

Aplicabilidad de la lingüística textual y la psicolingüística en la composición de los ítems de selección múltiple del examen de estado colombiano¹

Cristian Bohórquez

Universidad Tecnológica de Pereira
Risaralda, Colombia

Resumen

Se presentan los resultados de una investigación que indagó la aplicabilidad de la lingüística textual y la psicolingüística en la formulación de preguntas del examen de estado colombiano. El objetivo fue hacer un tanteo de error de construcción en una muestra de ítems proveniente del ICFES. El procedimiento aplicado permitió: 1. *explicar* el funcionamiento textual, semántico y cognitivo de la superestructura textual Ítem de Selección Múltiple, 2. *analizar* la muestra hallando una cifra de 23% de error de construcción y 3. proponer estrategias de *optimización* para la composición de cuestionamientos e interpretación de resultados.

Palabras clave: ICFES, evaluación, examen de estado colombiano, prueba Saber 11, diseño de pruebas de evaluación educativa, Ítem de Selección Múltiple.

Abstract

Applicability of Text Linguistics and Psycholinguistics in Multiple-Choice Questions of the Colombian State Exam

This article presents the results of an investigation on the applicability of text linguistics and psycholinguistics in the formulation of the Colombian state exam. The objective was to obtain a score of construction error in one corpus of ICFES items. The procedure aimed: 1. *to explain* the textual, semantic and cognitive functioning of the textual superstructure Multiple Choice Item, 2. *To analyze* the sample finding 23% of construction error, and 3. *to propose* strategies for optimizing the items composition and results interpretation.

¹ Este artículo se deriva de la investigación Intervención de la lingüística textual en la construcción de evaluaciones de tipo selección múltiple para el área de lenguaje, llevada a cabo por el autor, entre agosto de 2009 y noviembre de 2010, dentro del marco de presentación de trabajo de grado, Licenciatura en Español y Literatura, Universidad Tecnológica de Pereira.

Key words: ICFES, evaluation, Colombian state exam, Saber 11 test, educational test design, Multiple-Choice Questions

Résumé

Applicabilité de la linguistique textuelle et la psycholinguistique dans la composition des questions à choix multiple de l'examen d'état colombien

On présente ici les résultats d'une investigation sur l'applicabilité de la linguistique textuelle et la psycholinguistique dans la formulation des questions de l'examen d'état colombien. L'objectif était de faire un score d'erreur de construction dans un corpus de questions à choix multiple de L'ICFES. La procédure utilisée a permis de: 1. *Expliquer* le fonctionnement textuel, sémantique et cognitif de la superstructure textuelle Question à Choix Multiple, 2. *Analyser* le corpus avec un 23 % d'erreur de construction, et 3. *Proposer* des stratégies pour optimiser la composition des questions et de l'interprétation des résultats.

Mots clés : ICFES, évaluation, examen d'état colombien, test Saber 11, conception des tests, Questions à Choix Multiple.

INTRODUCCIÓN

La investigación se erigió sobre dos consideraciones básicas: la primera fue observar que los instrumentos de evaluación educativa empleados en la prueba de estado, los ítems de pregunta, son en esencia constructos de lenguaje. La segunda fue prestar atención a la controversia de origen popular que cuestiona la validez de un número indefinido de ítems de pregunta avalados y aplicados por el examen de estado colombiano. Con base en lo anterior, se pensó qué teorías y disciplinas lingüísticas podían aplicarse a una muestra de ítems para identificar instrumentos defectuosos.

En el estado del arte no se hallaron trabajos que se ocuparan de exponer ítems errados en las pruebas de estado desde alguna ciencia, disciplina o enfoque teórico. Tampoco se encontraron estudios que abordaran el tema del diseño de instrumentos desde teorías del lenguaje, lo que es coherente con la siguiente afirmación del profesor Haladyna (1990): "there is little scientific basis for multiple-choice item writing. Most item writing knowledge is based on personal experience or wisdom passed on from particular mentors".

La muestra observada consistió en treinta (30) ítems del área de lenguaje en la que se determinó error de construcción en siete (7) de ellos, lo que representó un 23% de error. Respecto a la procedencia de la muestra, los ítems se tomaron de los compartidos por el ICFES en su página web a manera de ejemplos de pregunta (ICFES, 2009a, 2010). El material usado propiamente en exámenes de estado es confidencial según la ley 1324 de 2009. Sobre el porcentaje de error hallado, para conocer si coincide con algún margen previsto por la institución, no se conoció alguna estadística pública que contabilice ítems descalificados con posterioridad a su aplicación. Un referente disponible es la anulación de dos ítems de pregunta de la prueba del concurso docente de 2009 (ICFES, 2009b). De cualquier modo, se entiende que cada pregunta del examen de estado debe ser un instrumento confiable.

La exposición de los yerros se dio a la luz de diversas teorías del lenguaje, como se detalla en el apartado de referentes teóricos, y se confirmó mediante un modelo de análisis que se compuso en este trabajo a partir de las teorías cognitivas de esquemas mentales, modelos de situación y procesos mentales que guían la coherencia (Vega, Díaz & León, 1999). Para llevar a cabo el análisis de la muestra fue fundamental la explicación textual del ítem de selección múltiple (I.S.M.) conseguida a través de los criterios de textualidad de Dressler (Renkema, 1999).

Los problemas hallados en los ítems erróneos no obedecían a deficiencias de redacción en el sentido gramatical más estricto, sino a dificultades de fondo que se manifestaban en la forma. El I.S.M. fue caracterizado como una superestructura textual susceptible de generar desacuerdos semánticos entre su enunciador y enunciatario, es decir los sujetos evaluador y evaluado, en la medida en que diverjan a nivel cognitivo sus respectivas estructuras semánticas. Los diversos esquemas mentales de los individuos, al aplicarse a la interpretación de un mismo modelo de situación (el cuestionamiento efectuado sobre el texto), generaban macroproposiciones diferenciadas, lo cual es un fenómeno intrínseco al funcionamiento de la cognición humana, según lo describe van Dijk (1983, pp. 75-76).

Se desconoce el motivo por el cual los ítems erróneos y el fenómeno cognitivo descrito han sido pasados por alto por el órgano de evaluación estatal. Con fundamento en esta circunstancia, se concluyó la necesidad

de fomentar investigación desde la ciencia lingüística, a través de sus diversas disciplinas, sobre el proceso de diseño de instrumentos en cualquier área del saber de la prueba. El lenguaje verbal escrito y la superestructura textual I.S.M. son mediadores de la medición cognitiva en todas las áreas que evalúa el examen de estado; por lo tanto los principios textuales, semánticos y cognitivos que gobiernan el ítem son intrínsecos de esta superestructura de manera independiente del marco cognitivo que establezca el interrogante.

La evaluación tipo selección múltiple (E.T.S.M.) es la metodología de medición cognitiva más estandarizada en el aparato educativo occidental. Para el caso de Colombia, el hecho se registra en la masa poblacional que evalúa. Según comunicados de prensa oficiales, alrededor de 600.000 escolares presentaron la prueba Saber 11 de septiembre de 2012, cerca de 165.000 universitarios presentaron la prueba Saber Pro de 2011 y 229.038 docentes fueron examinados en el reciente concurso magisterial de 2009. No obstante de la confianza gubernamental depositada en este tipo de examen y de la gestión del órgano estatal de evaluación, los resultados de esta investigación dieron razón a las quejas informales en esta ocasión, demostrando la existencia de ítems erróneos dentro de los avalados por el ICFES.

Este proyecto tuvo como propósito general colaborar en el mejoramiento de la práctica nacional de evaluación. Las anteriores cifras dan cuenta del enorme impacto poblacional de la evaluación estatal, a la que corresponde impartir juicios cognitivos sobre miles de personas cada año incidiendo en sus oportunidades académicas y laborales. Frente a esto, el hallazgo de una porción considerable de ítems de pregunta erróneos, demostrado desde la formalidad científica, constituye una señal de alarma. Corresponde a la ciencia lingüística abordar el constructo de lenguaje I.S.M. brindando apoyo a la pedagogía y a la psicometría² en la labor de diseño de instrumentos.

Es de agregar que los resultados de investigación conservan vigencia no obstante la anunciada reestructuración del examen de estado del año 2011, convirtiéndose en la prueba Saber 11. La revisión de la documentación de soporte expedida por el ICFES (2011) permitió

² Rama de la psicología que se ocupa de la elaboración de tests. Esta disciplina se ocupa de diseñar y aplicar procedimientos estadísticos para la interpretación de los resultados de los tests de medición cognitiva (Muñiz, 2010) y constituye uno de los principales soportes conceptuales de la prueba de estado (ICFES, 2000b).

conocer que no hubo reformas sobre la superestructura textual I.S.M. o de otra clase que afectara los resultados de esta investigación.

REFERENTES TEÓRICOS

Al respecto de la evaluación tipo selección múltiple existe una gran cantidad de producción bibliográfica, la cual convino clasificar para insertar adecuadamente los resultados de esta investigación en el estado del arte.

Según su fuente, la documentación se dividió en dos partes: a) la producida por expertos nacionales e internacionales en materia de evaluación y b) la producida por el ICFES, que se fundamenta en la anterior y por ello aportó aquí los principales referentes. Al mismo tiempo, se tomó de ambas fuentes solo la información relacionada con el diseño de instrumentos, que contiene dos clases de referencias a su vez: a) sobre la planificación de fondo del instrumento y b) sobre la forma o construcción textual del I.S.M., siendo esta última etapa, la escritural, la intervenida por el objetivo de investigación y los resultados.

Información sobre la práctica de diseño formal de ítems se encontró en documentación del ICFES. En *¿Cómo se elaboran las pruebas?* (ICFES, n.d.) se describe la “Metodología de diseño de especificaciones de las pruebas a partir del modelo basado en evidencias”, que puede resumirse en tres pasos: 1. la planeación de los conocimientos, habilidades o competencias que serán evaluadas, 2. *la construcción textual de las preguntas encargada a personal docente externo al instituto* y 3. la revisión y aplicación piloto de los ítems para su aprobación.

La fundamentación para la labor textual de los constructores de las preguntas es reseñada en un informe de Rocha (2008), integrante del Grupo de Evaluación del ICFES, que expone cada una de las fases del diseño de los instrumentos. Respecto a la fase de elaboración textual de las preguntas, compila una serie de reglas de composición derivadas de “la experiencia propia (...) combinada con la de entidades extranjeras y con los hallazgos de numerosos estudios” (Rocha, 2008, p. 11). Esta serie de lineamientos guarda coherencia con los determinados por expertos como Haladyna y Downing (1989) y Alderson, Clapham y Wall (2005).

El conjunto de criterios sustentados en estos trabajos recibió una valoración crítica negativa en esta investigación. Más que “reglas” son consejos (“suggestions”), como se logra apreciar en el fragmento: “General Item-Writing: 5. Use good grammar, punctuation and spelling consistently. (...) 9. Focus on a single problem. (...) 15. Avoid items based on opinions.” (Haladyna et al., 1989, p. 40). Desde diferentes disciplinas (semántica, pragmática, lingüística textual y psicolingüística) podría formularse un conjunto más específico de directrices en relación con el uso del lenguaje escrito en I.S.M.

El profesor Haladyna acepta que son escasas las bases científicas que direccionan la escritura de los ítems, que la mayor parte del conocimiento disponible tiene un origen empírico y que la atención gira sobre todo en torno al análisis de las respuestas:

The professional attention given to analyzing responses to multiple-choice test items is considerable and has led to advances in item response theories. In contrast, there is little scientific basis for multiple-choice item writing. Most item writing knowledge is based on personal experience or wisdom passed on from particular mentors. A paradox exists: we place more emphasis on analyzing responses than we do on how we obtain them. (Haladyna, 1990)

En la consulta del estado del arte no se hallaron estudios afines con el objetivo de identificar errores en I.S.M.; por lo tanto las incorrecciones de los ítems se analizaron mediante diversas teorías de las disciplinas del lenguaje: principios del signo lingüístico (Saussure, 1979), infortunios de Austin (Escandell, 1996), correferencia, intertextualidad (Renkema, 1999), progresión temática (Villegas, 1998), paráfrasis, transformaciones semánticas y macrorreglas (van Dijk, 1983, 1998), macroestructuras (van Dijk, 1995, 1998), cohesión (Renkema, 1999), coherencia (Vega et al., 1999; van Dijk, 1995), inferencias (Gárate, Gutiérrez, Luque, García & Elosúa, 1999), esquemas mentales y modelos de situación (Vega et al., 1999). De este modo el estudio exigió una descripción particular para cada caso de error. En consecuencia se construyó para este trabajo un modelo general de análisis basado en la exploración de los procesos cognitivos que guían la coherencia, de Vega, Díaz y León (1999).

Tras confirmarse la presencia de ítems errados, se incluyó en este trabajo una propuesta de optimización para el diseño de instrumentos

que partió de la serie de “problemas técnicos” identificados en el examen de estado (ICFES, 2000a). Se aportó adicionalmente un ejercicio experimental de interpretación de resultados que tuvo en cuenta las disposiciones de la ley 1324 de 2009 y la fundamentación del instituto en esta materia (ICFES, 2000b, 2011).

En la restante documentación del ICFES no se hallaron referentes relacionados con el objetivo de esta investigación (ICFES, 1999, 2000c, 2007), ni en los documentos de Lineamientos Curriculares (1998) y Estándares Básicos de Competencia (2003) expedidos por el Ministerio de Educación Nacional.

METODOLOGÍA

Tres etapas conformaron el procedimiento: *explicación* del objeto de estudio, *análisis* de la muestra y proposición de estrategias de *optimización* para el diseño de ítems e interpretación de resultados.

Dada la ausencia de teoría lingüística aplicada al diseño de instrumentos, la primera etapa tuvo por objetivo conformar las bases conceptuales sobre el I.S.M. que permitieran llevar a cabo la etapa de análisis.

Se centró la atención sobre el ítem de selección múltiple como unidad celular de este género evaluativo. Para la aplicación de la teoría textual se reflexionó previamente si el objeto merecía el tratamiento de *texto*, lo que se comprobó tras consultar los criterios de textualidad de Dressler (Renkema, 1999).

Seguidamente se procedió a caracterizar la superestructura que identifica al I.S.M. entre el conjunto de tipologías textuales, señalando y denominando sus respectivos componentes. Este procedimiento permitió observar las relaciones semánticas entre las partes y describir las leyes generales que gobiernan el I.S.M., dando explicación textual a los efectos cognitivos de: *acierto de respuesta*, *error de respuesta*, *error de construcción*, *efecto distractor de los enunciados incorrectos de respuesta*, *relación intertextual del ítem con su texto de evaluación*, *relación exofórica del ítem con el texto presente en la mente de los sujetos evaluador y evaluado*, y *relación semántica entre los enunciados del ítem*.

Esta etapa finalizó con la composición de un modelo general de análisis del I.S.M. que, en correspondencia con el fundamento de reproductibilidad del método científico, permitiría a cualquier investigador reproducir el análisis para corroborar los resultados obtenidos y llevar a cabo nuevos trabajos.

La conceptualización anterior posibilitó la segunda etapa denominada *análisis*, que guardó el objetivo principal del trabajo: indagar la existencia de errores de construcción en I.S.M. pertenecientes al examen de estado del ICFES.

La muestra observada consistió en treinta (30) ítems de selección múltiple tipo I, que componían el total del área de lenguaje de dos pruebas Saber 11 (ICFES, 2009a, 2010). Se eligió el marco de lenguaje para la extracción de los I.S.M. por constituir un ámbito neutro de análisis: a. por aislar la superestructura estudiada de contextos ajenos a lo lingüístico que pudieran influir en la apreciación del error, y b. porque se esperaba de la competencia comunicativa de los docentes evaluadores del área de lenguaje un control paradigmático del artefacto textual I.S.M.

En la toma de la muestra la mayor parte de variables no fue de libre elección: la cantidad de pruebas (2) y de ítems (30) sumaba todo el material disponible. Todos los ítems eran tipo I. No se consiguió acceder a material directo de evaluación atendiendo a lo contemplado en la ley 1324 de 2009 que determina su uso confidencial. No obstante, estos factores no constituyeron contrariedad alguna sino una circunstancia afortunada. La extensión de la muestra permitió una satisfactoria cantidad y calidad de observaciones y hallazgos. El tipo I de los ítems es la superestructura más común y representativa de la evaluación estatal; la inclusión de otros tipos de ítem (como el tipo X) habría restado unidad al trabajo que constituyó una primera aproximación al objeto de estudio. La observación de otros tipos de pregunta y hasta de otras áreas del saber requerirá de futuras investigaciones. También fue un evento propicio que la muestra consistiera en *ejemplos de pregunta* pues el material incluía la solución de los ítems.

Obtenida la muestra se detectaron los ítems sospechosos de error de construcción. La selección consistió en tomar aquellos I.S.M. en los cuales no se hacía posible obtener un razonamiento tal que coincidiera con la solución oficial del cuestionamiento.

Sobre la serie de ítems sospechosos de error (7 entre 30) se aplicó análisis de cada uno en dos etapas. Inicialmente se explicó la

incorrección particular por medio de las teorías lingüísticas, semánticas, pragmáticas, textuales o cognitivas que exigiera el caso y que fueron referidas anteriormente. Seguidamente, se aplicó el modelo de análisis de coherencia para constatar el resultado obteniendo una descripción más puntual del error en el total de los casos.

La tercera etapa, denominada *optimización*, consistió en proponer estrategias en la construcción de I.S.M. y en la interpretación de los resultados. Contuvo tres estrategias. La primera, proponer la creación de un *banco de errores*. La segunda, contribuir en la solución de la serie de “problemas técnicos” del examen de estado (ICFES, 2000a). La tercera, una propuesta de interpretación de resultados basada en un experimento textual y semántico aplicado al I.S.M.

RESULTADOS

Explicación

1. Se comprobó la naturaleza textual del I.S.M. al cumplir con los siete criterios de textualidad de Dressler (Renkema, 1999), y se encontró que juega un papel fundamental la triada formada por *cohesión, coherencia e intertextualidad*.
2. Se describió la superestructura textual I.S.M. con el fin de convenir denominaciones. Se halló que el esquema de un I.S.M. consiste en la disposición de cinco enunciados ubicados en orden descendente, haciendo de la distribución espacial del texto un rasgo visual distintivo. El primero de ellos encabeza la estructura desde los factores espacial, textual y semántico planteando un interrogante, por lo que se denominó *enunciado principal* o también *enunciado de pregunta*. Los cuatro restantes son *enunciados secundarios* porque se desprenden en una relación subordinada del primero en sentido espacial, textual y semántico, constituyendo opciones para dar solución al cuestionamiento, por lo que se denominaron también *enunciados de opción de respuesta*.³

3 Esta taxonomía controvierte la expuesta por Rocha (2008, p. 10) que divide el ítem en *enunciado y opciones de respuesta*, en razón de establecer coherencia con la clasificación de unidades discursivas que permitan el análisis de ítems desde las teorías del lenguaje, pues “opciones de respuesta” no es una unidad en ningún nivel de la lengua.

Figura 1. Ejemplo de I.S.M. "Adaptada de ICFES, 2010"

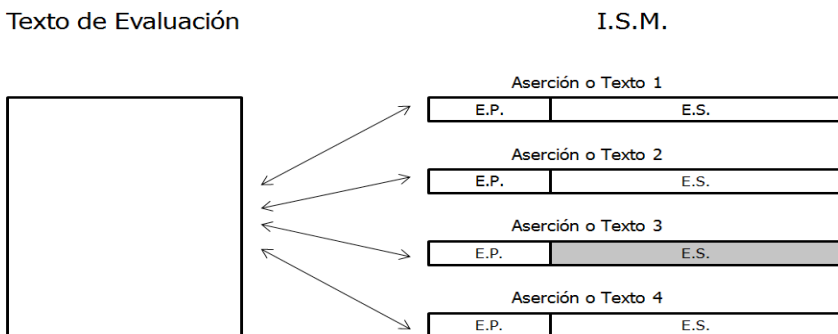
1. Con la expresión que inicia el segundo párrafo: "Pero eso no es todo...", el autor de la carta pretende } Enunciado Principal
 - A. negar lo que ha dicho en el párrafo anterior. }
 - B. reafirmar lo dicho hasta ese punto. }
 - C. completar una información que ha dado. }
 - D. contradecir lo que ha dicho. }

3. Sentar las anteriores convenciones hizo posible referir y describir la interacción y el funcionamiento de los componentes que integran el mecanismo texto-semántico I.S.M.
- 3a. Se observó que el enunciado principal se relaciona con los enunciados secundarios por medio de la *cohesión*, formando cuatro textos y no solo uno como se aparenta. Cada texto se denominó una *aserción*⁴. Para ejemplificar, las siguientes cuatro aserciones corresponden al ítem de la figura 1. La cursiva señala el enunciado secundario:
 - a. Con la expresión que inicia el segundo párrafo: "Pero eso no es todo...", el autor de la carta pretende *negar lo que ha dicho en el párrafo anterior*.
 - b. Con la expresión que inicia el segundo párrafo: "Pero eso no es todo...", el autor de la carta pretende *reafirmar lo dicho hasta ese punto*.
 - c. Con la expresión que inicia el segundo párrafo: "Pero eso no es todo...", el autor de la carta pretende *complementar una información que ha dado*.
 - d. Con la expresión que inicia el segundo párrafo: "Pero eso no es todo...", el autor de la carta pretende *contradecir lo que ha dicho*.
- 3b. Se observó una doble función de la cohesión en un I.S.M. Una es brindar economía evitando repetir el enunciado principal. Otra, más compleja, es aportar una lógica aparente a cada aserción, lo que explicó en parte el *efecto distractor* de los enunciados de opción: el sujeto evaluado cohesionaba en su mente los enunciados dispersos espacialmente en la superestructura y evalúa la lógica que representa cada unión.

4 El término usado por Rocha (2008, pp. 10-11) para referir esta unidad es proposición. Se propuso en su reemplazo la denominación aserción para precisar el carácter "asertivo" del enunciado (Searle, 1979, citado por Escandell, 1996, p. 65). Por otra parte el concepto proposición abarca exclusivamente el contenido lógico proposicional del enunciado (van Dijk, 1995, pp. 49-50).

- 3c. Lo que inspecciona en el fondo el sujeto evaluado es *la coherencia* de cada aserción. De ahí que el papel de la coherencia es, por una parte, dotar de sentido lógico la combinación de enunciado principal y secundario que resuelve el interrogante y, del otro lado, dotar de sentido aparente a las combinaciones erradas o distractoras.
- 3d. La coherencia de las aserciones se construye a partir del tercer elemento de la triada: *la intertextualidad*. Se definió como *texto de evaluación*⁵ a la unidad discursiva de la que parten los ítems a modo de cuestionamiento. En el examen de estado todo ítem parte desde un texto de evaluación (ICFES, 2011, p. 7) que puede ser un texto verbal o icónico (e.g. esquema, gráfico, mapa) (Rocha, 2008, p. 10). La siguiente figura representa el hecho:

Figura 2. Composición de aserciones e intertextualidad entre I.S.M. y texto de evaluación



La anterior figura explica como cada aserción del I.S.M. se compone a partir del texto de evaluación. La abreviatura E.P. indica el enunciado principal y E.S., el enunciado secundario. El E.S. resaltado en gris representa la opción correcta de respuesta, la que, en otras palabras, cohesiona una aserción coherente respecto del texto de evaluación.

4. Debe entenderse la coherencia como un proceso más cognitivo que textual: “La coherencia es una característica que, generalmente, se ha atribuido al discurso (v. gr., un texto es coherente o incoherente).

⁵ El término contexto empleado por Rocha (2008, p. 10) para referir esta unidad es un equívoco. No se trata de un “contexto”, que se refiere a las condiciones que rodean la situación comunicativa (van Dijk, 1983, p. 93).

Sin embargo, los psicolingüistas consideran que la coherencia es más bien una función cognitiva. Dicho de otro modo: *la coherencia está en la mente y no en el texto.*" (Vega et al., 1999, p. 273). La coherencia no reside en realidad en el I.S.M.; se construye en la mente a partir de la relación intertextual que conecta a cada aserción con el texto de evaluación y también con el sistema de informaciones depositado en la memoria a largo plazo del sujeto, en una relación exofórica.

5. Siendo la coherencia un proceso cognitivo más que textual, implica que el sentido lógico que adquiere determinada aserción del I.S.M. se fundamenta en el marco de conocimientos y experiencias del sujeto (evaluador o evaluado). Téngase en cuenta que "el lector de un texto entiende las relaciones de coherencia entre diversas partes del texto gracias, en gran medida, a sus conocimientos pragmáticos del mundo." (Vega et al., 1999, p. 273); estos conocimientos pragmáticos del mundo son los que dan origen a los *esquemas mentales* y a los *modelos de situación*, que pueden explicarse así:

Un esquema es una especie de promedio estadístico que resulta de procesar multitud de situaciones análogas, pero que no representa ninguna situación particular. El esquema es un prototipo que incluye información convencional sobre los objetos, personajes y roles, acciones características, etc., [...] En cambio, el modelo de situación [...] es una representación de un estado particular de sucesos, que incluye combinaciones únicas de parámetros del tipo «quién dijo qué (a quien)», «quién sabe qué», «dónde está qué o quién», etcétera. (Vega et al., 1999, p. 286)

Esta conceptualización es la que permitió entender cómo interviene el conocimiento en la resolución del I.S.M. En una prueba, la unidad discursiva denominada texto de evaluación establece un modelo de situación, es decir, una representación particular de sucesos. Mientras tanto, el conocimiento perteneciente a los distintos sujetos evaluador y evaluado constituye esquemas mentales, o sea las representaciones generales de la realidad.

6. La fundamentación expuesta en el anterior numeral permitió explicar el error de respuesta y de construcción del I.S.M. En la evaluación, los esquemas mentales de los sujetos evaluador y evaluado deben presentar una simetría respecto a la interpretación del modelo de situación. En otras palabras, el evaluado debe igualar los conocimientos requeridos por el evaluador para

interpretar el ítem y el texto de evaluación. En caso contrario el evaluado no podrá identificar la asección coherente, lo que da lugar al yerro de respuesta. Sin embargo, el sujeto evaluador es un hablante real, no ideal, por lo que su actuación (construcción del ítem) puede ser inadecuada, como se detalla en este artículo en el apartado de análisis.

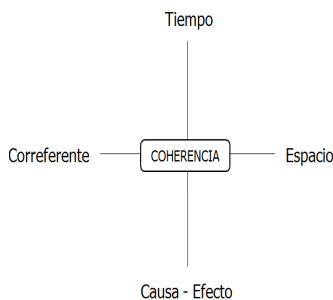
7. Indagando los principios cognitivos que guían la coherencia, pudo componerse un modelo de análisis que especifica a fondo por qué adquiere o no sentido cada asección.

Según los citados profesores, las estructuras mentales se actualizan guiándose por cuatro factores que determinan la coherencia: continuidad temporal, continuidad espacial, continuidad de causa-efecto entre sucesos y continuidad en el referente-correferente, lo que se explica de la siguiente manera:

Una vez fundada una estructura [mental], la información que se vaya procesando se incorporará a la misma, siempre que sea coherente con ella. Aunque no se define con mucha precisión qué se entiende por coherencia, se señalan cuatro factores que la determinan: la correferencia de las cláusulas, la continuidad temporal, la continuidad espacial y la continuidad causal entre sucesos descritos. (Vega et al., 1999, p. 284)

Se encontró que dentro de estas mismas cuatro dimensiones adquieren o rechazan la coherencia las asecciones de un I.S.M., dando origen al modelo:

Figura 3. Modelo de análisis de coherencia de cuatro dimensiones



8. El *error*, tanto de construcción como de respuesta de ítem, se definió como un razonamiento inadecuado del sujeto evaluado o evaluador dentro de la (s) dimensión (es) de coherencia que implique determinado I.S.M. Un razonamiento inadecuado quiere

decir pasar por alto fundamentos inherentes a la lógica causal, temporal, espacial o correferencial al interpretar o componer el I.S.M. con respecto al texto de evaluación.

Análisis

En la muestra de 30 ítems se descubrió error de construcción en 7 de ellos, lo que equivale a un porcentaje de 23 %. A continuación se resume cada caso clasificándolo según el tipo de error observado. En el orden de esta descripción, la ubicación de los ítems erróneos dentro de las pruebas es: numerales 9, 10 y 11 (ICFES, 2009a); numerales 4, 6, 12 y 5 (ICFES, 2010).

Ítem desbordado del texto de evaluación: se encontró en un ítem que solicitaba establecer la relación semántica entre el texto de evaluación (que era apenas un fragmento de un texto más extenso) y el título del texto original y completo. El cuestionamiento del ítem desbordaba la circunscripción estricta del texto de evaluación y cobraba sentido solo con el texto original depositado en la mente del sujeto evaluador en una relación exofórica. Para el sujeto evaluado, en cambio, el texto de evaluación (un mero fragmento) no tenía correspondencia con la macroestructura o título del texto completo tal como lo disponía la aserción que pretendía resolver el ítem. Se trató de una incoherencia en la dimensión causa-efecto: el texto de evaluación no proveía los suficientes datos (causa) que justificaran la coherencia de alguna de las aserciones de respuesta (efecto).

Relativismo macroestructural: se halló en un ítem que solicitaba identificar el tema del texto de evaluación. La definición del tema de un discurso es un proceso cognitivo en el que intervienen las macrorreglas para la síntesis semántica de la macroestructura. Van Dijk (1983, pp. 52, 75–76; 1998, p. 52) demuestra la posibilidad de que información proveniente del marco de conocimientos del sujeto se introduzca imprevistamente durante este proceso, siendo ésta la situación del ítem analizado. Se trataba de una divergencia entre las macroestructuras que conformaban los sujetos evaluador y evaluado. Desde el modelo se explicó que el texto de evaluación no ofrecía los suficientes datos (causa) que justificaran la coherencia de alguna aserción del ítem (efecto).

Paráfrasis inadecuada: partiendo de van Dijk (1983, pp. 211–216) pudo entenderse el parafraseo como un proceso cognitivo más que

textual expuesto a una “transformación semántica” por acción de las macrorreglas que median entre el input y el output. Al igual que en el caso anterior, en el trascurso de la omisión – adjunción – permutación – sustitución y (re) combinación pueden insertarse informaciones subjetivas. El ítem referido albergaba una incoherencia en la dimensión causa-efecto: ninguna de las aserciones del ítem constituía una paráfrasis (efecto) con sentido equivalente a la estructura semántica del enunciado original del texto de evaluación (causa).

Introducción arbitraria de correferente: error de construcción dado por el razonamiento inadecuado del sujeto evaluador en la dimensión correferencial de la coherencia. Algunos enunciados de opción de respuesta del ítem introducían un correferente arbitrario, no convencional al texto de evaluación ni a la lengua. La expresión problemática buscaba referirse a una situación del texto de evaluación extensa de explicar con palabras precisas. El correferente usado, aunque económico, resultaba coherente únicamente para el esquema mental del sujeto evaluador.

Ambigüedad o imprecisión semántica: de modo similar al yerro anterior, el sujeto evaluador interpretaba un hecho narrado en el texto de evaluación refiriéndolo bajo distintas expresiones como opciones de respuesta. El ítem requería que también el sujeto evaluado interpretara el hecho narrado con la misma denominación. La situación propiciaba la asimetría entre los marcos de conocimiento de los sujetos evaluador y evaluado, pues el sentido de los significantes utilizados como correferentes resultaban variables en función del trayecto vital de experiencias de cada sujeto. La ambigüedad se ubicó en la dimensión causa-efecto de la coherencia: los datos del texto de evaluación no constituían necesaria *causa* lógica de alguna de las aserciones del ítem (efectos).

Uso de correferente inadecuado para el contexto: error de construcción ubicado en la dimensión correferencial de la coherencia y que implicaba también las dimensiones de tiempo y espacio. En la aserción válida de respuesta el sujeto evaluador utilizó una expresión de correferencia inadecuada dentro del contexto espacio-temporal del texto de evaluación. Si bien en este ítem el equívoco no impedía la resolución de la pregunta, ya que procesos cognitivos de inferencia permitían llegar a la solución de manera indirecta, constituía sin embargo un error de formulación del cuestionamiento.

Error en la determinación y/o comunicación de la respuesta: se encontró que en un ítem correctamente construido (dado que podía resolverse) la opción comunicada como solución oficial no coincidía con la aseveración coherente que resolvía el cuestionamiento. La opción oficial en cambio originaba una incoherencia en la dimensión causa–efecto. Se concluyó la existencia de un error tipográfico o de transcripción en las soluciones.

Optimización

Se contemplaron tres estrategias de optimización del diseño de ítems a partir de las disciplinas lingüísticas. La primera consistió en la creación de un Banco de Errores en el cual se clasifique los diversos tipos de yerro y se los catalogue con descripción y explicación lingüística de su naturaleza. Investigaciones posteriores se encargarían de ampliar el banco cuyo objetivo es dar noticia, a evaluadores e investigadores, de los posibles tipos de error en I.S.M.

La segunda estrategia fue colaborar con la solución de los denominados “problemas técnicos” que observa el órgano evaluador en la prueba de estado (ICFES, 2000a, pp. 25-36). Para esto se halló aplicabilidad de los estudios del lenguaje en cada una de las problemáticas, como se explica a continuación:

Ítems de crédito parcial: desea el instituto “pensar en preguntas cuyas opciones (todas o algunas) sean respuestas correctas en cierta medida (...)” (p. 25). Este objetivo se apreció preocupante para la validez de la medición, teniendo en cuenta el hallazgo masivo de errores en ítems tipo I cuya característica es ofrecer única opción de respuesta. Se sugirió que tal problemática podría revisarse a partir de un trabajo exploratorio de las posibilidades textuales del ítem con respecto a la coherencia de diversas aseveraciones.

Comparación de resultados en el tiempo: es importante para el instituto determinar el grado de dificultad de un I.S.M. para ofrecer un rango de dificultad comparable entre los sucesivos exámenes de estado. Se propuso indagar la medición y representación numérica del factor dificultad teniendo en cuenta la clase de procesos cognitivos y el estado de maduración del esquema mental que exija el ítem al sujeto evaluado.

Reconocer la pluriculturalidad: para el organismo evaluador es importante introducir la pluriculturalidad del territorio nacional como

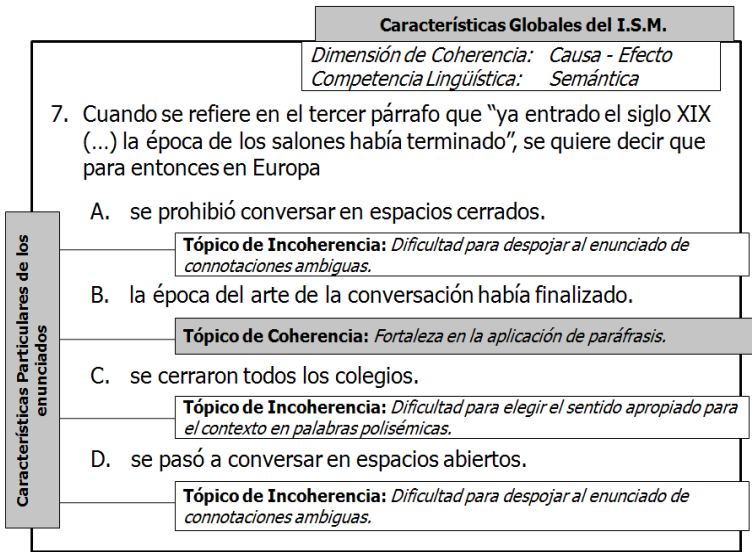
una variable para la composición de los exámenes de estado. Se planteó observar la variación diatópica, diastrática y diafásica a través de la sociolingüística y el análisis [crítico] del discurso, con el fin de hallar la mutua afectación entre la E.T.S.M. y el discurso social y lograr identificar nuevos criterios reguladores de la construcción del examen de estado.

Puntajes con significado y variedad en los resultados: en palabras del instituto, se busca ofrecer “Una información más detallada y, ojalá, de tipo cualitativo [que] podría contribuir a reflexiones importantes a nivel personal o institucional (...)” (p. 27). Por naturaleza el I.S.M. arroja un resultado de tipo binario (correcto/incorrecto). Este constreñimiento genera gran parte de la polémica informal en torno a la pertinencia de la E.T.S.M. debido a la escasa información que brinda al sujeto evaluado. La técnica de interpretación de resultados que implementó el instituto para solucionar esta contrariedad fue agrupar los diversos ítems en torno a *competencias y componentes* (ICFES, 2011) dando origen al formato actual de resultados (ICFES, 2000a, p. 41). Adoptando el anhelo del instituto y atendiendo al criterio de *cualificación* de la ley 1324 que promueve “la realización de ejercicios cualitativos, de forma paralela a las pruebas de carácter cuantitativo, que contribuyan a la construcción de explicaciones de los resultados en materia de calidad”, se experimentó con una modificación de la estructura semántica del I.S.M. para tratar de obtener resultados más amplios y precisos. Dicha tentativa se explica a continuación.

Experimento de interpretación de resultados

Sin modificar la superestructura del I.S.M. ni su carácter binario, la técnica experimental se centró en la planificación de fondo del ítem, de tal modo que cada aserción elegida por el sujeto evaluado indicaba de paso la circunstancia cognitiva que le hacía hallar coherencia en dicha opción de respuesta. Teniendo en cuenta que el ensamblaje de cualquier aserción del I.S.M. (E.P. + E.S.) genera una coherencia o incoherencia respecto del texto de evaluación, la estrategia consistió en definir para cada enunciado secundario del ítem un *tópico de coherencia*. Esto significa un enunciado que, en términos de fortaleza o dificultad —atendiendo al léxico educativo (M.E.N., 1998)— describa la circunstancia informativa que lleva al sujeto evaluado a elegir dicha opción, sea ésta correcta o incorrecta. Para representar la modificación de la estructura semántica del I.S.M., véase la siguiente figura:

Figura 4. Planeación del I.S.M. “Adaptada de MEN, 1998”



De la figura puede observarse que la superestructura del I.S.M. como tal se mantiene intacta: presenta un E.P. más cuatro E.S., o sea la pregunta y sus opciones de respuesta. Este es el texto que seguiría apareciendo dentro de la prueba. Los objetos adicionales que se aprecian en la figura se ubicarían en una planeación del ítem por lo cual no serían visibles.

La planeación consiste en dos apartados que explicitan las características necesarias del funcionamiento semántico del ítem para componer los enunciados de cualificación. En el apartado de *características particulares de los enunciados* (ver lateral izquierdo) se define el *tópico de coherencia* para cada uno. Toda aseveración del ítem tiene definido en el fondo el enunciado descriptor o tópico. Al resultar elegida la opción, el tópico de fondo constituye el juicio sobre la competencia del sujeto evaluado. La acumulación de dichos enunciados o tópicos de coherencia se listarían en la hoja de resultados del examen, brindando así al sujeto evaluado una cualificación de su estado cognitivo.

La explicitación de los tópicos de coherencia depende de las *características globales del I.S.M.* (ver esquina superior derecha), en donde se indican la dimensión de coherencia y la competencia lingüística (MEN,

1998) desde las que se desempeña el mecanismo semántico del ítem. A partir de estas características, el sujeto constructor del ítem puede enunciar los respectivos tópicos. Un tópico de coherencia es básicamente una inferencia del sujeto evaluador, un enunciado en el que trata de prever la circunstancia intelectual que llevará al evaluado a elegir una u otra opción de respuesta.

CONCLUSIONES

El examen de estado colombiano Saber 11 constituye un ejercicio de medición cognitiva de impacto trascendental en la sociedad colombiana. Con base en sus resultados se determina el acceso a la profesionalización de los cientos de miles de estudiantes que presentan la prueba en cada aplicación, y el pleno del aparato educativo nacional (esto abarca desde el M.E.N. hasta las instituciones educativas y sus docentes) moviliza periódicamente innumerables procesos académicos y administrativos que transforman la prestación del servicio educativo en el país. Esta tarea tan significativa solo puede tener éxito mediante instrumentos de evaluación confiables, calidad que se ve cuestionada por los resultados de esta investigación que encontraron una cifra de 23 % de error en una muestra de ítems aprobados por la metodología de diseño de instrumentos del ICFES.

No se concluyó que este porcentaje de yerro fuera indicador de una estadística constante ni tampoco resultante de un caso excepcional. Corresponde al ICFES observar estos resultados de investigación, rastrear el origen de los ítems defectuosos y tomar medidas correctivas.

Se consideró que una de las medidas que debe observar el instituto es impulsar investigación interna y externa sobre el diseño de ítems a partir de los estudios del lenguaje, tomando en cuenta que este enfoque demostró aportes relevantes en la identificación de ítems erróneos e indicó preguntas de investigación adicionales dirigidas a solucionar los “problemas técnicos” que encuentra el instituto en el examen de estado.

De conformidad con el artículo 3° de la ley 1324 que exige “al ente rector de la política de evaluación promover la participación creciente de la comunidad educativa en el diseño de los instrumentos y estrategias de evaluación”, el ICFES convoca anualmente desde el 2010 propuestas

de investigación para ser revisadas en un “seminario internacional de investigación sobre la calidad de la educación”. Tras observar que en las anteriores convocatorias no se registraron estudios sobre el tema de elaboración de ítems desde el marco conceptual de los estudios del lenguaje (Peña, 2011), se propone en este artículo que el ICFES podría incluir y anunciar este enfoque de *lingüística aplicada al diseño de instrumentos* como una de sus temáticas prioritarias para una próxima versión del evento⁶.

Finalmente, cabe aclarar que los resultados de esta investigación no se publican con el ánimo de orquestar una crítica corrosiva hacia el Instituto Colombiano para la Evaluación de la Educación, ICFES. Al contrario, se espera que estos resultados encuentren eco dentro de esta institución y de la comunidad científica, donde puedan ser discutidos con el objetivo colectivo de fortalecer la calidad de la evaluación en el país.

REFERENCIAS

- Alderson, J., Clapham, C. & Wall, D. (2005). Item writing and moderation. En *Language test construction and evaluation* (pp. 40-72). Reino Unido: Cambridge University Press.
- Dijk, T. (1983). *La ciencia del texto* (309 p.). España: Ediciones Paidós.
- Dijk, T. (1995). *Texto y contexto* (357 p.). España: Ediciones Cátedra.
- Dijk, T. (1998). Macroestructuras semánticas. En *Estructuras y funciones del discurso* (pp. 43-57). México: Siglo Veintiuno Editores.
- Escandell, M. (1996). Austin y la filosofía del lenguaje corriente. En *Introducción a la pragmática* (pp. 43-60). España: Editorial Ariel.
- Gárate, M., Gutiérrez, F., Luque, J., García, J. & Elosúa, M. (1999). Inferencias y comprensión lectora. En *Comprensión lectora y memoria operativa* (pp. 33-53). Barcelona: Ediciones Paidós.
- Haladyna, T. & Downing, S. (1989). A taxonomy of multiple choice item writing rules. *Applied measurement in education*, (2)1, 37-50. Consultado el 11 de octubre de 2012 en http://mdcbquestions.ahhuh.com/Haladyna_Item_Writing_Rules.pdf
- Haladyna, T. (1990). Advances in Item Design. En Institute for Objective Measurement (Comp.). *Rasch Measurement Transactions*. Consultado el 5 de marzo de 2012 en <http://www.rasch.org/rmt/rmt42b.htm>

⁶ En la versión del 2011 del seminario fue temática de investigación prioritaria “el aporte a la comprensión científica de las brechas encontradas entre hombres y mujeres”. (Peña, 2011)

- ICFES. (1999). Antecedentes y marco legal del examen de estado. Consultado el 12 de abril de 2012 en http://web.icfes.gov.co/index.php?option=com_docman&task=doc_view&gid=1177&Itemid=59
- ICFES. (2000a). Transformaciones en las pruebas para obtener resultados diferentes. Consultado el 13 de octubre de 2012 en http://www2.icfes.gov.co/examenes/component/docman/doc_download/105-examen-estado-2000-trasformaciones-en-pruebas-para-obtener-resultados-diferentes?Itemid=
- ICFES. (2000b). Sicometría. Consultado el 13 de octubre de 2012 en http://www2.icfes.gov.co/examenes/component/docman/doc_download/104-examen-estado-2000-sicometria?Itemid=
- ICFES. (2000c). Propuesta general. Consultado el 13 de octubre de 2012 en http://www2.icfes.gov.co/examenes/component/docman/doc_download/107-examen-estado-2000-propuesta-general?Itemid=
- ICFES. (2007). Sobre las pruebas saber y de estado: una mirada a su fundamentación y orientación de los instrumentos en lenguaje. Consultado el 13 de octubre de 2012 en http://www2.icfes.gov.co/examenes/component/docman/doc_download/102-marco-teorico-2007-area-lenguaje?Itemid=
- ICFES. (2009a). Ejemplos de preguntas. Consultado el 13 de octubre de 2012 en http://www2.icfes.gov.co/examenes/component/docman/doc_download/37-ejemplos-de-preguntas-nucleo-comun-saber-11?Itemid=
- ICFES. (2009b). Comunicado. Consultado el 13 de octubre de 2012 en http://www2.icfes.gov.co/examenes/component/docman/doc_download/116-comunicado-docentes-2009-reclamaciones-resultados?Itemid=
- ICFES. (2010). Ejemplos de preguntas. Consultado el 12 de abril de 2012 en http://www.icfes.gov.co/index.php?option=com_docman&task=doc_download&gid=4681&Itemid=59
- ICFES. (2011). Orientaciones para el examen de estado de la educación media ICFES Saber 11. Consultado el 13 de octubre de 2012 en http://www2.icfes.gov.co/examenes/component/docman/doc_download/32-guia-orientaciones-para-el-examen-de-la-educacion-media-icfes-saber-11?Itemid=
- ICFES. (n.d.). ¿Cómo se elaboran las pruebas? Consultado el 13 de octubre de 2012 en <http://www2.icfes.gov.co/examenes/acerca-de-las-evaluaciones/como-se-elaboran-las-pruebas>
- Ministerio de Educación Nacional. (1998). *Lineamientos curriculares lengua castellana*. Bogotá: Cooperativa Editorial Magisterio.
- Ministerio de Educación Nacional. (2003). *Estándares básicos de competencia del lenguaje*. Consultado el 13 de octubre de 2012 en http://www.mineducacion.gov.co/cvn/1665/articles-116042_archivo_pdf1.pdf
- Ministerio de Educación Nacional. (2009) Ley 1324 de 2009. DO. N°47.409

- Muñiz, J. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. *Papeles del psicólogo*, 31(1) 57-66. Consultado el 27 de abril de 2013 en <http://www.redalyc.org/articulo.oa?id=77812441006>
- Peña, M. (2011). Presentación del II seminario de investigación sobre la calidad de la educación. En *II seminario de investigación sobre la calidad de la educación*. Bogotá.
- Renkema, J. (1999). *Introducción a los estudios sobre el discurso* (pp. 52-59). España: Editorial Gedisa.
- Rocha, M. (2008). Diseño de pruebas evaluación educativa. Consultado el 13 de octubre de 2012 en http://www.colombiaaprende.edu.co/html/mediateca/1607/articulos-156625_archivo.unknown
- Saussure, F. (1979). Naturaleza del signo lingüístico. En *Curso de lingüística general* (pp. 127-145). Argentina: Editorial Losada.
- Vega, M., Díaz, J. M. & León, I. (1999). Procesamiento del discurso. En Vega, M. & Cuetos Vega, F. (Comp.). *Psicolingüística del español* (pp. 271 – 305). España: Ediciones Trotta.
- Villegas, C. (1998). El estudio de la estructuración discursiva por medio de la progresión temática. *Lenguaje*, 26, 76 – 92.

SOBRE EL AUTOR

Cristian J. Bohórquez Betancur

Licenciado en Español y Literatura, reconocido con la beca Jorge Roa Martínez de la Universidad Tecnológica de Pereira para estudios de posgrado. Miembro del grupo de investigación *Estudios del Lenguaje y la Educación*, categoría A de Colciencias. Áreas de interés académico: aplicabilidad de los estudios del lenguaje en el diseño de instrumentos de evaluación educativa, mecanismos cognitivos de la coherencia implicados en la comprensión del discurso.

Correo electrónico: castellblanco@hotmail.com

Fecha de recepción: 14-04-2012

Fecha de aceptación: 25-02-2013