaje oral, el autor no está presente para hacer preguntas sobre el texto, por eso, para que se dé una comprensión significativa del texto es necesario que el lector posea conocimientos sobre la estructura de los

textos y que disponga de conocimientos generales y específicos de aquellos contenidos que lo acercan al acto lector.

Siguiendo por esta línea De Zubiría (1996) afirma que leer no consiste en descifrar pasivamente los signos impresos, el buen lector descifra los textos de manera activa donde procesa el símbolo gráfico con los sentidos, luego se interioriza por medio del entendimiento y la razón, luego vuelve hacia el texto con el significado que el lector ha dado a los símbolos gráficos. Para que este proceso se dé de manera exitosa el lector requiere de ideas y conocimiento previo para que pueda producirse un pensamiento al respecto. Por ello, desde los diversos niveles de la educación inicial y básica se inicia la construcción de este proceso lector. En la educación preescolar se inicia el desarrollo de las habilidades de lectura fonética; en primaria se enfatiza en los decodificadores primarios y secundarios, los terciarios requieren de operaciones formales que se extienden en la educación básica secundaria, los mecanismos decodificadores complejos como los metasemánticos y categoriales continúan su evolución durante la formación universitaria y de posgrado. De esta manera, el proceso lector se resalta desde un plano cognitivo pero, con una incidencia del componente emocional: motivación, intereses y propósitos del lector se vinculan al texto, y como el contenido de éste -también- afecta los indicios emocionales del lector. Con esto, se subraya que la comprensión de lectura es un proceso dinámico y en continua evolución, que se da en periodos críticos de la formación académica del individuo y que continúa cualificándose a lo largo de toda la vida.

Continuando con estas tramas de significado en torno a la lectura y su comprensión, aparece Golder et al., (2002) cuya apuesta se centra en definir la lectura como una actividad que es familiar para el que lee pero que implica procesos cognoscitivos y perceptivos que son complejos. Desde este autor, podría decirse que la lectura es la recepción y la comprensión de las

ideas de un autor a partir de signos visuales; Golder, defiende la idea de que leer es ante todo descodificar debido a que este proceso debe ser automatizado para que se faciliten los procesos de alto nivel.

Finalmente van Dijk (1983, 1992) muestra un panorama interdisciplinario en el abordaje de la comprensión de lectura situando a la psicolingüística y a la psicología cognitiva como las disciplinas que pretenden explicar el funcionamiento de la lectura como un sistema lingüístico abstracto estableciendo la existencia de procesos cognitivos en la producción y comprensión de textos.

En su planteamiento menciona la existencia de **macroestructuras** definidas como las conexiones que se basan en el texto como un todo y no solo las conexiones entre oraciones aisladas y sus proposiciones, así, las macroestructuras son de orden semántico ya que aportan una idea de la coherencia global y del significado del texto, perteneciente a un nivel superior que al de las proposiciones por sí solas. La aplicación de **macrorreglas** de construcción permite sustraer del texto aquella información que es innecesaria para la comprensión del mismo. Dichas macrorreglas son a su vez, *la supresión*, cuya función es suprimir dicha información, *la generalización* puede aplicarse a un grupo de proposiciones que pueden derivar en una sola proposición sin perder el sentido del texto, *la construcción* permite sustituir una secuencia de hechos que pueden contenerse implícitamente en una sola proposición. La aplicación frecuente de estas macrorreglas según este autor, permite alcanzar distintos niveles de abstracción y generalidad, no obstante, estos principios de comprensión textual, se encuentran bajo el control de los esquemas y las superestructuras de los contenidos que se actualizan en narraciones argumentaciones y/o exposiciones.

Los **esquemas** se definen como unidades de información de carácter general que representan características de objetos, situaciones y hechos, jugando un papel importante en la organización y recuperación de la información, estos esquemas son activados durante la lectura dependiendo de los objetivos de la misma, el contexto y la información aportada por el texto. Por su parte, las **superestructuras** representan formas fijas de organización del texto en un conjunto de categorías funcionales entre los que se encuentran las narraciones, los cuentos, los informes de investigación, las argumentaciones y las exposiciones; de este modo, la superestructura se considera como un esquema abstracto al cual el texto se adapta, estableciendo el orden global del mismo que se compone de un serie de categorías cuyas combinaciones están basadas en reglas convencionales. Entre otras se encuentra la superestructura argumentativa que se compone de una introducción, una hipótesis, los argumentos que apoyan la hipótesis y la conclusión.

En síntesis, puede definirse la lectura como una actividad motivada, orientada a una finalidad y cuyo resultado depende de la interacción entre las características del texto y la actividad del sujeto que lo lee, es decir el intercambio que se establece entre sus conocimientos previos, propósitos y expectativas. En este sentido la comprensión puede concebirse como el objetivo fundamental de casi toda lectura ya que ésta no es un proceso que se da después del reconocimiento visual de las palabras sino que se da de forma paralela.

### **Procesos psicológicos involucrados en la comprensión de lectura**

Como se mencionó anteriormente, el enfoque cognitivo concibe a la comprensión de lectura como un proceso y un producto en el que es necesario considerar la interacción que el lector tiene con el texto; en este sentido Vallés (2005) afirma que para comprender un texto se ponen en marcha una serie de procesos perceptivos, procesos psicológicos básicos y procesos cognitivo lingüísticos.

Los **procesos perceptivos** son los encargados de obtener la información textual para transmitirla a las estructuras cerebrales que realizarán su procesamiento. La entrada de la información se da generalmente de manera visual, cuando este sentido se ve afectado la entrada de la información se da por medio de canales auditivos o táctiles; en el proceso lector se produce una discriminación de tipo visual que consiste en seleccionar grafías del repertorio existente en la memoria con el fin de identificar correctamente los grafemas que se leen y evitar así errores en la exactitud con que se leen las palabras, este proceso contribuye de manera significativa a lograr una buena comprensión lectora.

En esta *Percepción Visual* cobran importancia los movimientos oculares denominados movimientos sacádicos se manifiestan en la fijación y deslizamiento de la vista en las palabras del renglón que se está leyendo, se producen breves pausas en cada fijación lo que permite una focalización visual para identificar periféricamente las letras y palabras alrededor del punto de focalización.

Los **procesos psicológicos básicos** están relacionados con el reconocimiento de las palabras y su asociación a conceptos almacenados en la memoria, el desarrollo de las ideas significativas, la extracción de conclusiones y la relación entre lo que se lee y los conocimientos del lector; todos esto exige un procesamiento múltiple de información ya que hay una demanda de atención activa a muchas fuentes simultáneamente y está relacionada con la experiencia que pueda tener el lector, cuando esta capacidad es limitada se presentan dificultades en la comprensión de lectura.

Entre los procesos psicológicos que intervienen en la comprensión se encuentran la *Atención Selectiva* que le permite al lector focalizar su atención en el texto y rechazar cualquier otro estímulo tanto interno como externo que lo pueda distraer en su lectura; el *Análisis*

*Secuencial* donde el lector realiza una lectura continua palabra por palabra y va conectando los significados de cada una para luego dar significado, mediante las inferencias lingüísticas, a la secuencia leída en el texto; en el proceso de *Síntesis* el lector resume y atribuye significado a las unidades lingüísticas para que las palabras leídas se articulen en una unidad coherente y con significado, de esta manera se logra una comprensión del texto. Sin embargo, para que la comprensión sea eficaz es necesario el proceso simultáneo de *Análisis-Síntesis* ya que mientras se lee se produce un reconocimiento y una decodificación de las letras, luego estas letras se integran en sílabas, palabras, frases, párrafos y se otorga un significado al texto leído, dándose un proceso interactivo entre análisis y síntesis.

Otro proceso fundamental en la comprensión de lectura es el almacenamiento de la información en la *Memoria a corto y largo plazo* (Lorenzo, 2001). En la memoria a corto plazo se activan los mecanismos de asociación, secuenciación, linealidad y recuerdo del texto, siguiendo una disposición lógica y estructurada a medida que se va leyendo el texto, por otra parte, en la memoria a largo plazo se van estableciendo vínculos de significados con los conocimientos previamente adquiridos por el lector, construyendo así aprendizajes significativos. Para que se dé un proceso de comprensión de lectura es necesaria la interconexión de estos dos tipos de memoria, ya que la información procedente del texto es captada por la memoria sensorial que la almacena por un breve lapso de tiempo y la memoria de trabajo selecciona la información para almacenarla en la memoria a corto plazo durante algunos segundos, pasando posteriormente a la memoria a largo plazo mediante un mecanismo de integración de la memoria de trabajo; así mismo, el proceso de evocación de los conocimientos previos es realizada por la memoria de trabajo logrando que mediante las correspondientes inferencias se dé una comprensión parcial del texto a medida que el lector va realizando la lectura, para finalmente

depositar el producto comprensivo en la memoria a largo plazo. Este proceso interactivo entre la memoria a corto y largo plazo realizado por la memoria de trabajo se da de manera constante durante todo el proceso de lectura.

Finalmente, los **procesos cognitivo lingüísticos** le permiten al lector acceder al significado y extraer la información necesaria del texto para lograr la comprensión. Entre estos procesos se encuentra el *Acceso al léxico* que permite recobrar la información semántica y sintáctica almacenada en la memoria a largo plazo y que puede estar organizada según el uso frecuente de las palabras, su significado o su estructura silábica, lo que permite conseguir las relaciones estructurales entre las palabras que constituyen las frases, con el fin de obtener la información contenida en el texto. Luego de acceder al léxico, se accede a las relaciones estructurales entre las palabras que conforman las frases para obtener la información contenida en el texto denominado *Análisis Sintáctico*, las diferentes estructuras gramaticales están sintácticamente relacionadas a través de la secuencia de palabras lo que permite obtener la información del conjunto de proposiciones o frases, este proceso se lleva a cabo mediante la identificación de señales lingüísticas como el orden y estructura de la frase, el acceso sintáctico inmediato y la memoria de trabajo. Después de estas dos procesos debe darse la comprensión del texto, esto se logra mediante la *Interpretación Semántica* que consiste en producir inferencias, es decir, obtener información nueva del texto basándose en la interpretación de la lectura y de acuerdo con el contexto; este proceso se da en el mismo momento en que se realiza la lectura ya que se producen evocaciones, se verbalizan encubiertamente pensamientos, se accede a los significados de las palabras, etc, y también al finalizar la lectura porque se producen evocaciones, verificación de la información y contestaciones a cuestiones interpretativas de la lectura, entre otros.

Si bien existe una amplia revisión y discusión sobre los procesos que se dan en la comprensión de lectura, la intención de esta investigación es realizar un énfasis en la evaluación y de las diferentes aproximaciones que permitan su comprensión.

### **Evaluación de la comprensión de lectura**

La conceptualización de la comprensión de lectura ha sufrido varios cambios en el último siglo como se ha evidenciado en apartados anteriores, sin embargo, dados los avances e investigaciones sobre los procesos cognitivos implicados en la comprensión de lectura, puede considerarse que existe una interacción entre el lector y el texto, donde están implícitas estrategias de razonamiento que debe hacer el lector que le permiten construir el significado del texto a partir del texto, utilizando sus conocimientos previos y realizando inferencias de lo que no está explícito en el texto pero que le permite realizar una comprensión de lo que se lee (Perez, 2005).

Esta consideración es vital al momento de definir cómo y de qué manera se realizará la evaluación de la comprensión; al respecto autores como Alonso (1985) y Montanero (2004) menciona que la mayoría de estrategias de medición se han centrado en el *producto* o rendimiento, como se evidencia en el uso de las pruebas objetivas, más que en el *proceso* que se da en la comprensión, sin embargo, la utilización de alguna estrategia para la evaluación de la comprensión debe estar relacionada directamente con los objetivos de la evaluación.

Montanero (2004) propone cuatro niveles de comprensión de lectura que pueden ser evaluados y a su vez propone algunas estrategias para realizar la evaluación como proceso y como producto.

El primer nivel de comprensión hace referencia a la *Comprensión local y referencial* que se da a nivel léxico y sintáctico donde se realizan inferencias relacionadas con la conexión

sintáctica entre las diferentes oraciones que permite generar un hilo conductor entre las mismas, por ejemplo, busca que el lector reconozca y recuerde los detalles del texto ligados a nombres de los personajes, ideas centrales y secundarias, lugares, características de personajes primarios y secundarios, ideas explícitas del autor, personajes que intervienen en los distintos momentos de la narración, relaciones de igual manera explícitas que se establecen entre los personajes, entre otros eventos (Pérez, 2005).

Este nivel de comprensión puede evaluarse a través de pruebas objetivas mediante preguntas breves con opciones de respuesta cerradas donde se indague sobre la capacidad de seleccionar correctamente el antecedente de una proposición conectada referencialmente con otra anterior, solicitar al evaluado el parafraseo de alguna oración con el fin de evaluar si logra relacionar dos proposiciones, ofrecer varias oraciones que no aparecen literalmente en el texto para que el evaluado elija la opción que se ajusta a la idea que esta explícitamente en el texto, entre otras. Algunas de las limitaciones que conlleva el uso de estas pruebas objetivas tiene que ver con el proceso inferencial que interviene en la comprensión del hilo conductor porque éste se da durante la lectura y las preguntas son formuladas una vez el evaluado termina la lectura, por lo que es difícil evaluar si realmente se ha realizado una comprensión del texto o si simplemente se debe a un recuerdo inmediato, así mismo es difícil discriminar si el evaluado realiza las inferencias espontáneamente o si es un producto de la demanda a posteriori cuando se responde la pregunta o si la respuesta se da por el simple reconocimiento y ubicación en el texto.

Otras estrategias que pueden utilizarse para evaluar el proceso que se lleva a cabo en este nivel son la utilización del pensamiento en voz alta y el registro de observación que consiste en grabar y analizar las reflexiones que realiza el evaluado en voz alta durante la lectura de un texto, sin embargo, esta estrategia conlleva algunas limitaciones en su utilización como lo son la

cuantificación objetiva de lo mencionado por el evaluado, la necesidad de entrenamiento previo del evaluado corriendo el riesgo de que se realicen más inferencias de las que se realizarían en condiciones naturales.

En segundo lugar se encuentra la evaluación de la *Comprensión global y estructural* que se fundamenta en la actividad mental que el evaluado desarrolla para asignar un significado global al texto leído. La estrategia más utilizada para evaluar este nivel es mediante pruebas objetivas, donde se busca que el evaluado realice las inferencias sobre el significado global del texto o la distinción de ideas jerárquicamente diferentes, también puede plantearse la opción de elegir el título que condense la idea del texto. Una estrategia para evaluar el proceso de la comprensión global tiene que ver con el análisis del contenido de resúmenes, protocolos de recuerdo o mapas conceptuales, cuyo objetivo principal es evaluar la representación semántica generada por el evaluado utilizando diversos métodos de análisis del contenido proposicional del resumen que ha elaborado, sin tener acceso al texto que se leyó.

En tercer lugar, la evaluación de la *Comprensión pragmática y elaborativa* busca que el lector pueda unir el texto con su experiencia personal con el fin de poder realizar conjeturas e hipótesis y se espera que tenga la capacidad de deducir ideas, detalles de los personajes o características que no formula el texto y que no se encuentran de modo explícito en el mismo (Pérez, 2005); de este modo, el evaluado hace uso de sus conocimientos previos, inferencias y vivencias para realizar una interpretación de la nueva información. En las pruebas objetivas se utilizan preguntas que indagan sobre las intenciones comunicativas, la interpretación de una metáfora o moraleja, el género literario del texto o los recursos estilísticos del autor en el texto; todas éstas tienen en común la necesidad de desplegar inferencias elaborativas más allá del texto que se presenta al evaluado. En estudios realizados por Montanero (2004) se encontró que, al

realizar la evaluación mediante este tipo de pruebas, no existen diferencias significativas en las respuestas de los evaluados entre responder las preguntas teniendo a mano el texto y sin tener acceso a él.

También han sido utilizadas estrategias como el pensamiento en voz alta y las tareas de solución de problemas con textos inconsistentes para evaluar este nivel, basándose en la presentación de textos que contienen afirmaciones contradictorias entre sí o con respecto a los conocimientos previos que se sabe tiene el evaluado; se realiza mediante preguntas abiertas de manera oral o escrita donde el evaluado realiza un autoinforme durante la lectura.

Finalmente, se encuentra la evaluación de *Estrategias de regulación metacognitiva y apoyo a la comprensión de textos expositivos* que tiene relación, de manera indirecta, con la comprensión de textos académicos y se trata de la evaluación del uso de estrategias o hábitos de estudio que tienen un efecto en las capacidades de atención sostenida que está involucrada en la comprensión de textos expositivos. Las estrategias de evaluación utilizadas para este nivel son sensiblemente diferentes a las expuestas anteriormente y están basadas principalmente en técnicas de autoinformes y registro de observaciones, donde el evaluado responde una serie de preguntas sobre sus hábitos de estudio.

Como se mencionó al inicio de este apartado, la escogencia o el uso de alguna de las estrategias para evaluar estos niveles de comprensión debe guardar relación con los objetivos de la misma, reconociendo así mismo las ventajas y desventajas de cada una. Para cumplir los objetivos de esta investigación, se tomará la comprensión de lectura como un producto susceptible de ser evaluado mediante pruebas objetivas que permita realizar una evaluación masiva mediante test análogos (de lápiz y papel) o digitales.

### **Diseño y construcción de bancos de ítems**

En los últimos 60 años, y gracias a las nuevas tecnologías de la información y las comunicaciones, se han evidenciado progresos en la construcción y aplicación de test. Gracias a ésta evolución los bancos de ítems han cobrado una importante relevancia por ser un componente básico en la utilización de Test Adaptativos Informatizados (TAI's), necesitándose así, una cantidad de ítems cuyas propiedades sean conocidas para seleccionar el más apropiado en cada caso logrando así una adaptación continua a las respuestas de los evaluados y una medida confiable de su habilidad (Molina, Sanmartín, & Pareja, 2000).

En este apartado se mencionarán algunas consideraciones sobre el diseño y planeación de un banco de ítems, la construcción de ítems y el análisis estadístico bajo el Modelo de Rasch.

#### **Diseño y planeación de un banco de ítems**

Para evaluar conocimientos, habilidades y competencias existen varios tipos de instrumentos, el más utilizado es la prueba objetiva, definida como un conjunto de ítems intencionalmente articulado que permite una medición sistemática de una propiedad o atributo permitiendo su cuantificación, con los resultados obtenidos en la aplicación de dicha prueba se puede inferir el desempeño de los evaluados en el atributo medido (Brown, 1980; ICFES, 2009).

De este modo, una prueba objetiva es más que la suma de los ítems que la conforman, aunque su poder evaluativo proviene de la potencia con que evalúan todos y cada uno de ellos; siendo así, este tipo de pruebas están conformadas por preguntas cerradas o estructuradas presentadas de manera organizada, con unas opciones de respuesta posible y unas instrucciones precisas para que el evaluado pueda elegir la opción que considere la más acertada, conformando

así una unidad evaluativa completa y armónica que responda efectivamente a los propósitos y objetivos de la evaluación (ICFES, 2009).

Algunas de las ventajas en la utilización de este tipo de instrumentos de medición (Herrera, 1996) se relacionan con la posibilidad de minimizar el juicio subjetivo del calificador, ya que se cuentan con criterios previamente establecidos para su calificación; la aplicación de la prueba puede controlarse al máximo, lo que permite que los evaluados sean sometidos, en la medida de lo posible, a las mismas condiciones para responder la prueba garantizando que el proceso se lleve en condiciones de igualdad; son útiles para la evaluación de grupos grandes de evaluados y puedan incluirse un número amplio de preguntas, permitiendo que la aplicación y la obtención de resultados se dé en un tiempo corto con una baja tasa de error y una adecuada precisión en la medida del atributo que se desea evaluar. Sin embargo, este uso es limitado cuando se pretende evaluar aspectos como la capacidad de producción de textos, realizar juicios, opiniones o críticas de algún tema particular.

El proceso de construcción de este tipo de pruebas requiere planeación, organización y permanente revisión de los ítems que la conforman, como se mencionó anteriormente, no basta con la agregación de ítems aunque aparentemente tengan relación con el objetivo de la evaluación. Construir buenos ítems implica, a su vez, conocer y aplicar aspectos técnicos en su formulación como las reglas existentes para su correcta elaboración, los factores que pueden alterar su validez y las características de los distintos tipos de preguntas que pueden ser utilizadas; por estas consideraciones, la tarea más compleja del proceso evaluativo es la construcción de ítems que se ajusten a las especificaciones de la prueba y que respondan efectivamente a los propósitos de la evaluación (Hambleton & Zaal, 1994).

Dada la importancia de lo mencionado anteriormente, la construcción de una prueba objetiva implica una serie de fases que deben cumplirse; algunos autores como Thorndike (1995); Moreno, Martínez, & Muñiz (2004); Morales (2006); Herrera (2009), e instituciones como AERA; APA; NCME (1999) e ICFES (2009) han propuesto una serie de fases para la construcción de instrumentos objetivos de evaluación, así como algunas especificaciones psicométricas que deben ser tenidas en cuenta, como son la definición del atributo a evaluar, los componentes de este atributo, las competencias que debe tener el evaluado para responder la prueba, la definición de la estructura de prueba, la escogencia de un grupo de expertos para la construcción de los ítems, los tipos de preguntas que se utilizarán, las condiciones en las cuales se realizará el proceso de aplicación y los parámetros de calificación.

### ***Definición del atributo***

En primer lugar, previo a cualquier construcción de ítems, para la construcción de un banco de ítems se debe definir conceptualmente y/o teóricamente qué se va a evaluar, orientando así todo el proceso de construcción de los ítems y de la prueba en general. Esto permite determinar el objetivo de la evaluación, el contenido temático bien sea por componentes o bloques temáticos, la profundidad y dificultad con la que se hará la evaluación del contenido temático, la dimensión cognitiva que se evaluará que pueden ser competencias, conocimientos, aptitudes, actitudes, entre otros; a su vez, debe identificarse claramente la población a la que irá dirigida la prueba, teniendo presentes la edad, el nivel educativo, el contexto sociocultural, entre otros. La información contenida en esta fase guiará al equipo de construcción y validación y le brindará información a los evaluados sobre el contenido de la evaluación.

### ***Componentes y competencias a evaluar***

Los componentes de la prueba hacen referencia a los contenidos temáticos o disciplinares que serán evaluados y deben estar en perfecta relación con la definición del atributo; cada componente debe tener un número mínimo de ítems que permita realizar adecuadamente una caracterización del mismo desde un punto de vista estadístico y del contenido disciplinar.

Por otro lado, las competencias han sido entendidas como un saber situado o un saber hacer en un contexto determinado y que son puestas en acción en una tarea o resolución de problemas, de este modo, se privilegia lo que el evaluado hace con sus conocimientos cuando está en interacción con contextos específicos y la comprensión, análisis y crítica de los mismos (ICFES, 2009)

### ***Estructura del banco de ítems***

Una vez definido el atributo, las competencias y los componentes se realiza la estructura de prueba que es, a su vez, la directriz para la redacción de las preguntas, y sobre ella debe basarse la construcción de las mismas. Se resume, generalmente, en una Tabla de varias entradas donde se resumen las características importantes de la prueba como lo son: la *dimensión disciplinar*, donde se mencionan los contenidos que se abordarán organizado en áreas o componentes, y las *dimensiones cognitivas* que se evaluarán como aptitudes, conocimientos o competencias, entre otros., así como la *cantidad de preguntas* por cada componente y competencia, como se observa en la Tabla 2.

Tabla 2 Ejemplo de una estructura de prueba

| <b>Dimensión<br/>Disciplinar</b> | <b>Dimensión Cognitiva</b> |             |          | <b>Total<br/>preguntas</b> |
|----------------------------------|----------------------------|-------------|----------|----------------------------|
|                                  | Argumentar                 | Interpretar | Proponer |                            |
| Lenguaje                         | 5                          | 5           | 5        | 15                         |
| Matemáticas                      | 5                          | 5           | 5        | 15                         |
| Ciencias                         | 5                          | 5           | 5        | 15                         |
| Número total de preguntas        |                            |             |          | 45                         |

### ***Escogencia del grupo de Autores de preguntas***

Para la construcción de una prueba y de los ítems que la conforman se debe contar expertos en la teoría psicométrica, en técnicas de medición, evaluación y construcción de instrumentos, así como con un grupo de personas que sean especialistas en tema o atributo que será medido; estos expertos deben contar con algunas características deseables en un grupo de constructores, como lo son: el conocimiento de las variables educativas y psicológicas relevantes de la población a la que va dirigida la prueba; tener un alto grado de experticia en la dimensión disciplinar que se va a evaluar y en las diferentes aproximaciones teóricas, filosóficas, epistemológicas, entre otras; manejar una adecuada comunicación verbal que le permita expresar lo que desea de una manera sencilla y entendible para el evaluado; conocer y manejar los diferentes tipos de preguntas que se utilizarán en la evaluación; ser creativo y utilizar su ingenio en la creación de situaciones que requieran exactamente el conocimiento que se desea evaluar; poseer habilidades de trabajo en equipo, de manera que se sienta cómodo en el momento de someter su trabajo al juicio de sus compañeros y que pueda revisar con sentido crítico y constructivo el trabajo de los demás miembros del grupo; mantener el compromiso de no divulgar la información relativa a la conformación de los ítems y del material de la prueba al que tenga acceso.

### ***Tipos de pregunta***

Los tipos de pregunta corresponden a la forma en la cual se estructurarán las preguntas; en la Tabla 3 se muestran los diferentes tipos de preguntas que ha utilizado el ICFES (2009) en sus procesos de evaluación.

Tabla 3 Tipos de preguntas utilizadas en la evaluación objetiva (ICFES, 2009)

| <b>Tipo de Pregunta</b>                       | <b>Descripción</b>   |
|---|--|
| Selección Múltiple con Única Respuesta        | Consta de un enunciado y cuatro opciones de respuesta (A, B, C y D). Sólo una de estas opciones responde correctamente a la pregunta.  |
| Selección Múltiple con Múltiple Respuesta     | Constan de un enunciado, un problema o un contexto, a partir del cual se plantean cuatro opciones numeradas del 1 al 4. El evaluado debe seleccionar la combinación de dos opciones que mejor responda a la pregunta.        |
| Selección Múltiple con información suficiente | Este tipo de preguntas se basan en un problema y dos conjuntos de información denominados I y II. No se requiere que se solucione el problema, sino que se decida que información es suficiente y necesaria para resolverlo. |
| Análisis de Postulados                        | Consta de una tesis y dos postulados verdaderos. El evaluado debe decidir si los postulados se deducen lógicamente de la tesis.  |
| Análisis de Relación                          | Constan de dos proposiciones, así: una afirmación y una razón, unidas por la palabra PORQUE. El evaluado debe examinar la veracidad de cada proposición y la relación teórica que las une.                                   |

En este apartado se mencionarán las características y las reglas de construcción para las preguntas de Selección Múltiple con Única Respuesta debido a que es el tipo de pregunta más utilizado en aplicaciones masivas, permitiendo una facilidad en la aplicación y calificación, y las consideraciones para su construcción se ajustan a los demás tipos de pregunta.

Este tipo de pregunta consta de tres componentes fundamentales: el contexto, el enunciado y las opciones de respuesta. El *contexto* está compuesto por la información necesaria para dar solución a un problema; este puede ser un texto, una gráfica, una Tabla, un dibujo a partir del cual se deriva la pregunta. El *enunciado* plantea la problemática que será resuelta por el evaluado y puede ser una pregunta, una frase incompleta, un texto o una gráfica; en ocasiones el enunciado está contenido dentro del contexto y en otras se presenta de manera independiente haciendo referencia a la información contenida en el contexto. Las *opciones de respuesta* son las posibles, generalmente aparecen identificadas con letras mayúscula A, B, C y D y solo una de las opciones completa o responde correctamente el enunciado (conocida como clave), las demás opciones (llamadas distractores) no responden de forma completa o pertinente a la problemática planteada pero deben resultar atractivas para los evaluados que no dominan el tema evaluado.

### **Consideraciones para la construcción de ítems**

Como se ha mencionado anteriormente, el proceso de construcción de ítems es un trabajo dispendioso y requiere de unas habilidades y destrezas en el constructor; varios autores como Morales (2006); Moreno, Martínez, & Muñiz (2004); Cheung (2002); Haladyna & Downing (1989) manifiestan la necesidad de contar con algunas consideraciones que sirvan como guía en el proceso de construcción de ítems.

A continuación se mencionan una serie de consideraciones generales, así como, consideraciones específicas para la construcción del enunciado y las opciones de respuesta.

### *Consideraciones generales*

Al momento de planear la construcción de los ítems es recomendable contar con un tiempo suficiente para realizar la construcción y la validación de los mismos; es importante verificar que en el documento de la prueba no existan errores gramaticales, de puntuación, ortográficos, entre otros; es vital utilizar un lenguaje claro, preciso y directo que permita minimizar el tiempo que el evaluado demora en la lectura de los ítems; es imprescindible incluir preguntas que evalúen todos los niveles de habilidad (alta, media, baja) de la población a la que va dirigida la prueba; se recomienda evitar el uso de textos extensos y de preguntas que puedan favorecer o perjudicar a determinados tipos de evaluados.

### *Consideraciones para la elaboración de los enunciados*

En cuanto a la elaboración de enunciados es importante que el ítem suministre una idea clara y completa de la tarea que se está evaluando y que contenga la información necesaria y suficiente para que el evaluado pueda responder el ítem; debe evitarse incluir información trivial o poco relevante que pueda distraer al evaluado, de manera que la idea central se encuentre en el enunciado y no en las opciones de respuesta; es necesario que cada ítem corresponda a una tarea, tal como se ha definido en la estructura de prueba, y que sean independientes o que no brinde información para responder otros ítems; se deben evitar ítems que incluyan posturas ideológicas o contengan prejuicios y que indaguen sobre la opinión del evaluado, así como ítems que evalúen conocimientos muy específicos y aquellos cuya respuesta válida se determine solo bajo el criterio del constructor; también se aconseja redactar el enunciado evitando el uso de negaciones, en caso de ser necesario su uso, debe registrarse en mayúsculas y negrilla (**NO**); también debe evitarse el uso de nombres propios para referirse a personas o lugares

### ***Consideraciones para la elaboración de las opciones de respuesta***

Estudios previos (Haladyna & Downing, 1989) han analizado la cantidad de opciones de respuesta y se ha propuesto que cuatro (4) es un número adecuado; la opción correcta o clave debe diferenciarse de los distractores por su contenido y no por sutilezas del lenguaje, del mismo modo, la elaboración de los distractores pueden basarse en una afirmación verdadera pero que no contesta a la pregunta planteada o no satisface las exigencias del problema planteado y debe evitarse la inclusión de distractores fácilmente descartables, ya sea por obvios o por ser de otro orden semántico o gramatical, adicionalmente es necesario justificar de manera suficiente y adecuada cada una de las opciones de respuesta, con el fin de garantizar que solo hay una opción válida y las demás son atractivas para los evaluados que no dominan el tema.

Otras recomendaciones relacionadas con la forma de presentación de la clave y los distractores se relacionan con no repetir en las opciones de respuesta palabras o frases contenidas en el enunciado, verificar que todas las opciones de respuesta tengan la misma longitud, evitar la utilización de opciones como *Todas las anteriores* o *ninguna de las anteriores*, evitar el uso de adverbios como *Totalmente*, *Siempre* o *Nunca*, finalmente, es recomendable que la clave aparezca proporcionalmente en cada posición posible (A, B, C o D).

### ***Características de las preguntas***

Además de las consideraciones expuestas anteriormente, se recomienda tener en cuenta algunas características o aspectos que deben revisarse en el momento de la construcción de los ítems, y previo al ensamblaje y la aplicación de la prueba; estos aspectos resultan de mayor o menor importancia dependiendo del tipo y objetivo de la prueba:

El primero de ellos tiene que ver con *Aspectos formales* que se debe tener en cuenta la redacción, la precisión de las instrucciones, claridad en la presentación de la tarea, uso de un

lenguaje adecuado al nivel de la población, entre otros. El segundo se refiere a la *Pertinencia*, entendida como la relación del ítems con lo que la prueba pretende medir, es decir, se busca que el ítems esté en armonía con el plan general de la prueba y que pertenezca al dominio definido bien sea teóricamente o como manifestación del atributo que medirá la prueba. El tercero se define como *Relevancia* que se refiere al contenido específico de la pregunta en relación con la estructura de prueba, donde se indican los aspectos generales que se consideran importantes, pero es el constructor quien debe seleccionar aquellos aspectos específicos que considere importantes de acuerdo con el atributo que se pretende medir. Por último se encuentra la *Dificultad*, entendida como la apreciación subjetiva y a priori del nivel de habilidad o de magnitud de atributo necesarios para responder el ítem, esta apreciación de la dificultad está en estrecha relación con las características de la población que será evaluada y puede ser utilizada como criterio para ensamblar la prueba final.

### **Análisis estadístico con el Modelo de Rasch**

El Modelo de Rasch fue propuesto por George Rasch en 1960 basándose en la Teoría de Respuesta al Ítem (TRI), que es un paradigma de la psicometría que surgió hacia los años 50's con los aportes de Lord y Birnbaum (Muñiz & Hambleton, 1992), como una forma alternativa de conceptualizar la medición enfocando su atención en las propiedades de los ítems a diferencia del paradigma clásico basado en el estudio del test global. El objetivo fundamental de esta teoría es la construcción de instrumentos de medición que cuenten con propiedades invariantes entre poblaciones, es decir, si dos evaluados cuentan con el mismo nivel de habilidad tendrán igual probabilidad de dar la misma respuesta a un ítem, independientemente de la población a la que pertenezcan (Attoresi, Lozzia, Abal, Galibert, & Aguerri, 2009)

El paradigma clásico llamado Teoría Clásica de los Test (TCT) fundamenta la construcción, aplicación, validación e interpretación de distintos tipos de tests y que permite derivar escalas estandarizadas aplicables a una población (Hambleton & Zaal, 1994) presentaba algunas limitaciones, que pueden profundizarse en Embretson (1996), Hambleton & van der Linden (1982), Muñiz (2010), Cortada de Kohan (2004), que han llevado al desarrollo de modelos alternativos, entre los cuales el más parsimonioso es el Modelo de Rasch, ya que permite la medición conjunta de las personas y de los ítems que componen un test (Prieto & Delgado, 2003).

Este Modelo ha sido el recurso técnico más utilizado, desde mediados del siglo anterior, para el análisis de ítems que forman parte de evaluaciones de uso académico, aunque también se han utilizado en disciplinas como la psicología, la sociología, ciencias de la salud, entre otras; en este modelo el principal interés se centra en la medición del nivel de habilidad de los evaluados y la dificultad de los ítems que componen la prueba (Gonzalez, 2008).

### ***Definición del Modelo de Rasch***

La TRI brinda una fundamentación probabilística a las diversas discusiones en la medición de rasgos no observables, considerando al ítem como la unidad básica de medición y obteniendo la puntuación por medio del patrón de respuestas del individuo, y no por la sumatoria de las respuestas a ítems individuales como sucede en la TCT (Nunnally, 1987).

El Modelo propuesto por Rasch se fundamenta en tres supuestos básicamente, que de cumplirse garantizan la precisión en la medida (Prieto & Delgado, 2003; Reid, Kolakowsky-Hayner, Lewis, & Armstrong, 2007; Gonzalez, 2008; Attoresi et al, 2009; Solorzano & Montero, 2011).

El primero de los supuestos radica en que el atributo que se desea medir debe representarse en una única dimensión, en la que se sitúan simultáneamente los ítems y los evaluados, este supuesto ha sido definido como *Unidimensionalidad*; aunque en la práctica es difícil que este supuesto se logre satisfacer totalmente debido a que más de un factor puede afectar la respuesta a un ítem, se exige que la presencia de un rasgo fundamental o factor dominante que explique las respuestas de los evaluados.

El segundo supuesto se denomina *Independencia Local* y se refiere a que las respuestas de un individuo a cualquier par de ítems son independientes y no existe relación alguna entre los diferentes ítems presentados, esto significa que las respuestas dadas a los distintos ítems son estadísticamente independientes.

Por último, se considera que el nivel de habilidad de un individuo en cierto atributo y la dificultad del ítem determinan la probabilidad de que la respuesta a un ítem sea correcta. Si esta relación se puede describir mediante una función monótonicamente creciente, es decir, cuando la capacidad del individuo aumenta también aumenta la probabilidad de responder correctamente el ítem; la violación de este supuesto puede derivar en un desajuste importante entre las características del ítem y los datos observados.

Este supuesto hace referencia al parámetro de dificultad y ha sido definido por Rasch (Cortada de Kohan, 2004) mediante una función logística, así:

$$P_{i\theta} = \frac{e^{(\theta-b)}}{1 + e^{(\theta-b)}}$$

en donde:

$P_{i\theta}$  = es la probabilidad de que un examinado elegido al azar con aptitud  $\theta$  conteste correctamente el ítem  $i$ .

$b$  = parámetro de la dificultad del ítem  $i$



En este caso, el eje X representa el nivel de habilidad requerido para responder correctamente el ítem y el eje Y se ubica la probabilidad de responder el ítem correctamente; así, el punto de inflexión 0 en la escala de habilidad corresponde a una probabilidad de 0,5 de responder correctamente el ítem, como puede observarse en la curva número 4; de este modo, puede observarse que a mayor dificultad del ítem se requiere un nivel de habilidad mayor para responder correctamente el ítem con una probabilidad de 0,5 como es el caso de la curva número 2.

La escala utilizada para medir el nivel de habilidad tiene un rango teórico que va desde  $-\infty$  hasta  $+\infty$ , aunque generalmente sus valores se encuentran entre -4 y 4. Esta escala llamada Logit se aplica para ubicar tanto la dificultad de los ítems como la habilidad de los evaluados, suele ubicarse el punto cero (0) como el nivel medio de dificultad del ítem, por el contrario, en el nivel de habilidad de los individuos el punto medio es dejado de manera libre, ubicándose en cualquier punto de la escala.

### ***Estimación de parámetros***

La estimación de los parámetros de la habilidad de los evaluados y la dificultad de los ítems tiene como característica fundamental la invarianza en la medida, es decir, los parámetros que caracterizan a un ítem no dependen de las distribuciones de habilidad de las personas, y a la inversa, las estimaciones de la habilidad de las personas no depende del conjunto de ítems presentados en la prueba (Gonzalez, 2008).

Según Prieto (2003) la estimación de estos parámetros se realiza luego de la aplicación de la prueba y de contar con las respuestas de los evaluados y puede realizarse mediante la estimación condicional o la estimación conjunta. En la estimación condicional se pretende conocer el parámetro de habilidad de las personas a partir del parámetro de dificultad de los ítems

obtenida a partir de aplicaciones anteriores de la prueba, cuando no se cuenta con esta estimación se realizará una estimación conjunta. El método más utilizado para realizar la estimación es el de Máxima Verosimilitud, que consiste en determinar los parámetros que con una mayor probabilidad habrían producido los patrones de respuestas observadas.

Una vez se ha realizado el cálculo de las estimaciones de los parámetros de habilidad y de dificultad es necesario evaluar el ajuste de los datos al modelo que permite ver si existe una diferencia entre lo que el modelo predice y los datos observados; esta evaluación del ajuste se realiza mediante el análisis de la discrepancia entre la CCI teórica o dada por el modelo y la CCI empírica u obtenida a partir de los datos, cuando los datos se ajustan al modelo teórico se puede suponer que se representa de forma adecuada la relación entre el rasgo latente y la probabilidad de dar una respuesta al ítem, si este ajuste no se da puede suponerse que el modelo no es el adecuado para relacionar el rasgo latente con la respuesta al ítem o el ítem no tiene un comportamiento que permita predecir esta relación (Attoresi et al, 2009). Prieto (2003) menciona que este desajuste puede estar relacionado con diversos factores como multidimensionalidad, presencia de sesgo en los ítems, falta de claridad en el enunciado o en las opciones de respuesta, respuestas al azar, falta de motivación del evaluado, errores al anotar las respuestas, entre otras.

Uno de los procedimientos utilizados para detectar el desajuste es el que está basado en residuos o diferencias entre las respuestas observadas y esperadas, este estadístico recibe el nombre de *Infit* que es la medida de los residuos cuadráticos ponderados con su varianza y es sensible a los patrones de respuesta irregulares como, por ejemplo, personas que posean un nivel bajo en la escala de habilidad contesten correctamente uno o varios ítems que requieran de un nivel de habilidad alta. Para la interpretación de las puntuaciones obtenidas en este estadístico pueden tomarse como referencia los valores de la siguiente Tabla (Prieto & Dias, 2003)

Tabla 4. Rango de valores en el Infit que indican ajuste al modelo

| <b>Tamaño de la Muestra</b> | <b>Valores</b>  |
|-----------------------------|-----------------|
| Menos de 500 personas       | Entre 0,7 y 1,3 |
| Entre 500 y 1000 personas   | Entre 0,7 y 1,2 |
| Más de 1000 personas        | Entre 0,7 y 1,1 |

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

Los valores que se encuentran justo en el rango indican que existe un ajuste entre los datos y el modelo mientras que los valores inferiores a estos rangos indican una dependencia o determinismo lo que significa un desajuste, ya que aparentemente ajustan demasiado bien al modelo; de la misma manera los valores superiores también indican desajuste porque se relacionan con una alta variabilidad en los datos. Otro estadístico utilizado es el *Oufit*, que se define como el promedio de los residuales estandarizados tanto de los ítems como de los individuos y es una medida más sensible a respuestas no esperadas en los patrones de respuesta y a valores extremos, lo que permite determinar la discriminación de cada ítem; para esta medición los valores que se encuentran entre  $-1,5$  y  $1,5$  están dentro del intervalo aceptable que permite determinar un ajuste razonable tanto de ítems como de personas, los valores que se encuentran por fuera de este rango son indicadores de desajuste.

### ***Utilidades del modelo***

En cuanto al Modelo de Rasch varios autores como Prieto & Delgado (2003), Cortada de Kohan (2004), Reid, Kolakowsky-Hayner, Lewis, & Armstrong (2007), Attoresi et al (2009) y Solorzano & Montero (2011) han descrito la utilidad de realizar análisis estadísticos bajo este modelo, tanto desde la teoría como de la práctica.

Desde el punto de vista teórico este modelo brinda cuatro propiedades fundamentales. La primera de ellas tiene que ver con la *Medición Conjunta* de los parámetros de los ítems y de los evaluados brindando la posibilidad de ubicarlos en un mismo continuo de medición y en la

misma escala, permitiendo considerar la existencia de una varianza natural de los ítems desvirtuando la creencia de que los ítems miden el constructo de la misma manera; esta propiedad también permite realizar un análisis de la interacción entre los ítems y las personas brindando la posibilidad de hacer una interpretación en términos de probabilidad de acertar a un ítem dado un nivel de habilidad y no basándose en normas de grupo como se hace en la TCT. La segunda propiedad se relaciona con la *Objetividad Específica* donde los parámetros de las personas no estarán condicionadas por las características de los ítems y que los parámetros de los ítems no dependerán de las personas que respondan la prueba. La tercera se refiere a las *Propiedades de Intervalo* que se relacionan con la posibilidad de interpretar que las distancias iguales en la escala representan distancias iguales en el constructo medido, esto quiere decir que a diferencias iguales entre un sujeto y un ítem le corresponde las mismas probabilidades de una respuesta correcta. Por último se encuentra la *Especificidad en el Error Típico de Medida*, propiedad que permite cuantificar la cantidad de información con la que se mide cada punto del atributo, es decir, los ítems más fáciles estimarán con mayor precisión los parámetros de los sujetos de bajo nivel y de igual manera, los sujetos que presentan un alto nivel de habilidad serán medidos con mayor precisión por los ítems difíciles.

Las ventajas de utilizar este Modelo desde el punto de vista práctico se resumen en la conformación de bancos de ítems, la construcción de tests adaptativos y la posibilidad de realizar un análisis del funcionamiento diferencial de los ítems.

Con respecto a la construcción de **Bancos de Ítems** Muñiz (1997) lo define como un conjunto de ítems que miden un mismo rasgo y cuyos parámetros están estimados en la misma escala, y menciona que para la construcción del banco se deben tener aspectos como estructurar la construcción de los ítems en función del contenido que se pretende evaluar, contar con una

muestra representativa de la población para poder aplicar los ítems construidos y utilizar un modelo psicométrico para la calibración de los ítems.

Para el desarrollo del banco de ítems se deben tener en cuenta cuatro fases (Muñiz, 1997). La primera es la *Elección del diseño*, aquí se determina la forma de aplicación de los ítems en función de su cantidad y de la muestra a la cual se aplicarán los ítems; esto tiene como finalidad poner las puntuaciones en una métrica común para poder realizar comparaciones entre las estimaciones obtenidas. De esta manera se proponen cuatro clases de diseños, el diseño de un solo grupo puede utilizarse cuando son pocos ítems, el diseños de grupos equivalentes se utiliza, por lo general, cuando hay una cantidad considerable de ítems, el diseño de anclaje, bien sea de ítems o sujetos y el diseño de doble anclaje tanto de ítems como sujetos. La segunda es la *Evaluación del Ajuste*, donde se determina si los datos satisfacen los supuestos del modelo tales como la unidimensionalidad, la inexistencia de aciertos al azar, si se obtiene la invarianza de los parámetros y de los sujetos y el ajuste entre las puntuaciones observadas y las predicciones realizadas a partir del modelo. En tercer lugar se encuentra la *Calibración*, que permite evaluar el ajuste de los ítems al modelo y eliminar aquellos ítems que no cumplan con los supuestos, también se busca estimar los parámetros de los ítems y transformar las puntuaciones obtenidas a una misma escala, de manera que las puntuaciones obtenidas por los sujetos puedan ser comparables. Finalmente se encuentra la *Creación del Banco*, donde se almacena la información obtenida en la fase anterior para, posteriormente, construir instrumentos con los ítems más ajustados y que brinden una mayor información del atributo medido.

Otra de las ventajas, y tal vez la más destacable, tiene que ver con la posibilidad de contar con **Test Adaptativos Informatizados** (TAI), según Olea, Abad, & Barrada (2010) la principal característica de este tipo de test es que los ítems que se presentan al evaluado se van adaptando

al nivel de habilidad según las respuesta a ítems presentados previamente; para que esto sea posible se debe contar con un banco de ítems amplio que permita seleccionar únicamente los ítems que resultan altamente informativos para poder determinar con precisión el nivel de habilidad del evaluado en el rasgo medido. Algunas ventajas adicionales en la aplicación de este tipo de test tienen que ver con la *mejora de la seguridad del test*, esto es posible porque gran parte de los ítems presentados a los evaluados son diferentes debido a que la presentación de los ítems se va adaptando al nivel de habilidad del evaluado; también es posible *reducir el tiempo de aplicación*, ya que se consiguen niveles similares de precisión que los test convencionales con un número reducido de ítems; además se pueden realizar *estimaciones más precisas* con el mismo número de ítems de un test convencional de lapiz y papel, esto quiere decir que bajo las mismas condiciones tanto en tiempo de aplicación como en número de ítems, un TAI brinda menor error en la medida en el rasgo evaluado y por ende, respecto a las decisiones que se toman a partir de las puntuaciones en el test. Estos aspectos resultan especialmente relevantes cuando se trata de aplicaciones masivas de test de rendimiento o conocimientos y aún más cuando es necesario aplicar este tipo de test de manera continua.

Finalmente, este modelo permite realizar un análisis del **Funcionamiento Diferencial de los Ítems** (DIF) que se ha conceptualizado como la evidencia estadística que muestra si un ítem funciona de forma diferente en dos grupos distintos; esto quiere decir que dos evaluados con el mismo nivel de habilidad y que pertenecen a grupos diferentes, presentan una probabilidad distinta de acertar a un determinado ítem (Hambleton, Swaminathan, & Rogers, 1991; Clauser & Mazor, 1998; Muñiz, 1997; Herrera, 2005; Berrio, 2008). Es importante diferenciar la presencia de DIF con impacto, ya que este último se refiere a que las diferencias encontradas entre los dos grupos esta dada por una magnitud de atributo diferencial.

El procedimiento utilizado para la detección de DIF en el Modelo de Rasch es la *Diferencia de la dificultad*, donde se compara las dificultades estimadas de un ítem entre los grupos focal y de referencia, para ello se realiza la estimación del parámetro de dificultad para cada ítem examinado en cada uno de los grupos, de esta manera y por medio de la prueba T – Welch, se calcula si la diferencia es debida al azar o si depende de otros factores, un valor de T de 2,4 puede considerarse un buen indicador de que las diferencias encontradas no se deben al azar, especialmente si se trata de una prueba con más de 20 ítems (Berrio, 2008). En este punto es relevante diferenciar el concepto de DIF del concepto de sesgo ya que, como se mencionó anteriormente, el DIF es de naturaleza puramente estadística y no considera las causas de este funcionamiento diferencial; solo se considera que un ítem está sesgado cuando se haya dado una explicación debidamente fundamentada para el funcionamiento diferencial (Attoresi et al, 2009)

Este análisis de DIF es útil para la creación de instrumentos de medición invariantes entre poblaciones, por ejemplo, permite adaptar pruebas de una cultura a otra, también permite detectar diferencias entre grupos cuyas interpretaciones pueden generar hipótesis de interés psicológico (Attoresi et al, 2009)

## METODO

Para dar cumplimiento a los objetivos planteados, la metodología de esta investigación se desarrolló en dos fases. La primera tuvo como objetivo realizar los análisis estadísticos de subprueba de Lenguaje de la prueba SABER 11 con el fin de identificar las posibles diferencias en la evaluación de personas con y sin limitación visual, para esto se realizó el reconocimiento y la depuración de la base de datos facilitada por el ICFES, se realizó el análisis estadístico bajo el Modelo de Rasch, el análisis de DIF y un análisis cualitativo de éstos ítems. En la segunda fase se propone el diseño y construcción de ítems para evaluar la comprensión de lectura, tomando como base los procesos involucrados y teniendo en cuenta las posibles diferencias entre personas con y sin limitación visual; para el cumplimiento de este objetivo se realizó el diseño y planeación para la construcción de un banco de ítems y el desarrollo del banco de ítems.

### Fase 1

#### Instrumento

El estudio se realizó utilizando la bases de datos suministradas por el ICFES correspondientes a la subprueba de *Lenguaje* de la aplicación de la prueba SABER 11 en el cohorte 2008 II. Cada aplicación contenía 36 ítems (de selección múltiple con única respuesta) los cuales conformaban dos formas de prueba cada una de 24 ítems, con 12 ítems comunes. La Tabla 4 muestra la estructura de prueba diseñada por el ICFES (2009) para evaluar la subprueba de Lenguaje que evalúa tres componentes y tres competencias.

Tabla 5 Estructura de prueba subprueba de Lenguaje Saber 11 2008-II

| Componente                                 | Competencia            |                 |                |
|--|------------------------|-----------------|----------------|
|  | Propositiva            | Argumentativa   | Interpretativa |
| Función semántica de los elementos locales | 17,23,29,35            | 7,8,14,18,26,30 | 16,28          |
| Configuración del sentido global del texto | 1,3,5,9,10,15,21,27,33 | 6,19,31         | 11,22,34       |
| Del sentido del texto hacia otros textos   | 13,25                  | 2,20,32         | 4,12,24,36     |

Los componentes se han definido como la **Función semántica de los elementos locales** donde se indaga por la función que cumplen los elementos microtextuales (pronombres, adjetivos, adverbios, sustantivos) y locales en la construcción del sentido del texto. La **Configuración del sentido global del texto** hace referencia a cómo el evaluado hace un reconocimiento de la temática global de un texto (qué dice), identificando y estableciendo relaciones entre las informaciones locales de éste, así como entre frases, párrafos y títulos; entre el cuerpo y el título del texto o entre el autor y el texto mismo. Finalmente se encuentra el componente **Del sentido del texto hacia otros textos** que indaga por lo dicho en el texto en relación con otros textos, lo anterior quiere decir que el texto debe ser estudiado progresivamente, de modo que se dé cuenta de la información local, que se configure un sentido global del texto y que se establezca su sentido en relación con otros textos (ICFES 2009).

El ICFES (2009) a su vez ha definido las competencias de Interpretar, Argumentar y Proponer. En primer lugar, **Interpretar** hace referencia a las acciones que realiza una persona, con el propósito de comprender una situación en un contexto específico; la interpretación implica dar cuenta del sentido de un texto, de una proposición, de un problema, de un evento o de una

gráfica, mapa, esquema o simbología propia de la disciplina objeto de evaluación, e involucra acciones tales como: identificar las variables involucradas en una situación problema y sus relaciones, describir el estado, las interacciones o dinámica de un sistema, en términos gráficos o simbólicos, traducir información de un sistema de representación a otro (identificar el esquema ilustrativo correspondiente a una situación dada o, identificar la descripción más apropiada para determinada gráfica) y plantear conclusiones válidas a partir de un conjunto de datos o eventos.

En segundo lugar la competencia de **Argumentar** se refiere a las acciones que realiza una persona con el propósito de fundamentar o sustentar un planteamiento, una hipótesis, un diseño, el uso de un procedimiento, la solución a un problema o a una decisión. Al evaluado no se le solicita que exponga un discurso argumentativo personal, se le pide que se apropie de los sentidos que articulan un texto o discurso breve o un enunciado y encuentre los fundamentos que los explican. Esta competencia exige explicitar los por qué de lo que se hace, e involucra acciones tales como: fundamentar la ocurrencia de determinados fenómenos con base en planteamientos teóricos, sustentar conclusiones para diferentes eventos o fenómenos, mediante el establecimiento de cadenas de relaciones entre conceptos, postulados o principios teóricos, sustentar conclusiones a partir de datos o procedimientos experimentales, argumentar cuál es la información relevante para la resolución de un problema.

Finalmente se encuentra la competencia relacionada con **Proponer**, que hace referencia a las acciones que realiza una persona, con el propósito de plantear alternativas de decisión o de acción; establecer lo que puede ocurrir en un sistema o situación cuando cambian las condiciones y dilucidar relaciones o vínculos no evidentes entre eventos o conceptos. Esta competencia implica una actuación argumentada, crítica y creativa para seleccionar entre opciones o alternativas encaminadas a la solución de un problema, en un contexto determinado. También

involucra acciones tales como: predecir el comportamiento de un sistema o situación si se realizan cambios en las interacciones o en sus condiciones iniciales, conjeturar, a partir de un estado final, las condiciones o estados precedentes que lo hicieron posible, plantear el diseño experimental o la estrategia más adecuada y pertinente para realizar un determinado estudio o contrastar una hipótesis, establecer regularidades válidas para un conjunto de situaciones aparentemente desconectadas y plantear hipótesis.

### **Procedimiento**

En primer lugar se llevó a cabo el reconocimiento y depuración de las bases de datos donde se excluyeron los sujetos en cuya matriz de respuesta presentaban omisiones, multimarca o cuyos datos demográficos no contenían la información necesaria y podían interferir en el análisis. El número de individuos que presentaron la subprueba fueron de 256.348, una vez depurada la base de datos, se conservaron registros de 237.102 individuos que contenían la información necesaria para continuar con el análisis. 237.015 personas sin limitación visual y 87 personas con limitación visual

Como se puede apreciar la población con limitación visual representa menos del 1% de la población total en todas las aplicaciones, situación que puede afectar la precisión de las estimaciones de los parámetros tanto de los individuos como de los ítems, y la potencia estadística de la prueba en la detección del posible funcionamiento diferencial de los ítems (DIF, por sus iniciales en inglés). Basados en los resultados de estudios previos (Herrera, 2005; Arias, 2008; Berrio, 2008; Santana, 2009) se decidió utilizar una razón de tamaños de muestra de 1:25, es decir, 1 persona con limitación visual por cada 25 personas sin limitación visual. Utilizando el muestreo aleatorio se seleccionaron 10 muestras de 2175 evaluados de la población sin limitación visual, lo cual permitió repetir los análisis para la identificación de ítems con DIF.

En segundo lugar, se realizaron los análisis con el programa estadístico Winsteps 3.71.0 (Linacre, 1991, 2006) siguiendo el Modelo de Rasch el cual permite obtener una estimación conjunta, tanto de la habilidad de las personas como de la dificultad de los ítems, en una misma dimensión o constructo; se estimaron los parámetros de dificultad de los ítems, la medida de la habilidad de las personas evaluadas y la confiabilidad de las puntuaciones para cada una de las subpoblaciones, así como el ajuste de los datos al modelo. Además, se identificaron los ítems que presentaban posible Funcionamiento Diferencial considerando las dos subpoblaciones y tomando como grupo de referencia la muestra de examinados videntes.

Finalmente, una vez detectados los ítems con DIF para cada aplicación, se solicitó al ICFES la revisión de contenido de estos ítems con el fin de realizar un análisis cualitativo de los mismos e indagar sobre las posibles causas del funcionamiento diferencial. Esta revisión se llevó a cabo en una sesión de 5 horas en las instalaciones del ICFES, se contó con la información básica del ítem como la competencia y componente evaluado, el enunciado, las opciones de respuesta y la clave; se realizó una lectura rigurosa de cada uno de los ítems y se tomó como referencia de la evaluación las consideraciones para la construcción de ítems, mencionadas en el apartado anterior y se identificaron las posibles causas del funcionamiento diferencial de los ítems.

## **Fase 2**

Esta fase se basó en los resultados obtenidos en los análisis estadísticos de la Fase 1 y en la revisión bibliográfica realizada con el fin de proponer un diseño y construcción de ítems para evaluar la comprensión de lectura, tomando como base los procesos involucrados y teniendo en cuenta las posibles diferencias entre personas con y sin limitación visual y brindar algunas directrices o recomendaciones para la construcción de ítems que pretendan evaluar a poblaciones

con y sin limitación visual.

### **Diseño y planeación del Banco de ítems**

Inicialmente se realizó la identificación del atributo a evaluar y se tomó como referencia el Modelo Interactivo de la comprensión de lectura donde se afirma que el significado del texto está dado por el lector, quien lo construye de manera activa recurriendo a sus conocimientos y a su motivación al momento de interactuar con el texto.

Los componentes a evaluar se tomaron según la propuesta y definición del ICFES (2009) y son la *Función semántica de los elementos locales, el Sentido global del texto y Del sentido del texto hacia otros textos*.

Para definir las competencias se tomaron, de igual manera, las que han sido propuestas por el ICFES, pero se realizó una identificación de cada competencia con los procesos psicológicos básicos involucrados en la comprensión de lectura propuestos por Valles (2005)

En primer lugar, la competencia *Interpretativa – Recordar*, está relacionada de forma directa con aspectos como el vocabulario, idea principal, detalles, secuencias, inferencias y estructura; que le permiten al lector interpretar el texto a partir de un nivel de comprensión literal. Se busca determinar la identificación del significado de palabras y oraciones, evocar pasajes del texto, encontrar el sentido a palabras de múltiple significado, identificar sinónimos, antónimos y homófonos, reconocer y dar significado a los prefijos y sufijos de uso habitual, etc.

La competencia *Argumentativa – Análisis Síntesis* hace referencia a las acciones relacionadas con sustentar información, dar explicaciones, plantear juicios de valor y relaciones con conocimientos previamente adquiridos. Estas habilidades están relacionadas con un nivel de comprensión que incluye la formación de puntos de vista propios del lector a partir del texto y

sus conocimientos previos, plantear posiciones acerca del contenido de un texto, distinguir un hecho de una opinión, captar sentidos implícitos, interpelar la actuación de los personajes, analizar la intención del autor, reflexionar frente a un comportamiento, valorar la estructura de un texto, etc.

Finalmente, la competencia *Propositiva – Aplicación* está relacionada con la generación de hipótesis y la producción textual, busca evaluar si el lector tiene la capacidad de establecer relaciones entre partes del texto para inferir información, conclusiones o aspectos que no están escritos. Este nivel es de especial importancia, pues quien lee va más allá del texto, el lector completa el texto por medio de acciones como deducción de mensajes del texto, proponer títulos para un texto, reformar un texto variando aspectos sustanciales que lo componen, deducir el tema de un texto, elaborar finales diferentes, inferir secuencias lógicas del texto, interpretar lenguaje figurativo, configurar las intenciones del autor que están detrás del texto, que no son explícitas, entre otras situaciones.

Una vez realizada la definición de los componentes, competencias y procesos cognitivos presentes en la comprensión de lectura, se diseñó la estructura de prueba determinando el número de ítems para cada componente y competencia, como se observa en la Tabla 6.

Tabla 6 Estructura de prueba para evaluar comprensión de lectura

| Componentes  | Competencias                      |   |                                  | Total ítems por componente y competencia |
|--|-----------------------------------|---|----------------------------------|--|
|  | Interpretativa<br><i>Recordar</i> | Argumentativa<br><i>Análisis-Síntesis</i> | Propositiva<br><i>Aplicación</i> |  |
| Función semántica de los elementos locales                 | 3                                 | 4   | 3                                | 10                                       |
| Configuración del sentido global del texto                 | 4                                 | 3   | 3                                | 10                                       |
| Del sentido del texto hacia otros textos: Intertextualidad | 3                                 | 3   | 4                                | 10                                       |

### Desarrollo del Banco de ítems

Para la construcción de los ítems se contó con la participación de dos (2) personas expertas en la construcción de preguntas que evalúan comprensión de lectura y dos personas (2) expertas en la teoría psicométrica, en técnicas de medición, evaluación y construcción de instrumentos. Una vez se conformó el grupo de expertos, se realizó una sesión de capacitación a los expertos en consideraciones psicométricas para la elaboración de preguntas y se dio a conocer la definición del atributo, los componentes, las competencias, la estructura de prueba y las consideraciones para la construcción de ítems, además, se construyó un Formato de Preguntas (ver Anexo 1) donde se incluyeron algunos aspectos adicionales a los ya mencionados y basados en la evaluación PISA (Instituto de evaluación, 2010): Situación, tipos de texto, formato de texto y la justificación de la pregunta, la clave y los distractores

La característica de la **Situación** se refiere a los contextos y usos para los que el autor elaboró el texto; entre los tipos de situaciones se encuentran las categorías personales, públicas y educativas. La categoría *personal* se refiere a la utilización de textos dirigidos a satisfacer la curiosidad y puede estar relacionada con las actividades de ocio e intereses personales de un individuo, entre estas se encuentran las cartas personales, las biografías y los textos informativos. En la categoría *público* se utilizan textos que presuponen un contacto más o menos anónimo con otras personas y están relacionados con actividades e inquietudes de la sociedad en general que incluyen acontecimientos públicos y documentos oficiales. Por último, la categoría de texto *educativo* se elabora con el fin de dar instrucciones; este tipo de lectura incluye la adquisición de información como parte de una tarea de aprendizaje más amplia.

También se hace referencia a los **Tipos de texto** que pueden ser utilizados en el contexto, como son: descriptivo, narrativo, expositivo y argumentativo. El texto *descriptivo* hace

referencia a las propiedades de los objetos en el espacio y suelen responder a la pregunta qué y cómo es; para estas descripciones puede utilizarse cualquier formato de texto como párrafos o apartes de un texto así como diagramas e ilustraciones que describen lugares. En el texto *narrativo* la información hace referencia a las propiedades de los objetos en el tiempo y suele responder a la pregunta cuándo o en qué orden; en este tipo de texto se encuentran las novelas, las noticias, una biografía, una tira cómica, una crónica periodística, entre otras. El texto *expositivo* es el tipo de texto en el que la información se presenta en forma de conceptos y/o ideas, o de elementos en los que se pueden inferir éstos y responde a la intención dar a conocer: un fenómeno, una realidad, un descubrimiento, etcétera. En este tipo de texto se encuentran las definiciones que explican en qué modo los términos o nombres se interrelacionan con las ideas o conceptos, las explicaciones que muestran cómo un concepto puede estar vinculado a palabras o términos, los resúmenes y/o mapas conceptuales. El texto *argumentativo* presenta una relación entre conceptos o proposiciones y suelen responder a la pregunta por qué y bajo qué razones sustenta lo dicho; aquí se relacionan los conceptos sobre acontecimientos, objetos e ideas con un sistema personal de pensamientos, valores y creencias; por ejemplo, una carta al editor, una crítica de un libro o película. La toma de posición sobre un evento o tema en particular, es quizás la característica más reveladora en este tipo de textos.

La clasificación del **Formato de texto** hace una distinción entre texto continuo y discontinuo. El *texto continuo* hace referencia a la utilización de oraciones que se organizan en párrafos que forman parte de un texto, en este tipo de textos se pueden encontrar los ensayos, las novelas, las reseñas. El texto *discontinuo* se organiza de forma distinta a los textos continuos; al igual que la oración es la parte básica constitutiva del *texto continuo*, en el texto discontinuo están formados por una serie de listas que pueden ser únicas o simples pero la mayoría se

componen por una combinación de varias de estas listas, un ejemplo de estos textos puede encontrarse en las Tablas, gráficos, anuncios, formularios, mapas, etc.

También se solicitó a los constructores que registraran la **justificación** de por qué la pregunta pertenecía a la competencia y componente a evaluar, así como las acciones cognitivas que debe realizar el evaluado para resolver correctamente la pregunta; del mismo modo, debían justificar cada una de las opciones de respuesta.

Una vez realizada la capacitación se procedió a la construcción individual de preguntas por los constructores y se realizaron 7 sesiones grupales para la validación de las preguntas; en estas sesiones de validación se reunieron simultáneamente a los constructores de preguntas para población con y sin limitación visual con el fin de garantizar un proceso equivalente de construcción de preguntas. El procedimiento realizado en cada sesión de validación fue el siguiente: se presentaba un ítem y se pedía a los constructores-validadores que lo leyeran y contestaran cuál consideraban que era la respuesta correcta; luego se discutía sobre la pertinencia, relevancia y dificultad del ítem así como aspectos formales de la redacción y claridad del ítem y se decidía si el ítem cumplía con la consideraciones psicométricas para ser validado, si eran necesarias algunas modificaciones o si debería ser reemplazado por otro ítem.

Finalizadas las sesiones de validación y contando con los 30 ítems validados, se procedió a la redacción de un manual de construcción preguntas donde se incluyeron aspectos relevantes a tener en cuenta en la construcción de ítems para este tipo de poblaciones, con el fin de garantizar una equidad en la evaluación desde el momento de la construcción de los ítems.

## RESULTADOS

De acuerdo con la metodología planteada, la obtención de los resultados se logró en dos fases: En la primera, se presentan los análisis estadísticos mediante el Modelo de Rasch, la estimación de parámetros de dificultad de los ítems y de habilidad de las personas, el ajuste de los datos al modelo, la identificación de los ítems con DIF y el análisis cualitativo de los mismos; en la segunda fase se mencionan los resultados de la propuesta del diseño, planeación y desarrollo del banco de ítems.

### Fase 1

En la Tabla 7 se presentan los estadísticos descriptivos de las estimaciones del parámetro de dificultad de los ítems y del parámetro de habilidad de los individuos cuando se ajustó un Modelo de Rasch para la población general de evaluados y para cada una de las subpoblaciones con limitación visual (Con LV) y sin limitación visual (Sin LV).

Tabla 7 Estadísticos descriptivos de ítems y sujetos para la población general y subpoblaciones

| Análisis                       | Parámetro           | Estadísticos Descriptivos |            |       |       |               |      |
|--------------------------------|---------------------|---------------------------|------------|-------|-------|---------------|------|
|                                |                     | Media                     | Desviación | Min   | Máx   | Confiabilidad |      |
| Población General<br>N=237.102 | Individuos          | -0,68                     | 0,66       | -3,31 | 3,29  | 0,48          |      |
|                                | Ítems               | 0                         | 0,61       | -1,08 | 0,77  | 1             |      |
| Subpoblaciones                 | Sin LV<br>N=237.015 | Individuos                | -0,68      | 0,67  | -3,31 | 3,29          | 0,48 |
|                                |                     | Ítems                     | 0          | 0,61  | -1,08 | 0,77          | 1    |
|                                | Con LV<br>N=87      | Individuos                | -0,85      | 0,71  | -3,28 | 0,55          | 0,5  |
|                                |                     | Ítems                     | 0          | 0,57  | -1,07 | 1,04          | 0,8  |

Puede observarse que la media de las estimaciones de habilidad para la población general se encuentra por debajo de las estimaciones de dificultad de los ítems; además, hay una mayor variabilidad en las estimaciones de habilidad que en las de dificultad de los ítems, mientras que los individuos se distribuyen a lo largo de toda la escala de habilidad, lo que indica que los ítems

se encuentran agrupados en un nivel medio. Otro resultado llamativo es la diferencia de la confiabilidad de las medidas de los ítems y de los individuos; mientras que las estimaciones de los parámetros de los ítems son altamente confiables, la de los sujetos es baja. Obviamente este resultado está relacionado con el número de examinados y la longitud de la prueba, mientras los tamaños de muestra son suficientes para garantizar estimaciones confiables de los parámetros de los ítems, la longitud de prueba no es suficiente para garantizar estimaciones igualmente confiables de los parámetros de habilidad de los individuos.

En cuanto al análisis de las subpoblaciones puede observarse que las personas Con LV presentan una media más baja que las personas Sin LV; adicionalmente, ésta población presenta mayor variabilidad en las estimaciones de habilidad mientras que las personas Con LV se ubican en un rango medio – bajo de habilidad. La diferencia de la confiabilidad es similar a la observada en la población general

En cuanto al procedimiento de selección de la muestra con razón de tamaño 1:25, se realizaron 10 muestreos aleatorios para la población sin limitación visual, obteniendo una muestra de 2175 personas Sin LV y se conservó la muestra de 87 personas Con LV; se estimaron nuevamente los parámetros de habilidad de los individuos y de dificultad para los ítems, promediándose los resultados obtenidos para la población Sin LV, como se observa en la Tabla 8.

Tabla 8 Estadísticos descriptivos de ítems y sujetos para las subpoblaciones Sin LV y Con LV

| Población           | Parámetro  | Estadísticos Descriptivos |            |       |       | Confiabilidad |
|---------------------|------------|---------------------------|------------|-------|-------|---------------|
|                     |            | Media                     | Desviación | Min   | Máx   |               |
| Promedio remuestreo | Individuos | -0,69                     | 0,66       | -3,31 | 1,77  | 0,46          |
|                     | Sin LV     | Ítems                     | 0          | 0,61  | -1,08 | 0,8           |
| Con LV              | Individuos | -0,85                     | 0,71       | -3,28 | 0,55  | 0,52          |
|                     | Ítems      | 0                         | 0,57       | -1,08 | 0,98  | 0,8           |

Al igual que en el análisis anterior, la población Sin LV presenta una media mayor en el nivel de habilidad que la población Con LV; sin embargo, se observa una disminución en la variabilidad de las estimaciones de habilidad de la población Sin LV, en cuanto a la confiabilidad se siguen manteniendo las diferencias entre ítems y sujetos dado el tamaño de la muestra y el número de ítems

En lo referente al ajuste de los datos al modelo se estimaron los valores de Infit y Oufit tanto para los ítems como para los individuos. En la Tabla 9 se observan los valores de ajuste para los ítems en cada una de las poblaciones.

Tabla 9 Ajuste de los ítems al modelo en cada subpoblación

| Item | Sin LV     |       |       | Con LV     |       |       |
|------|------------|-------|-------|------------|-------|-------|
|      | Dificultad | Infit | Oufit | Dificultad | Infit | Oufit |
| 1    | -0,74      | 1,04  | 1,04  | -0,72      | 0,99  | 1,10  |
| 2    | -0,07      | 1,05  | 1,08  | -0,21      | 1,00  | 1,01  |
| 3    | 0,07       | 1,03  | 1,04  | -0,15      | 0,93  | 0,88  |
| 4    | -0,93      | 1,00  | 1,01  | -0,88      | 1,00  | 1,00  |
| 5    | -0,62      | 0,96  | 0,97  | -0,62      | 1,02  | 1,00  |
| 6    | -1,17      | 0,94  | 0,92  | -0,15      | 1,04  | 1,05  |
| 7    | 0,12       | 1,00  | 0,99  | 0,13       | 0,96  | 0,96  |
| 8    | 0,70       | 0,96  | 0,95  | 0,98       | 0,92  | 0,82  |
| 9    | -0,27      | 0,91  | 0,89  | -0,21      | 1,07  | 1,18  |
| 10   | 0,54       | 0,97  | 0,96  | 0,51       | 0,95  | 0,86  |
| 11   | -0,67      | 1,00  | 0,99  | -0,47      | 1,00  | 0,97  |
| 12   | 0,56       | 1,09  | 1,18  | 0,07       | 1,09  | 1,09  |
| 25   | 0,68       | 1,02  | 1,03  | 0,25       | 1,06  | 1,12  |
| 26   | -1,05      | 0,98  | 0,96  | -0,83      | 1,04  | 1,04  |
| 27   | -0,88      | 1,01  | 1,02  | -1,08      | 0,90  | 0,88  |
| 28   | 0,07       | 0,92  | 0,90  | -0,21      | 1,02  | 1,03  |
| 29   | 0,46       | 1,01  | 1,00  | 0,98       | 0,99  | 0,98  |
| 30   | 0,42       | 0,96  | 0,95  | 0,51       | 0,88  | 0,76  |
| 31   | -0,17      | 1,02  | 1,03  | 0,38       | 1,00  | 0,98  |
| 32   | 0,43       | 1,06  | 1,10  | 0,01       | 1,10  | 1,06  |
| 33   | 0,46       | 0,98  | 0,98  | 0,51       | 1,02  | 1,06  |
| 34   | 0,69       | 1,03  | 1,05  | 0,73       | 0,99  | 1,07  |
| 35   | 0,75       | 1,02  | 1,01  | 0,19       | 0,94  | 0,90  |
| 36   | 0,62       | 1,01  | 1,02  | 0,25       | 0,99  | 1,01  |

Para el ajuste de los ítems al modelo en la población Sin LV se tomo como referencia lo mencionado por Prieto & Dias (2003) teniendo en cuenta el tamaño de la muestra y se observa que para la totalidad de los ítems los valores de Infit se encuentran en un rango entre 0,7 y 1,1, lo que indica que hay un ajuste al modelo; de la misma manera, los valores de Infit para los ítems en la población Con LV se encuentran en un rango entre 0,7 y 1,3, indicando así un ajuste al modelo. Con respecto los valores de Oufit también se evidencia un ajuste de los datos al modelo. Se realizó el mismo procedimiento para analizar el ajuste de los individuos y se encontraron valores de Infit y Oufit acordes con lo esperado, lo que indica un ajuste al modelo. Cabe resaltar la importancia del ajuste de los datos al modelo, ya que si este ajuste no se presenta, los valores carecen de significado teórico y las ventajas de la utilización del Modelo de Rasch se disipan (Prieto & Delgado, 2003)

El análisis de DIF se realizó estimando la diferencia de la dificultad de cada ítem para cada uno de los 10 remuestreos aleatorios de la población Sin LV con la estimación de la dificultad de los ítems para las personas Con LV, como se observa en la Tabla 10. Según la *regla de medio logito* (Berrio, 2008) se encontraron diferencias por encima de medio logit en promedio para el ítem 6, de la misma manera ocurre para los ítems 29 y 31

Si bien esta regla permite sospechar la presencia de DIF no se convierte en prueba suficiente para afirmar la existencia de un funcionamiento diferencial, se utilizó la prueba T – Welch para confirmar si las diferencias encontradas se deben al azar o dependen de otros factores.

Tabla 10 Estimación de la diferencia de la dificultad en población con y sin limitación visual

| Ítem | Remuestreo |       |       |       |       |       |       |       |       |       | Promedio |
|------|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
|      | 1          | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |          |
| 1    | 0,1        | -0,03 | -0,02 | 0     | 0,03  | 0,05  | 0,03  | 0,14  | 0,01  | 0,02  | 0,03     |
| 2    | -0,39      | 0,16  | 0,14  | -0,02 | 0,01  | 0,14  | 0     | 0,13  | 0,01  | 0,05  | 0,02     |
| 3    | 0,27       | 0,23  | 0,22  | 0,26  | 0,25  | 0,31  | 0,17  | 0,11  | 0,28  | 0,19  | 0,23     |
| 4    | -0,16      | -0,15 | -0,05 | -0,18 | -0,13 | -0,16 | -0,15 | -0,13 | -0,17 | 1,94  | 0,07     |
| 5    | 0,07       | 0,11  | 0     | 0,17  | 0,13  | 0,12  | -0,01 | -0,01 | 0,18  | 0,1   | 0,09     |
| 6    | -0,9       | -0,94 | 1,32  | -0,79 | -0,78 | -1,04 | -0,95 | -1,03 | -0,74 | -0,88 | -0,67    |
| 7    | 0,08       | -0,04 | -0,01 | 0,13  | 0,12  | -0,1  | -0,01 | 0,02  | 0,12  | 0     | 0,03     |
| 8    | -0,17      | -0,24 | -0,28 | -0,06 | -0,08 | -0,24 | -0,23 | -0,39 | -0,12 | -0,22 | -0,20    |
| 9    | 0,01       | 0,12  | -0,06 | 0,21  | 0,14  | 0,07  | 0,02  | 0,03  | 0,25  | 0,15  | 0,09     |
| 10   | 0,23       | 0,12  | 0,03  | 0,19  | 0,2   | 0,22  | 0,19  | 0,1   | 0,14  | 0,17  | 0,16     |
| 11   | -0,36      | -0,24 | -0,2  | -0,34 | -0,34 | -0,31 | -0,24 | -0,22 | -0,35 | -0,32 | -0,29    |
| 12   | 0,34       | 0,47  | 0,49  | 0,06  | 0,07  | 0,42  | 0,53  | 0,66  | 0,03  | 0,24  | 0,33     |
| 25   | 0,37       | 0,45  | 0,43  | 0,24  | 0,25  | 0,33  | 0,45  | 0,48  | 0,24  | 0,41  | 0,37     |
| 26   | -0,23      | -0,2  | 1,88  | -0,11 | -0,04 | -0,27 | -0,27 | -0,19 | -0,08 | -0,24 | 0,03     |
| 27   | 0,28       | 0,26  | 0,2   | 0,25  | 0,31  | 0,18  | 0,18  | 0,23  | 0,22  | 0,25  | 0,24     |
| 28   | 0,36       | 0,29  | 0,28  | 0,44  | 0,4   | 0,32  | 0,29  | 0,15  | 0,45  | 0,44  | 0,34     |
| 29   | -0,52      | -0,6  | -0,52 | -0,64 | -0,59 | -0,58 | -0,46 | -0,6  | -0,67 | -0,52 | -0,57    |
| 30   | -0,06      | -0,12 | -0,09 | -0,13 | -0,11 | -0,09 | -0,09 | -0,13 | -0,12 | -0,12 | -0,11    |
| 31   | -0,64      | -0,54 | -0,55 | -0,53 | -0,5  | -0,5  | -0,56 | -0,58 | -0,51 | -0,45 | -0,54    |
| 32   | 0,33       | 0,31  | 0,42  | 0,22  | 0,21  | 0,42  | 0,45  | 0,46  | 0,23  | 0,29  | 0,33     |
| 33   | -0,06      | 0,02  | -0,05 | 0,03  | -0,01 | 0,07  | -0,02 | 0     | 0,02  | 0,02  | 0,00     |
| 34   | -0,26      | -0,18 | -0,04 | -0,21 | -0,3  | -0,17 | -0,18 | -0,01 | -0,23 | -0,18 | -0,18    |
| 35   | 0,46       | 0,43  | 0,56  | 0,54  | 0,45  | 0,43  | 0,52  | 0,47  | 0,53  | 0,37  | 0,48     |
| 36   | 0,33       | 0,33  | 0,37  | 0,31  | 0,33  | 0,4   | 0,37  | 0,34  | 0,29  | 0,42  | 0,35     |

Se realizaron las estimaciones con la prueba T – Welch para los 24 ítems que conforman la prueba, se encontraron diferencias significativas en los ítems 6, 29 y 31 como se muestra en la Tabla 12, confirmando los hallazgos con la regla de medio logito. Sin embargo en el primer análisis no se detectó presencia de DIF en el ítem 35, como se observa en la Tabla 11, pero con la prueba T – Welch se detectó presencia de DIF.

Tabla 11 Estimaciones de la dificultad de ítems identificados con DIF

| Ítems<br>con DIF | Dificultad del ítem |        |           |
|------------------|---------------------|--------|-----------|
|                  | Con LV              | Sin LV | T - Welch |
| 6                | -0,15               | -1,19  | 4,22**    |
| 29               | 0,98                | 0,34   | 2,13*     |
| 31               | 0,38                | -0,26  | 2,42*     |
| 35               | 0,25                | 0,75   | -2,23*    |

\*p 0.05; \*\*p 0.01

Estos cuatro ítems detectados con DIF obtuvieron un valor T - Welch suficiente como para justificar una cuidadosa revisión con miras a identificar las posibles razones del DIF. Por lo menos los resultados informan sobre la posible valoración diferencial de ciertos aspectos evaluados a través de los ítems, factores que pueden referir a un concepto o constructo que se efectúa en detrimento de la población invidente.

Se realizó un análisis cualitativo de estos cuatro ítems donde se revisaron aspectos formales de construcción de los ítems como pertinencia, relevancia, dificultad y las consideraciones para la construcción de ítems. Cabe resaltar que por razones de seguridad de la información y protocolos establecidos por el ICFES no fue posible extraer alguno de estos ítems para poder ilustrar lo encontrado en esta revisión, sin embargo, a continuación se muestran algunos ejemplos de textos y/o preguntas similares a las revisadas.

### **Ejemplo 1**

*Sobre el Cuento* (Julio Cortazar)<sup>1</sup>

(...) nadie puede pretender que los cuentos sólo deban escribirse luego de conocer sus leyes. En primer lugar, no hay tales leyes; a lo sumo cabe hablar de puntos de vista, de ciertas constantes que dan una estructura a ese género tan poco encasillable; en segundo lugar, los teóricos y los críticos no tiene por qué ser los cuentistas mismos, y es natural que aquéllos sólo entren en escena cuando exista ya un acervo, un acopio de literatura que permita indagar y esclarecer su desarrollo y sus cualidades.

<sup>1</sup> Fragmentos del Escrito de Julio Cortazar *Sobre el Cuento* Texto tomado de [http://www.ciudadseva.com/textos/teoria/opin/sobre\\_el\\_cuento.htm](http://www.ciudadseva.com/textos/teoria/opin/sobre_el_cuento.htm)

(...) Los cuentistas inexpertos suelen caer en la ilusión de imaginar que les bastará escribir lisa y llanamente un tema que los ha conmovido, para conmover a su turno a los lectores. Incurren en la ingenuidad de aquél que encuentra bellísimo a su hijo, y dar por supuesto que los demás lo ven igual de bello. Con el tiempo, con los fracasos, el cuentista capaz de superar esa primera etapa ingenua, aprende que en la literatura no bastan las buenas intenciones. Descubre que para volver a crear en el lector esa conmoción que lo llevó a él a escribir en cuento, es necesario un oficio de escritor, y que ese oficio consiste, entre otras cosas, en lograr ese clima propio de todo gran cuento, que obliga a seguir leyendo, que atrapa la atención, que aísla al lector de todo lo que lo rodea para después, terminado el cuento, volver a conectarlo con su circunstancia de una manera nueva, enriquecida, más honda o más hermosa. Y la única forma en que puede conseguirse ese secuestro momentáneo del lector es mediante un estilo basado en la intensidad y en la tensión, un estilo en el que los elementos formales y expresivos se ajusten, sin la menor concesión, a la índole del tema, le den su forma visual y auditiva más penetrante y original, lo vuelvan único, inolvidable, lo fijen para siempre en su tiempo y en su ambiente y en su sentido más primordial.

(...) Para entender el carácter peculiar del cuento se le suele comparar con la novela, género mucho más popular y sobre el que abundan las preceptivas. Se señala, por ejemplo, que la novela se desarrolla en el papel, y por lo tanto en el tiempo de lectura, sin otros límites que en el agotamiento de la materia novelada; por su parte, el cuento parte de la noción límite, y en primer término de límite físico, al punto que en Francia, cuando un cuento excede de las veinte páginas, toma ya el nombre de *nouvelle*, género a caballo entre el cuento y la novela propiamente dicha. En este sentido, la novela y el cuento se dejan comparar analógicamente con el cine y la fotografía, en la medida en que en una película es en principio un “orden abierto”, novelesco, mientras que la fotografía lograda presupone una ceñida limitación previa, impuesta en parte por el reducido campo que abarca la cámara y por la forma en que el fotógrafo utiliza estéticamente esa limitación. No sé si ustedes han oído hablar de su arte a un fotógrafo profesional; a mi siempre me ha sorprendido el que se exprese tal como podría hacerlo un cuentista en muchos aspectos. Fotógrafos de la calidad de un Cartier-Bresson o de un Brassai definen su arte como una aparente paradoja: la de recortar un fragmento de la realidad, fijándole determinados límites pero de manera tal que ese recorte actúe como una explosión que abre de par en par una realidad mucho más amplia, como una visión dinámica que trasciende espiritualmente el campo abarcado por la cámara. Mientras que en el cine, como en la novela, la captación de esa realidad más amplia y multiforme se logra mediante el desarrollo de elementos parciales, acumulativos, que no excluyen, por supuesto, una síntesis que dé el “climax” de la obra, en una fotografía o un cuento de gran calidad se procede inversamente, es decir que el fotógrafo o el cuentista se ven precisados a escoger y limitar una imagen o un acaecimiento que sean significativos, que no solamente valga por sí mismos sino que sean capaces de actuar en el espectador o en el lector como una especie de apertura, de fermento que proyecta la inteligencia y la sensibilidad hacia algo que va mucho más allá de la anécdota visual o literaria contenida en la foto o en el cuento...

Este primer ejemplo muestra un texto con algunas características que pueden afectar de manera negativa la comprensión de lectura. En primer lugar este texto está fragmentado, es decir, cada párrafo es un fragmento del escrito, lo que implica que no haya una relación lógica y estructurada entre los párrafos y que sigan una linealidad y asociación entre los mismos; esto afecta negativamente la comprensión de lectura porque se perjudican algunos de los *procesos psicológicos básicos* implicados en la comprensión como lo son el análisis – síntesis y la memoria a corto y largo plazo (Vallés, 2005) que son vitales para el desarrollo de ideas significativas, extracción de conclusiones y relación de lo que se lee con los conocimientos previos.

En segundo lugar, el texto reviste una complejidad que no está acorde con el nivel de escolaridad de los evaluados, sería más adecuado y provechoso para lectores que tengan algún vínculo con la literatura, por ejemplo. Esta complejidad puede afectar la comprensión desde dos perspectivas, una de ellas relacionada con los *procesos cognitivos lingüísticos* (Valles, 2005) involucrados en la comprensión de lectura que se encargan de acceder al significado de las palabras y de extraer la información necesaria del texto para lograr la comprensión, ya el texto puede incluir palabras que no sean familiares para el lector y el acceso al léxico se vea afectado; la otra perspectiva está relacionada con la motivación del lector que mencionan los *Modelos Interactivos* de comprensión de lectura porque los conocimientos previos facilitan la asimilación de la información de orden inferior consistente con las expectativas del lector y los datos del texto aseguran que el lector esté atento a la información nueva o que no encaje con sus hipótesis previas; en este modelo el significado está dado por el lector, quien lo construye de manera activa recurriendo a sus conocimientos y a su motivación al momento de interactuar con el texto (Alonso et al, 1985).

Así como el texto analizado reviste algunas dificultades en la comprensión, las preguntas que pueden realizarse con base en éste pueden llegar a evaluar de manera imprecisa la comprensión de lectura, como puede observarse en los siguientes ejemplos.

### **Ejemplo 2**

1. En el párrafo 3, la palabra *acaecimiento* hace referencia a:
  - A. Un acontecimiento
  - B. Una interrupción
  - C. Una complicación
  - D. Un inconveniente

Esta pregunta evalúa el componente *Función Semántica de los Elementos Locales* y la competencia *Interpretativa*, sin embargo, podría decirse que no evalúa comprensión de lectura porque no se requiere de la comprensión del texto para elegir la respuesta correcta, basta con recordar el significado de algunas palabras para responder correctamente. También puede afirmarse que la utilización de textos extensos o complejos demanda que los evaluados deban volver a leer el texto cada vez que respondan a una pregunta ligada a ese texto, de la misma manera que la pregunta invita indirectamente a releer el párrafo 3.

### **Ejemplo 3**

2. El cuento se compara con la novela porque:
  - A. La novela se desarrolla en el papel
  - B. La novela se compara con el cine
  - C. El cuento se compara con la fotografía
  - D. No hay relación entre el cuento y la novela

Esta pregunta evalúa el componente *Configuración del sentido global del texto* y la competencia *Interpretativa*. Aunque esta pregunta evalúa comprensión de lectura y corresponde a la intención de evaluar el componente y la competencia, presenta por lo menos dos inconvenientes importantes: en primer lugar no se logra evidenciar cuál es la intención del constructor de la pregunta y podría decirse que no discrimina entre las personas que dominan el tema y las personas que no, por ejemplo, las opciones A, B y C pueden ser correctas y la opción D es obviamente descartable; y esto puede darse por el segundo inconveniente, en las opciones de respuesta se incluyen palabras presentes en el enunciado lo que puede inducir en el evaluado que esa es la respuesta correcta, en el caso de la opción A.

## Fase 2

Luego de los análisis realizados en la Fase 1 y del procedimiento realizado en la Fase 2 mencionado en el apartado anterior se construyeron las 30 preguntas que conformarán, inicialmente, el banco de ítems para evaluar comprensión de lectura basado en los procesos psicológicos involucrados en la misma; además, se elaboró un Manual para la Construcción de preguntas que evalúen comprensión de lectura en personas con y sin limitación visual (ver Anexo 2).

A continuación se presentan algunos ejemplos de las preguntas construidas que pueden ilustrar las consideraciones que se deben tener en cuenta para la construcción de este tipo de ítems que evalúan la comprensión de lectura en personas con y sin limitación visual

### Ejemplo 4

Rubén Salazar, director de Ingeniería Aeronáutica de la Universidad San Buenaventura y quien desarrolló junto a un grupo de ingenieros de esta institución un dron utilizado por Ecopetrol para monitorear sus oleoductos y por Codensa para vigilar sus redes eléctricas en Cartagena, resume en dos las ventajas de estos aparatos: se reducen los costos de operación

frente a naves tripuladas y no se pone en riesgo la vida de seres humanos en situaciones peligrosas.

El Navigator, dron diseñado y construido por Salazar y su equipo, alcanzó un costo de 1.200 millones, tiene una autonomía de vuelo de 4 horas, una carga útil de 10 kilogramos y puede recorrer un rango de 61 kilómetros, además de ser capaz de despegar en espacio del tamaño de una cancha de fútbol.

“En el caso de erupción de un volcán como el Galeras, se podría enviar un dron para monitorear el área y analizar el comportamiento del volcán. Lo mismo podría hacerse en caso de una inundación”, dice Salazar. A la lista de aplicaciones añade el estudio del tráfico en una ciudad como Bogotá en la que un dron equipado con las cámaras adecuadas podría enviar información en tiempo real a un computador para ser analizada y compartida con los ciudadanos<sup>2</sup>

A diferencia del texto presentado en el Ejemplo 1, este texto no está fragmentado lo que facilita una lectura fluida porque tiene una disposición lógica y coherente, ejemplifica la *Situación* en la categoría *Público* que tiene como característica el uso de acontecimientos de dominio público y que no reviste mayor complejidad en el lenguaje (Instituto de evaluación, 2010), la extensión del texto está acorde permite que no haya una mayor demanda de recursos atencionales y de memoria que puedan interferir con la comprensión del mismo.

Con respecto a las dificultades en las preguntas, se presentan algunos ejemplos que brindan posibles soluciones en cuanto a evaluar correctamente la comprensión de lectura, minimizar la lectura repetida del texto, el uso de las mismas palabras en el contexto y las preguntas, definir claramente la intención de la pregunta y las acciones cognitivas que se espera del evaluado.

#### Ejemplo 5

1. Cuando se dice de El Navigator: “además de ser capaz de despegar desde un espacio del tamaño de una cancha de fútbol”, se busca
  - A. mostrar que ese dron ahorra espacio.
  - B. presentar una razón para comprarlo.
  - C. resaltar que gasta mucho combustible.
  - D. señalar la practicidad de su diseño.

<sup>2</sup> Tomado y adaptado de: El lado bueno de los drones. El Espectador, 18 de Junio de 2013

Esta pregunta brinda una posible solución a los problemas que presenta la pregunta del Ejemplo 2 dado que evalúa el mismo componente y competencia; en este caso la pregunta evalúa correctamente el componente y competencia porque se requiere de una comprensión del texto para identificar cuál sería la respuesta correcta y no se induce a que el lector deba devolverse al texto, además, se resuelve el inconveniente de inducir a una lectura repetida del texto al incluir en la pregunta el fragmente del texto por el que se quiere indagar. Cabe aclarar que la intención no es evitar que el lector haga una relectura del texto, ya que esto depende del entrenamiento que cada evaluado tenga para leer, lo que se busca es que desde la pregunta no se obligue a una relectura para poder dar la respuesta a la pregunta, sino que sea una decisión del lector hacerlo. En relación con la intención de la pregunta, el autor de la misma manifiesta que esta pregunta pretende que el lector identifique la frase como un descriptor más de las cualidades que tiene El Navigator; da cuenta de la comprensión de los elementos locales del texto, para lo que se debe interpretar adecuadamente la intención de la frase.

#### Ejemplo 6

2. De acuerdo con el texto, se puede inferir que un dron es
  - A. un nuevo control de navegación aérea.
  - B. una nave aérea manejada a distancia.
  - C. un robot capaz de manejar un avión.
  - D. una estrategia de investigación satelital

Esta pregunta evalúa el mismo componente y competencia del Ejemplo 3, esta pregunta muestra de manera clara y concreta la intención del autor de la pregunta porque busca que el lector infiera lo que no está contenido en el texto en relación con los drones, para ello tiene que tomar los elementos claves del texto e interpretarlos de manera global. También se observa que las opciones de respuesta no incluyen frases textuales del texto evitando dar pistas sobre la respuesta correcta y sobre opciones fácilmente descartables.

## DISCUSIÓN Y CONCLUSIONES

La evaluación y medición mediante pruebas objetivas de diversos rasgos o atributos en los ámbitos psicológicos y educativos, ponen de manifiesto la importancia de tener instrumentos que cuenten con las calidades psicométricas que permitan realizar inferencias válidas y confiables de los resultados obtenidos. Aun más cuando se realiza la evaluación a poblaciones con distintas características que pueden afectar el desempeño de las personas en la evaluación.

Los estudios de DIF se han enfocado tradicionalmente en identificar las diferencias en cuanto a género o diferencias culturales con poblaciones minoritarias (Hidalgo, et al 1999; Berrio, 2008) pero son escasos, e incluso inexistentes, los estudios que se enfoquen en identificar diferencias que se refieren a condiciones de discapacidad, y para este caso, limitación visual. Aunque se han desarrollado estrategias que permitan la adaptación tanto de los instrumentos como de las condiciones de evaluación estas no garantizan que se evalúa el mismo constructo en poblaciones con y sin limitación visual.

Lo anterior resulta interesante teniendo en cuenta que la entrada de información por vía auditiva no es una variable nueva para las personas con limitación visual, y que sus procesos de desarrollo y aprendizaje han incluido esta forma de recepción de información, lo cual podría plantear la posibilidad de que sean aspectos característicos de las pruebas utilizadas para evaluarlos lo que esté influyendo en el desempeño, y que el diseño de éstas no estaría contemplando este mecanismo de recepción de la información, el cual requiere por parte de quien responde, un mayor uso de recursos cognitivos como memoria y atención.

Una evidencia teórica la brindan Crowder (1985) y Hoover & Gough (1990) quienes afirman que, según la modalidad sensorial utilizada para la entrada de la información, se presentan diferencias en el procesamiento de un texto, bien sea de manera visual o auditiva. Del

mismo modo, los diferentes análisis estadísticos brindan evidencia empírica de un comportamiento diferencial de los ítems para las dos poblaciones evaluadas, se observa que la distribución de las respuestas de las personas sin limitación a lo largo del continuo de habilidad es más amplio, indicando que estas personas respondieron acertadamente a ítems que para las personas con limitación resultaron más difíciles. Estos resultados muestran la relevancia de generar mayor conocimiento sobre la calidad de la evaluación dirigida a población minoritaria, ya que con las actuales metodologías de examen se obtienen resultados que aparentemente no les favorecen.

Una recomendación técnica que se deriva directamente de los análisis de Rasch y la detección de DIF, es la necesidad de elaborar pruebas que permitan evaluar el constructo medido a lo largo de la dimensión de la habilidad, con el fin de tener información en sus diferentes niveles ya que en las pruebas actuales solamente se están midiendo apropiadamente los niveles medios de habilidad. Si además se tiene en cuenta que la población con limitación visual tiende a agruparse en niveles bajos de habilidad en comparación con los examinados sin limitación, estos resultados muestran que las dos poblaciones no están siendo evaluadas con la misma calidad. También se recomienda utilizar otros procedimientos para detectar ítems con DIF ya que este Modelo solamente contempla la posibilidad de detectar DIF Uniforme (Berrio, 2008) y es posible que se presente DIF no uniforme en los ítems de la prueba; adicionalmente se recomienda el uso de metodologías cualitativas, como los grupos focales, en las cuales se pueda indagar desde la perspectiva del evaluado si las diferencias se deben a las características de los individuos o de los ítems

De la misma manera, para tener mayor precisión en la estimación de la habilidad de las personas, es importante el diseño de ítems que tengan en cuenta las características de los ítems, a

la luz de las consideraciones propuestas para la construcción de ítems, el tipo de texto utilizado y el uso de un lenguaje acorde con el nivel educativo de los evaluados, para determinar si se está exigiendo un tipo de competencia diferente a la que se pretende evaluar, o si éstos no están teniendo en cuenta las características de los procesos psicológicos involucrados en la comprensión de lectura para las dos poblaciones .

Al respecto resulta relevante considerar la importancia de los procesos psicológicos presentes en la comprensión de lectura mencionados por Vallés (2005), toda vez que la interrupción en alguno de los procesos perceptivos, en los procesos psicológicos básicos o en los procesos cognitivos lingüísticos implican una dificultad en la comprensión porque son procesos que se dan de manera simultánea. Es importante tener en cuenta estos procesos no solamente para evaluar la comprensión de lectura sino para evaluar cualquier otro rasgo o atributo, ya que se debe considerar la comprensión de lectura como un proceso transversal a cualquier área de conocimiento y a todo proceso de evaluación mediante pruebas escritas y de aplicación masiva.

Finalmente, es de resaltar las ventajas que representan los modelos basados en la Teoría de Respuesta al Ítem, como el Modelo de Rasch, en el desarrollo de nuevas estrategias que permiten el uso de nuevas tecnologías informáticas para optimizar los procesos de evaluación. Aunque los alcances de esta investigación solamente proponen un diseño para el desarrollo de un banco de ítems y ofrece 30 ítems, puede considerarse una cuota inicial para la creación del banco de ítems y se puede alimentar a través del tiempo mediante la calibración de nuevos ítems y en un futuro, no muy lejano, permitir el desarrollo de TAI's que le permitan a las personas con limitación responder este tipo de pruebas objetivas de manera autónoma, buscando evaluar de manera más equitativa.

## REFERENCIAS

- AERA; APA; NCME . (1999). Standards for educational and psychological testing. Washington: American Educational Research Association.
- Alonso, J., & Mateos, M. (1985). Comprensión lectora: Modelos, entrenamiento y evaluación. Infancia y aprendizaje, 5-19.
- Arias, E. (2008). Efecto de la razón de tamaño y el ajuste del modelo sobre el estadístico Mantel Haenszel y su métrica delta en la detección de DIF. Tesis de grado de maestría, Universidad Nacional de Colombia. Bogotá, Colombia.
- Attoresi, H; Lozzia, G; Abal, F; Galibert, M & Aguerri, M. (2009) Teoría de respuesta al ítem: Conceptos básicos y aplicaciones para la medición de constructos psicológicos Revista Argentina de Clínica Psicológica
- Berrio, A. (2008). La razón de tamaños y desajustes al modelo en la detección de ítems con funcionamiento diferencial mediante el procedimiento diferencia de la dificultad. Tesis de maestría para optar por el título de magister en psicología, Departamento de Psicología, Universidad Nacional de Colombia. Bogotá, Colombia.
- Brown, F. (1980). Fundamentos de Medición y evaluación en psicología y educación. México: El Manual Moderno.
- Castillo, M. (2004). La evaluación: una estrategia a nivel internacional para el mejoramiento de la calidad educativa. Secretaría de educación de Bogotá: Recuperado de <http://www.eduteka.org/EvaluacionBogota.php> el 4 de Septiembre de 2010
- Cheung, D. (2002). Science and technology education conference. How can we construct good multiple choice (pp. 1-11). Hong Kong: Recuperado de <http://www3.fed.cuhk.edu.hk/chemistry/files/constructMC> el 1 de noviembre de 2013.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 31-44.

Crowder, R. (1985). *Psicología de la lectura*. Madrid, España: Alianza Editorial.

De Zubiría, M. (1996). *Teoría de las seis lecturas*. Bogotá, Colombia: Editorial Presencia.

Duro, A. (1992). La coherencia textual en los modelos para la comprensión de textos: Kintsch et al Vs. Sanford y Garrod. *Cognitiva* 4 (2), 227-244.

Gaceta C. (1991). Constitución Política de Colombia. Recuperado de [http://www.secretariassenado.gov.co/http://www.secretariassenado.gov.co/senado/basedoc/cp/constitucion\\_politica\\_1991.html](http://www.secretariassenado.gov.co/http://www.secretariassenado.gov.co/senado/basedoc/cp/constitucion_politica_1991.html)

Golder, D., & Gaonac'h, D. (2002). *Psicología de la lectura. Leer y comprender*. México: Siglo Veintiuno Editores.

Gonzalez, M. (2008). El análisis de reactivos con el Modelo de Rasch Manual técnico A. Obtenido de Obtenido de Universidad de Sonora - Instituto Nacional para la Evaluación de la Educación. México D.F: <http://www.winsteps.com/a/recursos-offline.pdf>

Goodman, K. (1967). Reading a Psycholinguistic Guessing Game. Recuperado de [http://readbysight.com/images/Reading\\_a\\_Psycholinguistic\\_Guessing\\_Game.pdf](http://readbysight.com/images/Reading_a_Psycholinguistic_Guessing_Game.pdf) el 5 de febrero de 2013

Goodman, K. (1982). El proceso de la lectura: Consideraciones a través de las lenguas y del desarrollo. In E. Ferreiro, & M. Gómez Palacio, *Nuevas perspectivas sobre los procesos de lectura y escritura*. México: Siglo XXI.

Guzmán, C. (2010). La evaluación: Una propuesta para formar competencias. Recuperado de <http://www.eduteka.org/EvaluacionBogota.php> el 4 de Septiembre de 2010

Haladyna, T., & Downing, S. (1989). A taxonomy of multiple-choice item-writing rules. *Applied*

measurement in education 2 (1), 50-62.

Hambleton, R., & Zaal, J. (1994). *Advances in educational psychological testing: Theory and applications*. Boston, Londres: Kluwer Academic Publisher.

Hambleton, R., & Zaal, J. (1994). *Advances in educational psychological testing: Theory and applications*. Boston, Londres: Kluwer Academic Publisher.

Herrera, A. (2005). *Efecto del tamaño de muestra y la razón de tamaños de muestra en la detección de funcionamiento diferencial de los ítems*. Tesis de Doctorado en Evaluación y Tecnología y Tecnología Informática en Ciencias del Comportamiento. Universidad de Barcelona. Barcelona, España.

Herrera, A. N. (1996). *Algunas consideraciones técnicas sobre la construcción de ítems de pruebas objetivas según la clasificación de objetivos educativos de Bloom*. Recuperado de <http://cmapspublic3.ihmc.us/rid=1GLSW1LYG20GY7YBH24/Consideraciones%20C3%A9cnicas%20sobre%20la%20Construcci%C3%B3n%20de%20Items%20de%20Pruebas%20Objetivas%20seg%C3%BAAn%20la%20Taxonom%C3%ADa%20de%20Bloom.pdf> el 1 de noviembre de 2013

Hoover, W., & Gough, P. (1990). *The simple view of reading*. *Reading and writing. Interdisciplinary* 2, 127-160.

ICFES. (2009). *Lineamientos generales SABER 2009 grados 5° y 9°*. Bogotá, Colombia. From Recuperado de [http://www.mineducacion.gov.co/proyectos/1737/articles194703\\_recurso\\_1.pdf](http://www.mineducacion.gov.co/proyectos/1737/articles194703_recurso_1.pdf) el 16 de noviembre de 2010

ICFES. (2009). *Manual para la construcción de ítems tipo selección de respuesta*. Recuperado de [http://www.ascolfa.edu.co/archivos/C1\\_1\\_M01%20Manual%20de%20construcci%C3%B3n%20de%20EDtems.pdf](http://www.ascolfa.edu.co/archivos/C1_1_M01%20Manual%20de%20construcci%C3%B3n%20de%20EDtems.pdf) el 1 de noviembre de 2013

- ICFES. (2010). Examen de estado de la educación media-ICFES SABER 11°, qué se evalúa y cómo se interpretan los resultados individuales. Bogotá, Colombia. Recuperado de [http://www.icfes.gov.co/index.php?searchword=saber+11&option=com\\_search&Itemid=307](http://www.icfes.gov.co/index.php?searchword=saber+11&option=com_search&Itemid=307) el 8 de Septiembre de 2010
- ICFES. (2010). Orientaciones para el examen de Estado de calidad de la educación superior SABER PRO (ECAES). Bogotá, Colombia. Recuperado de [http://www.icfes.gov.co/index.php?option=com\\_content&task=view&id=351&Itemid=44](http://www.icfes.gov.co/index.php?option=com_content&task=view&id=351&Itemid=44) el 16 de noviembre de 2010.
- Instituto de evaluación . (2010). La lectura en PISA 2009. Marcos y pruebas de evaluación. Recuperado de <http://www.mecd.gob.es/dctm/ievaluacion/internacional/lectura-en-pisa.pdf?documentId=0901e72b8072f8d9>
- Linacre, J. (1991). Rasch-Model Computer Programs ©. Chicago: Electronic Publication [www.winsteps.com](http://www.winsteps.com).
- Linacre, J. (2006). A User's Guide to Winsteps Ministeps: Rasch-Model Computer Programs. Chicago: Electronic Publication. [www.winsteps.com](http://www.winsteps.com).
- Lorenzo, J. (2001). Procesos cognitivos básicos relacionados con la lectura. Segunda parte: La memoria y su implicancia en el rendimiento lector. *Interdisciplinaria* 18 (2), 115-134.
- Molina, J., Sanmartín, J., & Pareja, I. (2000). Los bancos de ítems en el escenario actual de la medición psicológica con test. *Revista de Psicología General y Aplicada* 53 (1), 127 - 145.
- Montanero, M. (2004). Cómo evaluar la comprensión lectora: alternativas y limitaciones. *Revista de educación* 335, 415 – 427.
- Morales, P. (2006). Las pruebas objetivas: normas, modalidades y cuestiones discutidas.

- Universidad Pontificia Comillas. Recuperado de <http://www.upcomillas.es/personal/peter/otrosdocumentos/PruebasObjetivas.pdf> el 1 de noviembre de 2013
- Moreno, R., Martínez, R., & Muñiz, J. (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema* 16 (3), 490-497.
- Muñiz, J., & Hambleton, R. (1992). Medio siglo de Teoría de Respuesta a los Ítems. *Anuario de Psicología*, 52, 41-66.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Psicología Pirámide
- OCDE. (2006). PISA 2006. Marco de la evaluación. Conocimientos y habilidades en Ciencias, Matemáticas y Lectura. Recuperado de <http://www.oecd.org/pisa/39732471.pdf> el 6 de septiembre de 2012
- Perez, M. (2005). Evaluación de la comprensión lectora: dificultades y limitaciones. *Revista de educación*, 121-138.
- Santana, A. (2009). Efecto de la Razón de Tamaños sobre la Detección del Funcionamiento Diferencial del ítem mediante Regresión Logística. Tesis de grado de maestría, Universidad Nacional de Colombia. Bogotá, Colombia.
- Solorzano, J., & Montero, E. (2011). Construcción y validación de una prueba de comprensión de lectura mediante el modelo de Rasch. *Actualidades Investigativas en Educación* , 11 (2) 1-27.
- Tapia, J. (2005). Claves para la enseñanza de la comprensión lectora. *Revista de educación*, 63-93.
- Thorndike, R. (1995). *Teoría psicométrica*. México: Limusa.
- Vallés, A. (2005). Comprensión lectora y procesos psicológicos. *Liberabit. Revista de Psicología*, 49-61.

Van Dijk, T. (1992). La ciencia del texto, un enfoque interdisciplinario. Barcelona, España: Ediciones Paidós.

Van Dijk, T., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

## **ANEXOS**

**Anexo 1 FORMATO DE CONSTRUCCIÓN DE ÍTEMS**

|                   |          |         |             |
|-------------------|----------|---------|-------------|
| Constructor:      |          |         |             |
| Fecha:            |          |         |             |
| Situación:        | Personal | Público | Educativo   |
| Formato de texto: | Continuo |         | Discontinuo |
| Dificultad:       | Baja:    | Media:  | Alta:       |

|                    |  |  |  |
|--------------------|--|--|--|
| Proceso Cognitivo: | Recordar                                   | Análisis-Síntesis                          | Aplicación                               |
| Componente         | Función semántica de los elementos locales | Configuración del sentido global del texto | Del sentido del texto hacia otros textos |
| Competencia        | Interpretativa                             | Argumentativa                              | Propositiva                              |

|                 |
|-----------------|
| <b>Contexto</b> |
|                 |

|                  |
|------------------|
| <b>Enunciado</b> |
|                  |

|                              |
|------------------------------|
| <b>Opciones de respuesta</b> |
|                              |

|                                   |
|-----------------------------------|
| <b>Clave</b> (Respuesta correcta) |
|-----------------------------------|

|   |   |   |   |
|---|---|---|---|
| A | B | C | D |
|---|---|---|---|

| Justificación de la clave |            |            |  |
|---------------------------|------------|------------|--|
| La opción A es            | Correcta   | Justifique |  |
|                           | Incorrecta |            |  |
| La opción B es            | Correcta   | Justifique |  |
|                           | Incorrecta |            |  |
| La opción C es            | Correcta   | Justifique |  |
|                           | Incorrecta |            |  |
| La opción D es            | Correcta   | Justifique |  |
|                           | Incorrecta |            |  |

**Justificación de la pregunta** (explique la finalidad de la pregunta incluyendo competencia, componente y proceso cognitivo, así como las acciones cognitivas que debe realizar el evaluado para resolver correctamente la pregunta)

**Observaciones de construcción:**