

# ¿Qué criterios subyacen a la evaluación en pruebas de E/LE? Construcción a partir de una investigación exploratoria

## What Criteria Underlie the Evaluation of Tests in Spanish as a Foreign Language tests? Construction from Exploratory Research

Claudia Forero

claudia.acasa1@gmail.com

Docente contratista programa de Maestría en Educación, Universidad Libre, Bogotá, Colombia

### Resumen

El presente artículo describe un estudio piloto cuyo objetivo inicial buscaba desvelar rasgos característicos de pruebas de E/LE en un marco de principios teóricos contemporáneos. El estudio se llevó a cabo con un corpus de exámenes de cuatro instituciones educativas de Bogotá, entre 2009 y 2010. Con una metodología exploratoria intentamos identificar aquellos rasgos que de manera distintiva ayudaran a configurar la estructura interna y externa de un examen. Mediante un instrumento construido para el efecto, se examinaron las muestras y se obtuvieron resultados reveladores en dos áreas mayores: a) Contexto y fines de la prueba y b) Configuración lingüística. El análisis de los hallazgos manifestó que los postulados teóricos vigentes, que presumiblemente subyacían a las pruebas de E/LE, se interpretaron y/o se representaron de manera paradójica. Estos resultados fueron asociados también con las dificultades enfrentadas durante el estudio, pues al inicio del mismo, se apuntó a validar pruebas de certificación locales, sin embargo, en la exploración se recopiló sólo un corpus reducido de exámenes de clasificación y aprovechamiento. Esta dificultad se subsanó parcialmente, ya que en la segunda fase sólo tuvimos acceso a algunas muestras de tres modelos internacionales de certificación de E/LE del contexto panhispánico. No obstante, y aunque las conclusiones extraídas han sido provisionales, pueden contribuir al campo de la evaluación y en consecuencia generar un impacto en la didáctica. Reiteramos que el objetivo inicial se proyecta a un instrumento de metaevaluación que se publicará una vez la segunda fase del estudio proporcione todos los efectos de validación y evaluación.

**Palabras clave:** *evaluación, prueba, Español como Lengua Extranjera, instrumento, metaevaluación.*

### Abstract

The aim of this pilot study was to explore and try to show the constitutive features of language tests in Spanish as a foreign language in a current theoretical and conceptual framework. The study was carried out with the corpus of Spanish exams collected in four educative institutions in Bogotá city, between 2009-2010. With the exploratory methodology implemented, we attempted to identify the constitutive internal and exter-

nal features from the exam structure. This X-ray was taken with a grill criterion built for that purpose. The results were revealing in two main areas; a) Context and purposes of the test, and b) linguistic structure. The analysis showed that the theoretical postulates and principles underlying the test were interpreted and represented in a paradoxical way. These results are also associated with the problems faced during the study. It means that at the beginning of the study the objective was to validate proficiency tests from the local context. However, it was only possible to collect a reduced number of placement and achievement tests. These problems were partially overcome in the second phase of the research because just some samples from three models of international proficiency exams in the Spanish speaking countries were gathered. Nevertheless, even though the conclusions achieved were temporary, they allowed us to contribute to the field of evaluation and, as a result, impact the didactics. We further reiterate that the initial aim has been projected to a meta-evaluation instrument that will be published once the second phase of the research has accomplished the evaluation and validation process.

**Key words:** *evaluation, test, Spanish as a foreign language, instrument, meta-evaluation.*

El cambio de perspectiva que han sufrido los procesos de enseñanza-aprendizaje de lenguas extranjeras desde la mitad del siglo XX y lo que llevamos del siglo XXI, nos ha llevado a examinar la lente con que solíamos evaluar a nuestros estudiantes. De hecho, si el mismo proceso “enseñanza-aprendizaje” cambia de orden a “aprendizaje-enseñanza” se abandona la idea de aprendizaje supeditado a la enseñanza y muy seguramente el actor principal del proceso, que ahora es el estudiante, asume un rol verdaderamente protagónico.

Dentro de este marco es innegable que el giro que se le da a la lengua desde el modelo estructuralista y desde el enfoque comunicativo, impulsa e impacta las diferentes áreas constitutivas de la evaluación. Aquí se incluyen tres aspectos básicos: la competencia comunicativa, la tarea y el constructo. A la fecha de realización del estudio constatamos la escasez bibliográfica que diera cuenta de investigaciones empíricas que abordaran la evaluación en Español como Lengua Extranjera E/LE<sup>1</sup> en el contexto colombiano, y específicamente un método que ayudara a configurar la estructura interna de las pruebas de evaluación certificativa. El único estudio relacionado fue llevado a cabo por Mejía y otros (2009) del Instituto Caro y Cuervo y la Universidad de los Andes sobre *los Exámenes de evaluación en Colombia* en el que presenta una parte de la panorámica de la evaluación en E/LE y sugiere una serie de reestructuraciones. Otro documento es el de Yagüe Barredo, A. (2010), *Examinando exámenes. Consideraciones acerca de las pruebas de medición en ELE* y su preocupación por los exámenes, en el cual se plantean algunas características de las pruebas y se hace un análisis de sus posibles debilidades. Desde otra perspectiva, que también motivó el objeto del estudio, se encontraban los planteamientos del Sistema Internacional de Certificación de Español

---

<sup>1</sup> Aquí no se hace la distinción Español Lengua Extranjera y/o Segunda por prestarse a debate. Utilizamos la sigla E/LE de manera genérica.

como Lengua Extranjera (SICELE) que buscan no sólo apoyar el creciente interés por la lengua española, sino fomentar servicios de calidad en el marco del español, incluidos la construcción de herramientas de verificación de la calidad.

Con lo anterior, de alguna manera, se constataba que dicho campo de estudio había sido poco explorado en el contexto colombiano e implicaba que no se contaba con puntos de referencia metodológica similar o con resultados análogos de los cuales se pudiera partir, aspectos que hacían de este estudio un valioso aporte al tema. En consecuencia, con nuestra investigación no sólo se evidenciaba una contribución para la evaluación en E/LE, sino para la didáctica. Al apoyar la reflexión en pro de la articulación entre didáctica y evaluación y ampliar el debate creciente en el campo, nos interesa ayudar a minimizar el corte que se hace patente entre la dinámica del aula y la evaluación (qué se busca que el estudiante aprenda y qué se busca enseñar) y la manera en la que se operativizan los instrumentos. Igualmente, nos exige promover la formación permanente en evaluación: evaluadores, diseñadores, docentes-evaluadores, y evaluadores-investigadores que impulsen una evaluación de calidad, afín con las necesidades de usuarios y como respuesta a las exigencias del mundo globalizado y a sus políticas lingüísticas.

Es necesario puntualizar que el estado de la cuestión es amplio y rico en cuanto a marco teórico y conceptual en la lengua inglesa y en producción de pruebas de dominio y/o certificación. Exámenes en Inglés como Lengua Extranjera y/o Segunda de gran reconocimiento se encuentran en el mercado académico, así como numerosos estudios de validez y fiabilidad es el caso del TOEFL 2000 del documento de Chapelle, Grabe y Berns (1997), conocido como Modelo COE, que analiza los componentes más importantes desde la competencia comunicativa. En español<sup>2</sup>, por su parte, además de los estudios alrededor del DELE, se destacan los coloquios, proyectos e investigaciones relacionados con el examen Certificación del Español Lengua y Uso<sup>3</sup> (CELU) del consorcio Argentina-Brasil que *certifica mediante un examen oral y escrito la competencia lingüístico-comunicativa oral y escrita en Español como Lengua Extranjera en situaciones cotidianas, familiares, sociales, laborales o académicas, de hablantes alóglotas, sin tomar en consideración sus lenguas maternas ni el tipo de adquisición o aprendizaje que hayan realizado* (Prati, 2007, p.62). En cuanto a la Universidad Autónoma de México (UNAM), además de su amplia trayectoria en la disciplina, les ha apuntado a proyectos de largo alcance que incluyen las pruebas Examen de Español como Lengua Extranjera para el Ámbito Académico (EXELEAA) y el Servicio Internacional de Evaluación de la Lengua Española (SIELE) que merecen mencionarse por su impacto y lugar en el contexto panhispánico. En suma, los elementos anteriores respaldaron los objetivos propuestos en nuestra investigación, pues documentaban las necesidades en el campo que nuestro contexto inmediato podría requerir.

Desde el ámbito general se tiene que el objetivo general de la investigación exploratoria estaba encaminado a descubrir rasgos que caracterizaran pruebas de Español

---

<sup>2</sup> En la fase dos de la investigación se estudiarán argumentos teóricos y conceptuales relacionados con la certificación en E/LE y en otras lenguas.

<sup>3</sup> <http://www.celu.edu.ar/>

como Lengua Extranjera dentro de un marco de principios teóricos vigentes. Derivados del anterior, el primer objetivo específico buscaba determinar, con base en la lingüística aplicada y la didáctica de las lenguas extranjeras, qué constructos teóricos convendría incorporar a una estructura general de análisis que sirviera como instrumento para orientar procesos de construcción de pruebas internacionales de dominio y/o certificación en Español como Lengua Extranjera dentro de los modelos de evaluación actuales. El segundo objetivo pretendía elaborar y pilotear un protocolo de análisis que, producto del marco referido, permitiera identificar los rasgos más distintivos de las pruebas de dominio y/o certificación y de los principios que sustenta el modelo pedagógico que les subyace. El tercer objetivo consistía en aplicar dicha herramienta de análisis a un corpus de pruebas de E/LE con el fin de validar la utilidad y versatilidad de la herramienta en el análisis de categorías de construcción de pruebas.

En esta perspectiva general, situamos la investigación que buscaba llegar a un retrato de caracterización de una prueba de Español como Lengua Extranjera más allá de su estructura superficial. El corpus utilizado dio cuenta de muestras representativas de exámenes de cuatro instituciones de educación superior de carácter privado y de educación no formal de la ciudad de Bogotá, con la mayor trayectoria en la disciplina. Con el fin de explorar e identificar tales características distintivas y a través de una parrilla de criterios - a modo de cuestionario - construida previamente para el propósito, nos planteamos dos preguntas básicas en el desarrollo de la investigación. En la primera, nos preguntamos si era posible evidenciar criterios taxonómicos característicos particulares para cada prueba; y, en la segunda, si tales criterios manifestaban algún patrón que se pudiera asociar con un tipo particular de propuesta metodológica de aprendizaje-enseñanza.

Los resultados del análisis apuntaron a pruebas de diferente naturaleza en las que se evidenciaron aprovechamiento, dominio y clasificación y por lo mismo señalaron objetivos diversos y operativización variada. Dado que nuestra investigación era de carácter exploratorio y descriptivo en un contexto particular en el que se enfrentaron dificultades relacionadas con un acceso restringido a pruebas de lengua extranjera, bien por carencia de las mismas, por falta de difusión de información o por políticas de privacidad, los resultados se interpretaron con la debida cautela y las conclusiones y reflexiones no se pueden entender como absolutas.

No obstante, consideramos que aun así, nuestro estudio puede resultar de interés y utilidad para la coyuntura actual que a partir del SICELE (2007-2015) y SICELE Colombia se ha planteado en relación con la construcción de pruebas de lengua en clave de calidad y como respuesta a las demandas de certificación actuales. Adicionalmente, surge un planteamiento referido a la obligatoria interrelación didáctica-evaluación, y no a la evaluación como producto secundario de la didáctica o como proceso independiente de la misma, especialmente la evaluación con fines de clasificación o certificación. Es más, acudimos a la evaluación como una ciencia compleja y con identidad propia, como lo sugieren los expertos en el campo. Así, en el presente artículo se abordará en un primer momento los instrumentos de análisis utilizados en el presente trabajo, la descripción del corpus, la metodología de la investigación, la construcción de la herramienta de análisis y se presentarán y se discutirán algunos resultados.

## Metodología

El constructo teórico que sustentó el estudio aquí descrito resultó amplio y complejo pues requirió entretener un esquema compatible entre los diversos fundamentos teóricos de la evaluación, la evaluación certificativa, la didáctica de las lenguas extranjeras y la lingüística aplicada y sus ciencias afines. El punto de partida teórico contempló los planteamientos y las reflexiones perfiladas de los documentos y proyectos de los organismos que se consideran rectores en el área de evaluación y evaluación certificativa: la Association of Language Testers in Europe (ALTE)<sup>4</sup>, la European Association for Language Testing and Assessment (EALTA)<sup>5</sup> y la International Testing Association (ILTA)<sup>6</sup>. El Marco Común Europeo de Referencia para las Lenguas o MCER y el sistema SICELE<sup>7</sup>, por su parte, nos proporcionaron argumentos que favorecían el desarrollo de nuestra propuesta.

ALTE (2005, p. 46), por ejemplo, proporciona un modelo para el desarrollo de pruebas de seis etapas que incluyen desde la necesidad de construcción, pasando por la planeación, el diseño, el desarrollo, la operativización, hasta el monitoreo en la construcción de la prueba. EALTA (2008) se ha enfocado en promover la evaluación como disciplina profesional a través de principios base que incluyen objetivos, redacción de ítems, control de calidad, administración y revisión de pruebas, efecto washback y anclaje con los niveles del MCER. ILTA (2007) con su guía práctica y su código de ética propone una guía en pro de la profesión del evaluador y su profesionalización al tocar aspectos sensibles como lo que implica una buena práctica, las responsabilidades del evaluador y los derechos del usuario. Al MCER (2002) en su capítulo 9, se le reconoce un compromiso con el campo de la evaluación; especificaciones de contenido, niveles de dominio, ámbitos y tipos de evaluación. Con SICELE, desde la firma en Medellín en 2007, se concretan políticas lingüísticas que van orientando las tareas que asumen las entidades adscritas al sistema. El sello SICELE presume ventajas de calidad al abogar porque los sistemas de gestión y calidad se cumplan.

Entre los fundamentos teóricos tomados directamente y/o reformulados a partir de otras fuentes científicas se encuentran autores como Bachman, L. (1990); Bachman y Palmer, A. (1996, 2010); Bordón, T. (2004, 2006); Buck, G. (2001); Bygate, M., Skehan, P., y M. Swain (Eds.) (2001); Cárdenas Martínez, A. (2007); Celce-Murcia, M. (2007); Ellis, R. (1998, 2003); Figueras, N. (2008); Hughes, A. (2008); Jurado Salinas, M., y otros (2012); Nation, P. (2001); Nunan, D. (2004); Oxford, R. (2001); Parrondo, J. R. (2004); Pérez Ruiz, J. (2013); Prati, S. (2007); (Purpura, J. (2004); Skehan, P. (1998, 2001); y Yagüe Barredo, A. (2010), entre otros.

---

<sup>4</sup> <http://www.alte.org/> Association of language testers in Europe.

<sup>5</sup> <http://www.ealta.eu.org/> European Association for Language Testing and Assessment.

<sup>6</sup> <http://www.iltaonline.com/> International Testing Association.

<sup>7</sup> <https://sites.google.com/a/sicele.org/sicele/sobre-el-sicele/mision-vision-valores>.

Las fuentes mencionadas, además de arrojar luces en el tema de nuestro interés, nos llevaron a construir la plataforma que determinaría los ejes centrales subyacentes a las pruebas de lengua extranjera. Destacamos básicamente la competencia comunicativa, el uso de tareas de lengua y el constructo. No obstante, consideramos aquí necesario aclarar que en español es evidente que requerimos abordar algunos conceptos desde la lengua inglesa, no sólo por ser ésta referente de amplias investigaciones en el área, sino porque de alguna manera se ofrecen distinciones importantes para el tema de nuestro interés. Empezamos entonces por plantear las sutiles diferencias entre “Assessment”, “Evaluation” y “Test”, así:

**Assessment:** In language testing, the measurement of one or more aspects of language proficiency, by means of some form of test or procedure (Multilingual glossary of language testing terms: studies in language testing 6. ALTE<sup>8</sup> Members, 1998, p.135).

**Evaluation:** gathering information with the intention of using it as a basis for decision-making. In language testing, evaluation may focus on the effectiveness or impact of a programme of instruction, examination, or project (Multilingual glossary of language testing terms: studies in language testing 6. ALTE Members, 1998, p.144).

**Test:** A procedure for testing specific aspects of proficiency or knowledge.

- a) A set of components which together constitute an assessment procedure, often used-to-mean-the-same-as-examination.
- b) A single task or component for assessing an area of skill or knowledge, e.g. speaking.or-writing.
- c) An assessment procedure which is relatively short and easy to administer often devised and administer within an institution or used as part of a research programme or for validation purposes. (Multilingual glossary of language testing terms: studies in language testing 6. ALTE Members, 1998, p.166).

Con base en lo anterior, en español se habla de valoración, apreciación, recogida de información relacionada con el dominio de lengua, con desempeños, con efectividad, con impacto y con habilidades para tomar decisiones. Bachman, por su parte, refiere a los términos de la siguiente manera:

“Measurement” and “test” involves the quantification of observations, and are thus distinct from qualitative descriptions. Tests are a type of measurement designed to elicit a specific simple of behavior. “Evaluation” involves decision-making, and is thus distinct from measurement, which essentially provides information. Thus, neither measures nor tests are in and of themselves evaluative, and evaluation need not involve measurement or testing (Bachman, 1990, p.49).

---

<sup>8</sup> Association of Language Testers in Europe.

Esto nos lleva al término *evaluación* en sentido amplio de proceso:

Un proceso permanente y sistemático de análisis, de los eventos que ocurren en y como consecuencia del proceso enseñanza-aprendizaje y de los demás procesos inherentes, que permite potenciar las capacidades y habilidades de los participantes en el proceso educativo; afianzar los aciertos; analizar y corregir los errores, reorientar los procesos escolares, motivar, construir el conocimiento teórico y práctico, afianzar valores y actitudes y liderar (educadores y educandos) el uso social del conocimiento (Carrillo, 2002, p.p. 4-5).

Una vez tenemos una aproximación conceptual de donde partir, seguimos con la distinción de las perspectivas con base estructuralista y sociolingüística que han influido ampliamente y de diferente manera en la evaluación en lengua extranjera. Desde los enfoques estructuralistas, Saussure, el lingüista suizo, sienta las bases de la lingüística moderna en el siglo XX que servirán de inspiración para otras obras y trabajos, incluidos los del estructuralismo norteamericano de Bloomfield, quien, en su obra *Language* (1933), presenta una visión conductista de la lengua y propone algunos lineamientos sobre cómo se aprenden las lenguas y cómo se pueden enseñar. En la psicología, Skinner (1957) y el conductismo tienen gran influencia en el aprendizaje de las lenguas: este aprendizaje es similar al de otros aprendizajes y se genera a partir de un estímulo que produce una respuesta para convertirse en un hábito. Sus tesis se constituyeron en paradigma por décadas ya que ejercieron gran influjo en la enseñanza, el diseño de materiales y, por supuesto, en la evaluación de lenguas extranjeras con la aparición de pruebas psicométricas. Con Chomsky (1957-59) se produce una revolución en la lingüística por la posición innatista de la lengua que plantea, ya que se aparta totalmente del conductismo y el estructuralismo dominantes en décadas anteriores. Al respecto, se supone que Chomsky, además de aportar los conceptos de competencia y actuación, tenía una visión mentalista de la lengua que estaba en consonancia con los principios básicos del cognitivismo y que tenían que ver con la capacidad creativa de la mente: entender y producir expresiones nuevas. La predisposición a la que se refiere Chomsky, y que poseemos todos los seres humanos, es inconsciente e innata y esa gramática universal (GU) se actualiza en una gramática inconsciente de una lengua, llamada competencia lingüística, es decir el conocimiento implícito que tiene el hablante del sistema lingüístico como tal (Da Silva Gomes, 2005, p.19). El enfoque estructuralista se constituye con base en sílabos, materiales y métodos hasta finales de los 70 y se constituyó en la base para desarrollar pruebas de lengua con orientación de puntos discretos. ALTE cita por ejemplo, obligatoriamente, a Lado (1961), Valette (1967), Harris (1969) y Heaton (1975).

Por su parte, una nueva perspectiva desde la sociolingüística y la etnografía viene con Hymes (1967), quien retoma las aportaciones de Austin, Grice y el mismo Labov al llamar la atención aspecto sociocultural que no había sido considerado por las ideas chomskianas; al adquirir su lengua materna, el niño adquiere la competencia para formar oraciones gramaticales, para saber cuándo hablar y cuándo no, y para saber qué hablar dependiendo de su interlocutor, del contexto y del lugar. Ese conocimiento o competencia comunicativa para Hymes (1995), está compuesta de cuatro dimensiones

relacionadas con:

- a) La competencia lingüística (actuación lingüística)
  - b) La factibilidad (de las expresiones según edad, formación, fatiga...)
  - c) La aceptabilidad o adecuación (según reglas sociolingüísticas del contexto)
  - d) La realidad (probabilidad de uso del enunciado por un nativo en el contexto).
- (En: Llobera, M., 1995. p.p. 13-22).

El modelo ha sido reformulado por autores como Canale y Swain (1980), quienes retoman la propuesta y la traducen al conocido modelo compuesto por cuatro (sub) competencias: gramatical (uso de recursos lingüísticos para una buena actuación lingüística); sociolingüística (adecuación del discurso con el contexto, estatus de los participantes, interacción); discursiva (cohesión y coherencia); y estratégica (recursos lingüísticos y no lingüísticos). Otra reformulación es el modelo de habilidad lingüística comunicativa – CLA - de Bachman y Palmer (1996, 2010), quienes proponen una modificación al modelo de habilidad de lengua en el cual se activan e interactúan tres características mentales; el conocimiento de la lengua (sistema), el esquema del conocimiento (conocimiento y experiencia del mundo) y el esquema afectivo (memorias emocionales).

Con este modelo, que se caracteriza por su dinamismo, se plantea el contexto para cumplir con tareas comunicativas porque ayuda a perfilar el escenario base para definir el área de competencia que se va a evaluar:

A model of language ability is of use to the language tester because it provides a basis for defining the area of competence to be tested. Having a clear idea of what is being tested is a prerequisite for being able to decide whether or not a test is valid, and makes possible the devising of such useful tools as checklists for test content (ALTE, 2005, p.19).

Hasta aquí hemos sugerido que existe una interrelación evaluación - didáctica y que una aproximación a los conceptos de evaluación, competencia comunicativa y tarea de lengua es básica para la construcción de pruebas de lengua. De hecho las tendencias actuales de evaluación de lenguas, en el marco de la teoría vigente, tienen un trasfondo que debe reflejar una visión de la lengua y la lengua de uso del momento, según ALTE:

What is being tested and the kind of task or item type chosen as a means of testing can be expected to show the influence of current ideas on what constitutes language ability and what exactly we are doing when we use language in our everyday lives (ALTE, 1995, p.9).

Ahora bien, los llamados métodos comunicativos, el objetivo primordial que se deriva de estos es el desarrollo de la competencia comunicativa del hablante para diferentes fines y en contextos variados. Al respecto, Siguán (1995, p.8) sostiene que:

Los métodos que en conjunto conocemos como comunicativos no sólo se

proponen como primer objetivo conseguir una competencia comunicativa adecuada a las necesidades del aprendiz sino que basan su metodología en la explotación sistemática de situaciones comunicativas en las que participan los alumnos. Pero en realidad y desde una perspectiva más profunda, lo que hacen estos métodos es intentar que el alumno aprenda una segunda lengua en forma parecida a como adquirió la primera.

Es clara la apreciación de una lengua real para un uso real de la misma que debiera prevalecer en los procesos de aprendizaje de las lenguas, lo que viene a materializarse de alguna manera desde la perspectiva del MCER que, por su parte, adopta el término de “tarea” definida como: “Cualquier acción intencionada que un individuo considera necesaria para conseguir un resultado concreto en cuanto a la resolución de un problema, el cumplimiento de una obligación o la consecución de un objetivo (de comunicación)”.(MCER, 2002, p.20).

El concepto de tarea, que ha circulado por más de dos décadas, ha sido caracterizado por diferentes autores (Long, 1985; Crookes, 1986; Richard y Rogers, 1986; Breen, 1987; Nunan, 1989 y 2004; Peris, 1995; Bachman y Palmer, 1996; Bygate, 2001) desde variados ángulos: tareas de diversa clase que suponen un desarrollo y activación de competencias específicas y estrategias; tareas que implican una cadena de actividades, se llevan a cabo en diferentes ámbitos, pueden tener como objetivo la solución de problemas, la comprensión de información, la participación en un debate, la organización de un texto, y la respuesta a un correo electrónico, por ejemplo; y, tareas que pueden ser sencillas o complejas e incorporar unas *sub-tareas*, lo que hace difícil caracterizarlas. En otras palabras, las tareas son ante todo *comunicativas*, pues los participantes realizan actividades de interacción, comprensión y mediación:

Las tareas forman parte de la vida diaria en los ámbitos personal, público, educativo o profesional. La realización de unas tareas por parte de un individuo supone la activación estratégica de competencias específicas con el fin de llevar a cabo una serie de acciones intencionadas en un ámbito concreto con un objetivo claramente definido y un resultado específico (MCER, 2002, p.155).

Las tareas pueden ser tan variadas que incluyen la creatividad, la planeación, la interpretación, la resolución de problemas, la participación en debates y la respuesta a un texto, por ejemplo. Ya que las tareas comunicativas son una forma de operativizar las competencias y/o habilidades de lengua, tienen un carácter pedagógico en la medida en que buscan desarrollar competencias útiles para la vida real del estudiante. Así, para la realización de una tarea se requieren ciertas condiciones y restricciones: la activación de competencias generales y específicas, el seguimiento de reglas preestablecidas, la consideración de factores afectivos, cognitivos y lingüísticos, conocimientos en tema específicos, y destrezas organizativas, interpersonales e interculturales (MCER, 2002, p.p. 155-166).

Finalmente, en cuanto al “constructo”, se sugiere que antes de evaluar se comprenda su naturaleza:

An understanding of what we are trying to measure is called a construct, and our test will be useful and valid only if it measures the right construct. Thus, the first task of the test developer is to understand the construct, and then, secondly, to make a test that somehow measures that construct. This is construct validity, and ensuring that the right construct is being measured is the central issue in all assessment (Buck, 2001, p.1).

Se requiere, entonces, medir el “constructo” correcto y para hacerlo hay que comprenderlo y hacer una prueba que de alguna forma lo mida, es decir, la validez del constructo. Chapelle (1997) sostiene que hay dos formas básicas de definir un constructo. La primera es definir la competencia - conocimientos, destrezas y habilidades - que se cree los estudiantes que toman una prueba deben tener, y la segunda, es definir las tareas que deberían ser capaces de ejecutar. Ambos métodos representan posiciones filosóficas diferentes, y por obvias razones presentan ventajas y desventajas. La tercera opción consistiría en tratar de combinar los dos métodos. Entonces, para comprender la naturaleza del constructo en cada una de las habilidades (escuchar, leer, hablar y escribir), se debe tener claridad sobre los aspectos de tipo cognitivo que entran en juego y cómo funcionan en el desarrollo de estas habilidades en la L2. Es decir, los procesos que se siguen, las habilidades que se busca desarrollar, las etapas y la evaluación y operativización más adecuadas.

Matizando todo lo anterior con nuestro objetivo inicial, es evidente que lo expuesto, además de ayudarnos a allanar el camino en la construcción y depuración de la herramienta de análisis, nos puso de manifiesto que la estructura subyacente a los instrumentos de evaluación en lengua extranjera requería (y requiere) de un mecanismo que ayudará a detallar rasgos de manera eficiente y desde diferentes ópticas, como se describe en el apartado.

### **Instrumento de análisis**

El instrumento de análisis para la validación de pruebas de lengua extranjera que se construyó, nació como un cuestionario sencillo de rasgos que de manera descriptiva reflejarán un examen de E/LE. Dichos rasgos se formularon con un tratamiento binario: es decir, el rasgo aparecía o no en la prueba. El cuestionario fue evolucionando hasta constituirse en una estructura jerárquica de niveles operativos diversos que revelaban las características de los mismos exámenes de lengua y los postulados teóricos esbozados a lo largo del marco teórico. Dicho cuestionario se enfocó en dos aspectos generales: *Propósitos y fines de la prueba* y *Configuración lingüística de la prueba*. El primer aspecto refiere al contexto, al usuario y a los objetivos; el segundo, se orienta a las tareas de lengua, las competencias, las actividades, las muestras de lengua y los conocimientos. Cada una de estas categorías está formulada en términos de descriptores implementados, reformulados y adaptados a partir del marco teórico, de la estructura de las muestras del corpus y de la construcción personal del investigador. Naturalmente, los descriptores *No se aplica (no observable o no pertinente)* y *Otro(s)*, según el caso, también se incluyeron. El procedimiento utilizado para recoger los datos a partir de las categor-

ías preestablecidas apuntó a un número absoluto (0 o 1) y desde esta perspectiva binaria, la estadística descriptiva nos evidenciaba el peso de cada una de las categorías.

### Descripción del corpus

El corpus del estudio estaba constituido por cuatro pruebas de E/LE de diferente naturaleza (clasificación y aprovechamiento) para adultos, que se recogieron en Bogotá (Colombia) en dos universidades y en dos institutos de lenguas extranjeras. Estos exámenes fueron diseñados por los coordinadores del área o por el profesor encargado de cada nivel o curso particular. En su confección participaron máximo dos profesores, licenciados en lenguas extranjeras, con experiencia en E/LE y algunos con postgrado en lenguas extranjeras. Cada una de las cuatro muestras se codificó y segmentó para facilitar la tarea de análisis. En la siguiente tabla aparece una maqueta general del corpus donde se detalla el tipo de examen (clasificación, dominio y/o aprovechamiento), nivel, institución de origen y actividades o ejercicios. En este último aspecto y, en aras de sistematizar la información, fue necesario rotular según lo que propusiera la prueba: ejercicio, tarea, ítem, pregunta, sección, reactivo (ver Tabla 1).

Cabe decir que, si bien los exámenes recogidos no constituyen *per se* un corpus amplio o de gran impacto, como lo esperábamos al inicio del estudio, los hallazgos pueden despertar interés pues se lograron consideraciones desde diferentes ángulos. De una parte, se recopilaban 457 reactivos que, analizados a la luz de un cuestionario de 347 descriptores, arrojaban aproximadamente 158.000 datos. Tales datos, una vez categorizados, nos proponen algunas reflexiones teóricas que podrían llevarnos por el camino de replantear la evaluación y la didáctica en lenguas extranjeras, como se esboza más adelante.

Tabla 1

*Descripción general de las muestras de E/LE analizadas.*

	Tipología	Institución	Preguntas/ ejercicios/ actividades/ tareas/ prueba/ sección/ reactivos
01	Clasificación Secciones: 9 Reactivos: 86	Universidad privada con 10 años de experiencia en el área de E/LE.	1.1.A - Gramática (diferentes categorías). Llenar espacios en blanco. Escogencia múltiple. 30 reactivos 1.1.B - Gramática (verbos). Llenar espacios en blanco. 10 reactivos. 1.1.C (Sin instrucción). Gramática (diferentes categorías). Llenar espacios en blanco. Escogencia múltiple. 33 reactivos 2.1. Gramática. Transformar oraciones. 5 reactivos. 2.2. Gramática. Transformar oraciones. 4 reactivos. 3. Producción escrita. 3.a. Ejercicio de uso del presente. 3.b. Ejercicio de uso del pasado. 3.c. Ejercicio de uso del futuro. 4. Producción oral (sin información)

02	Aprovechamiento (nivel intermedio) Secciones: 9 Reactivos: 120	Instituto privado. 30 años de experiencia en el área de E/LE.	<ol style="list-style-type: none"> <li>1. Gramática (verbos). Llenar espacios en blanco. 20 reactivos .</li> <li>2. Gramática (verbos en presente y pasado). Llenar espacios en blanco. 20 reactivos.</li> <li>3. Gramática. Preguntas de respuesta corta. 10 reactivos.</li> <li>4. Lectoescritura. Responder con frases largas sin omitir verbos. 9 reactivos.</li> <li>5. Escribir (8) frases sobre las rutinas en pasado con los verbos indicados.</li> <li>6. Escribir (10) frases sobre las rutinas en pasado de un amigo, con los verbos indicados.</li> <li>7. Gramática. Llenar espacios en blanco (preposiciones). 10 reactivos.</li> <li>8. Gramática. Llenar espacios en blanco (verbos). 10 reactivos.</li> <li>9. Gramática. Llenar espacios en blanco con diversos tiempos verbales. 23 reactivos.</li> </ol>
03	Clasificación y aprovechamiento Secciones: 15 Reactivos: 171	Institución privada. 10 años de experiencia en el área de E/LE.	<ol style="list-style-type: none"> <li>1. Gramática. Llenar espacios en blanco de tres diálogos (diferentes categorías). 26 reactivos.</li> <li>2a Gramática-léxico. Llenar espacios en blanco (artículos y verbos). 30 reactivos.</li> <li>2b. Gramática-léxico. Llenar espacios en blanco (artículos). 14 reactivos.</li> <li>3a. Gramática-léxico. Llenar espacios en blanco (artículos). 14 reactivos.</li> <li>3b. Léxico. Llenar espacios en blanco (escribir números). 9 reactivos.</li> <li>3c. Léxico. Llenar espacios en blanco (escribir números). 8 reactivos.</li> <li>4. Escritura. Responder en forma corta. 3 reactivos. Sin instrucciones explícitas.</li> <li>5. Gramática. Transformar oraciones. 2 reactivos.</li> <li>6. Gramática. Identificar corrección gramatical. 6 reactivos.</li> <li>7. Gramática: Completar los espacios en blanco (pronombres demostrativos). 5 reactivos.</li> <li>8. Gramática: Completar los espacios en blanco (comparativo-superlativo). 8 reactivos.</li> <li>9. Escritura-Gramática. Contestar preguntas en forma corta (presente y futuro de indicativo). 22 reactivos- *en la prueba no todas son preguntas, aunque así aparece en las instrucciones.</li> <li>10. Léxico. Escribir los “opuestos” de adjetivos y adverbios. 10 reactivos.</li> <li>11. Gramática. Llenar los espacios en blanco de oraciones (verbo ser o estar). 10 reactivos.</li> <li>12. Gramática. Corrección gramatical. Corregir y responder. 4 reactivos.</li> </ol>

04	Aprovechamiento (nivel intermedio) Secciones: 11 Reactivos: 80	Universidad privada. Más de 40 años de experiencia en el área de E/LE.	1. I. Lectura de un texto (539 palabras). Preguntas de falso o verdadero. Corrección oraciones falsas. 6 reactivos. 2. Escritura y ejercicio de comprobación de la comprensión y producción mediante preguntas de respuesta corta. 2 reactivos . 2.1.a. II. Gramática. Llenar espacios en blanco (verbos). 10 reactivos. 2.1.b. Escritura (párrafo). Sin repetir estructuras. 10 oraciones. 3.1.a. Gramática. Llenar espacios en blanco (preposiciones). 15 reactivos. 3.1.b. Gramática. Llenar espacios en blanco (modo indicativo y subjuntivo y perífrasis verbales). 10 reactivos. 4. III. Léxico. Sinónimos de palabras y expresiones. 10 reactivos. 5. IV. Escritura (carta personal de 120 palabras). Criterios: contenido, gramática, vocabulario, estructura, puntuación y ortografía. 6.1.a. Escucha (video). Escogencia múltiple. 2 reactivos. 6.1.b. Escucha (video). Respuesta corta. 4 reactivos. 6.1.c. Escucha (video). Identificar la respuesta (10 opciones).
----	--	--	--

## Metodología de la investigación

La metodología adoptada para el trabajo fue de tipo exploratorio, pues buscaba aproximarse a un fenómeno poco abordado en el contexto particular: la evaluación certificativa y la caracterización de pruebas de dominio en E/LE. Igualmente toma elementos de la metodología descriptiva y explicativa. Aunque se inició con un estudio piloto (primera fase), se evidenció que el objeto de estudio requería más de un enfoque, de ahí la necesidad de llegar a una investigación mucho más rigurosa que se proyectara a una segunda fase, y exigiera tanto de la investigación cualitativa como de la investigación cuantitativa. Larsen-Freeman y Long (1994, p.21) señalan que, si reflexionamos sobre estos paradigmas, *no hay nada inherente a ninguno de los dos enfoques que prohíba practicarlos coherentemente junto con el otro paradigma*. A través del paradigma cualitativo se hace un acercamiento a las pruebas de lengua en E/LE; se analiza la estructura del examen, se interpretan los elementos subyacentes o evidentes al diseño de pruebas dentro de los marcos de evaluación actuales y la metodología de aprendizaje-enseñanza que se podría asociar a tales pruebas.

La metodología seguida para la validez de la herramienta se hizo mediante una estadística descriptiva simple en la que se registra el valor relativo del rasgo: es decir, si la característica está o no está en la muestra analizada. Con esta estadística se detectaron los rasgos más característicos de cada muestra de examen, las diferencias que podrían considerarse más marcadas y las categorías que ayudarían a predecir un perfil más detallado o una orientación más precisa de una prueba de E/LE. Subrayamos que dicho re-

trato no equivale a lo más representativo de la “calidad”, sino que se presume pone de manifiesto los rasgos más inmediatos y característicos de una prueba de lengua extranjera y su peso relativo en el marco de características afines con la fundamentación teórica y conceptual vigente en el campo disciplinar. No obstante, en la primera fase del estudio, el enfoque de exploratoria nos obligó a adaptarnos a los hallazgos y a echar mano de una estadística descriptiva básica. Es importante aclarar que, en la segunda fase de la investigación (en etapa de evaluación), se cambia el corpus a modelos de exámenes de certificación internacionales del contexto panhispanico (el examen Certificado de Español Lengua y Uso CELU de Argentina, el Examen de Español para Propósitos Específicos EXELEAA de México y el examen Diploma de Español como Lengua Extranjera DELE de España), se amplían la herramienta de análisis y la perspectiva técnica de la evaluación y se implementan paquetes estadísticos que incluyen el uso del sistema SPSS19, los programas de Análisis Multivariante de Datos SPAD y de Análisis de Correspondencias Múltiples ACM y el Alpha de Cronbach para determinar aspectos como fiabilidad y validez.

### **Construcción de la herramienta de análisis**

La trayectoria general seguida para la construcción de la herramienta de análisis se desarrolló en cuatro fases mayores: planeación, diseño, construcción y pilotaje. En la primera, apuntamos a un modelo que se planteó dentro de un marco general de postulados teóricos vigentes en el campo de la lengua, la enseñanza-aprendizaje, la evaluación, la gestión y el diseño de pruebas de lengua extranjera, la lingüística aplicada y otras disciplinas afines, a partir de autoridades científicas en el área. En otras palabras, la naturaleza y estructura de la herramienta propuesta admite entrecruzar los fundamentos teóricos con los datos empíricos recabados, a partir de la experiencia del investigador y sus pares académicos. Así, buscamos un modelo flexible y dinámico y a su vez estructurado jerárquicamente que nos permitiera hacer un mapeado de rasgos.

En el diseño y construcción, y como parte de la fase exploratoria inicial empezamos un acercamiento intuitivo y libre de prejuicios, considerando qué características básicas debería tener un examen de corte comunicativo dentro de los sistemas de evaluación actuales. Ese punto de partida permitió la elaboración de un banco de descriptores con base en la experiencia y el corpus recogido; en segundo lugar, con base en las lecturas, precisiones y sugerencias planteadas por pares evaluadores se involucran diferentes puntos de vista teóricos que ayudaron a matizar y definir categorías de análisis que fue necesario formular y reformular. Las categorías que se establecieron compartían rasgos y características y estaban relacionadas con todos los aspectos que ayudaron a definir la base teórica. Los criterios anteriores se enriquecieron con los de los anejos propuestos en el MCER (2002, p.p.200-224) para el desarrollo de los descriptores de dominio de la lengua y las escalas ilustrativas y por ALTE, ILTA<sup>9</sup> y EALTA<sup>10</sup> en relación con la construcción de pruebas, auditorias y validación de pruebas. Un referente privilegiado lo constituyo, de hecho, el SICELE (2007), que entre cuyos objetivos pro-

---

<sup>9</sup> International Language Testing Association.

<sup>10</sup> European Association for Language Testing and Assessment.

pone desarrollar e implantar procedimientos y herramientas de verificación de la calidad de los sistemas de evaluación certificativa, enunciado al inicio del artículo.

De todo lo anterior se desprenden entonces las categorías generales de análisis, a saber (ver Tabla 2):

Tabla 2  
Descripción general de las categorías (8) de análisis que conforman la parrilla modelo.

Categoría	Descriptorios constitutivos (347)	Aspectos analizados
1. Contextos de aplicación	21	Refiere a un enfoque general de la prueba: tipo de prueba, quién, cómo y dónde se presenta.
2. Usuario	7	Refiere al perfil y a los requisitos
3. Objetivos de la prueba	31	Refiere a los objetivos desde diferentes perspectivas
4. Tareas aportadas y generadas	39	Refiere a tipologías y características
5. Competencias implicadas	51	Refiere a la competencia comunicativa y a sus componentes
6. Actividades lingüísticas	74	Refiere a las destrezas básicas y lo que implican
7. Muestras de lengua aportadas y generadas	47	Origen de los textos orales y escritos y variantes que representan
8. Conocimientos evaluados	67	Refiere a la lengua sistema: morfosintácticos, fonético fonológicos; gramática y léxico; discursivo-textuales; pragmáticos y socioculturales.

Los 347 descriptorios que se acopiaron, fueron categorizados jerárquicamente en las ocho categorías y las mismas se agruparon en dos áreas de análisis. En términos generales se buscó que los descriptorios tuvieran significado y sentido por sí mismos; que hubiera coherencia y cohesión en la formulación; que los enunciados presentaran variaciones sintácticas mínimas; y que hubiera homogeneidad en el estilo, características válidas para auditoria de materiales (Ezeiza, 2006, p.p.541-542). Igualmente, los descriptorios se formularon de manera afirmativa y buscando evitar cualquier ambigüedad o sesgo. En la última fase, una vez la herramienta se había depurado, las pruebas de lengua que constituyeron el corpus se analizaron a la luz del instrumento creado, es decir, se hizo un análisis frente a la teoría, las preguntas y los objetivos planteados. Aquí, se presume, por tanto, que el papel protagónico de la teoría adquirió su verdadera dimensión en la práctica, pues llevó a nutrir dicho modelo desde diferentes fuentes y planos, ayudando a ratificar su esencia holística e integral, abierta y dinámica.

## Resultados y discusión

En relación con el procedimiento y los instrumentos utilizados correspondientes al pilotaje de la parrilla de criterios propuesta y a la validación de las cuatro muestras analizamos los hallazgos generales interpretados esencialmente de manera cualitativa para cada una de las categorías.

En la Tabla 3 aparecen registrados los datos de cada una de las pruebas en cuanto a su peso relativo. Según se evidencia, el peso relativo varía en cada muestra y nuestra intención no es que se comparen estos interpruebas pues, dado el formato tan diverso (en número de tareas/ejercicios) para cada uno, no sería útil hacerlo. Sin embargo, estos datos sí nos dan una idea general de que los conocimientos evaluados y las actividades de comunicación son los de mayor peso en cada prueba. Igualmente nos sugieren un orden de categorías cercano.

Tabla 3  
*Peso relativo de las categorías en las muestras analizadas.*

Categoría	Muestra 01	Muestra 02	Muestra 03	Muestra 04
Conocimientos evaluados (67)	81	109	143	100
Actividades lingüísticas de comunicación implicadas (74)	56	72	96	83
Competencias implicadas (51)	45	58	82	83
Muestras de lengua (47)	41	54	72	69
Objetivos de la prueba (31)	34	46	69	76
Contextos de aplicación (21)	33	36	60	55
Tareas aportadas y generadas (39)	28	36	48	70
Usuario (17)	21	27	36	33

Se encontró que, aun cuando los exámenes de E/LE para adultos eran de diferente origen y naturaleza y se rotulaban (algunos) como de clasificación y aprovechamiento, su uso incluía diagnóstico y dominio. Es decir, se percibía cierta ambigüedad en la naturaleza y uso de las pruebas, aun cuando en términos generales compartían las mismas características estructurales.

### Categoría 1:

Para esta categoría se evidencia que los ámbitos de las pruebas abarcan lo institucional pero no alcanzan los ámbitos local y/o nacional. Son de carácter particular y de uso privado. En cuanto a la naturaleza y función, algunos tipos de prueba tienen doble funcionalidad: una prueba puede ser de diagnóstico y de clasificación al mismo tiempo,

o de dominio y de clasificación. En relación con las condiciones físicas y temporales, este aspecto representa el mayor peso relativo. La presentación con lápiz y papel es común a todas las muestras y la estética no se evidencia como un aspecto que se destaque (25% de las muestras).

#### Categoría 2:

Para esta categoría los resultados obtenidos señalan que la edad de los candidatos usuarios oscila entre adultos jóvenes y adultos. El perfil sociocultural no se observa, sin embargo, los exámenes se dirigen en general a estudiantes extranjeros de Español como Lengua Extranjera y en un caso se deduce que es a usuarios cuya lengua materna y/o primera lengua es el francés. En el 50% de las muestras no se indican exigencias previas y, en el otro 50% se deduce que el examen es un requisito posterior a un curso formal de lengua.

#### Categoría 3:

Para esta categoría los datos arrojados parecen indicar que el enfoque general de las pruebas se centra en los conocimientos lingüísticos. Los aspectos pragmáticos y socio-culturales no se hacen explícitamente evidentes y, si lo están en algunos apartados, éstos no alcanzan a representar el 2% de los apartados totales de cada muestra. En relación con los objetivos generales, se espera que los estudiantes “sepan ser/estar” para desempeñarse académicamente. En cuanto a los objetivos operativos de “saber hacer”, de manera genérica no se evidencia qué es lo que los candidatos deben “saber hacer”. En relación con las destrezas de comprensión y producción no es claro el peso que éstas tienen en las pruebas. Solamente en una prueba se evalúa explícitamente lectura, escucha y escritura, y la habilidad de habla se enuncia pero no se indica cómo se evalúa. En el 50% de las muestras aparecen tareas de evaluación de la producción escrita en términos de composición sobre información personal a un amigo y una carta de invitación a visitar el país de la lengua meta con sugerencias para ello.

#### Categoría 4:

Los resultados obtenidos para esta categoría señalan que no se podría hacer referencia a tareas comunicativas en el sentido en que lo propondrían autores como Prabhu (1987); Nunan (1989, 2004); Bachman y Palmer (1996); Bygate (2001), entre otros, que aluden a actividades de procesamiento de pensamiento (información) en las que se busca un producto (meta). Como reflexión válida con respecto de las tareas, Bordón (2006) cita a Purpura (2004, p.288): *¿qué buscamos con las tareas? desarrollar la conciencia gramatical, comprender o interiorizar formas o que el usuario realice una tarea comunicativa y que para ello tenga necesariamente que valerse de formas específicas de la lengua.*

Según se describe en la Tabla 1, los ejercicios de las tres muestras (75%) se buscan recuperar a través de estructuras lingüísticas especificadas previamente. En el 100% de las muestras de exámenes se proponen ejercicios de rellenar huecos para diferentes categorías: verbos, preposiciones, artículos, y léxico.

#### Categoría 5:

Para esta categoría, la competencia lingüística ocupa el primer lugar (85%) con respecto del conjunto de competencias que se evalúa. La competencia estratégica le sigue a la lingüística; esto se puede deber a que aparece sólo en una muestra pero en varios ejercicios. En una muestra se evidencia la competencia sociocultural en un 25%. Las competencias pragmática, textual-discursiva no se evidencian como tales.

#### Categoría 6:

En esta categoría el mayor peso relativo lo tiene la expresión escrita. La interacción no se evalúa en ninguna de sus formas y no hay una destreza que se destaque especialmente.

#### Categoría 7:

Los datos obtenidos indican que en el 75% de las muestras no se evidencia el origen de los textos orales y/o escritos. El restante 25% señala que el origen de los textos proviene de Internet. En cuanto al tipo de variante proporcionada, se presume que se dirige genéricamente a la variante del español de Colombia y no se pide una variante específica a los potenciales candidatos.

#### Categoría 8:

En esta categoría, la gramática representa el mayor peso, seguido por los conocimientos ortográficos y léxicos, estos dos últimos con el mismo peso relativo. El parámetro fonético-fonológico tiene el menor peso relativo. Los parámetros socio-cultural y discursivo-textual tienen el valor mínimo.

Desde otra perspectiva, los hallazgos se deben tomar con cautela pues el mismo corpus sugiere cierta ambigüedad. Los resultados de la parte empírica de este trabajo parecen indicar que no se reportan diferencias marcadas importantes en cuanto al tipo de exámenes y/o estructura y objetivos entre las diferentes muestras del corpus; por el contrario, llaman la atención los múltiples rasgos en común que tienen al tratar de determinar un mapa de las mismas.

Se tiene entonces que el conocimiento declarativo (saber) aparece con el mayor peso en las muestras. Le sigue el saber hacer - destrezas y habilidades - y no aparecen el saber ser - competencia existencial - ni el saber aprender - capacidad de aprender -. Por su parte, los resultados indican que el conocimiento gramatical se evalúa a nivel oracional. La morfología y la sintaxis están en primer orden y los tiempos gramaticales en cuanto a inflexiones verbales y valencias cobran toda la atención. En el parámetro relacionado con el léxico, los datos indican que el léxico que se evalúa se restringe a su significado denotativo y no al connotativo. Es decir, no se verifica su uso en contextos discursivos más amplios, sino que se enfoca en oraciones aisladas, o como sinónimos u antónimos en listas de palabras. Por su parte, se evidencia que la dimensión semántica es restringida y en consecuencia el candidato debe saber el significado de las palabras y usarlas con cierta pertinencia en contextos oracionales. En otras “actividades”, dentro de este mismo parámetro, se pide al candidato hacer correcciones semánticas y/o sintácticas en algunas oraciones.

En general, la producción que se le pide al candidato no demanda una variación léxica, según se evidencia en el análisis del corpus. En relación con la norma culta, ésta no se pide explícitamente, sin embargo se exige conciencia lingüística en la mayoría de los ejercicios; se busca distinguir los tiempos verbales y los modos. No se evalúa el uso de conectores y no se pide una estructura textual básica: introducción, desarrollo o cuerpo y cierre, según se deduce de los datos. Desde una perspectiva sociolingüística, en las pruebas realizadas, no hay evidencia de que el estudiante tenga que tomar decisiones en cuanto a entender y/o producir un registro con base en una situación comunicativa particular. Todo parece indicar que la lengua no se asume como mediadora entre el hombre y su entorno ni es la que le permite aprehender e interpretar la realidad dentro del panorama de lo posible y adecuado.

Se podría discutir, con base en los resultados, si la lengua que se evalúa en las muestras analizadas se podría denominar auténtica en este sentido, ya que se presupone que es a través de la interacción en una comunidad lingüística en la que la lengua adquiere su auténtica dimensión social. Según lo registrado, se encontró también que al solicitársele al candidato algún tipo de producción, ésta no se ubica en un contexto social - en cuanto a un grupo social de terminada edad, nivel sociocultural, procedencia u ocupación -. En cuanto a la distancia social, según los resultados, se evidencia a través de la forma de tratamiento de “usted” en tres exámenes, el uso del “tu” aparece en dos exámenes. Sin embargo, en una de las pruebas analizadas se detecta cierta inconsistencia en el uso de estas dos formas de tratamiento en las instrucciones dadas.

Según los datos, en cuanto a la competencia pragmática en las pruebas analizadas, se ha detectado que a los usuarios no se les pide generar una variedad de actos de habla. Los actos que se piden se solapan con una instrucción estructuralista: por ejemplo, *narrar en pasado usando las estructuras aprendidas en clase o hacer una comparación usando las estructuras adecuadas*. Estos datos denotan que al estudiante se le pide dar cuenta de los aspectos formales de la lengua y sus elementos más discretos; se deduce por lo tanto que no se busca que el estudiante modifique el *educto*<sup>11</sup>. Los resultados también señalan que no se piden explícitamente producciones cohesionadas y coherentes.

En cuanto a los criterios de evaluación de las competencias no se presentan en forma explícita, o, en el caso de que se haga, no son suficientemente claros. Igualmente ocurre con las rúbricas, las cuales no son transparentes en cuanto no indican los requisitos previos, ni el tiempo que se requiere para desarrollar la prueba, ni el valor para cada uno de los apartados, destrezas, secciones, ni el peso de cada una de las destrezas evaluadas, entre otros. En cuanto al número de tareas, los datos dejan entrever que este aspecto no es manifiesto: una prueba de aprovechamiento presupondría más tareas que una de clasificación, por ejemplo; los exámenes de clasificación y de nivel de dominio requieren de información clara porque implican toma de decisiones importantes y trabajo en equipo; los exámenes de aprovechamiento son de menor impacto porque se cuenta con otras tareas que han permitido evaluar continuamente al estudiante. Con base en lo

---

<sup>11</sup> Output (en inglés).

anterior, se podría presumir que la múltiple funcionalidad en la aplicación de los exámenes puede generar problemas en los resultados que arroje.

De otra parte, el análisis del corpus plantea la percepción de que los exámenes surgen de una metodología con base estructural y que a partir de esta metodología se vislumbra una posición pedagógica, un paradigma de lengua, una postura didáctica igualmente estructurales que sustentan, de una u otra forma, las decisiones tomadas en el diseño de las pruebas descritas. No obstante, debemos ser cautos y presuponer que podría haber un análisis de resultados diferente comoquiera que la herramienta propuesta, aunque ha sido sometida a diferentes procesos de validación por parte de expertos, podría construirse y ser susceptible de diferentes interpretaciones.

Finalmente, consideramos que los datos aquí presentados son aproximaciones y no se pueden asumir como irrefutables y concluyentes. Sin embargo, el aval que nos permitimos a partir de la etapa empírica y el análisis de datos tiene como fin capital contribuir al campo de la investigación de E/LE en Colombia que, aunque marcha lentamente, ya muestra tímidos progresos en el área de evaluación. En efecto, esperamos que los esfuerzos hechos y que se concretaron en un instrumento teórico-práctico de validación con principios sólidos, con el fin de aproximarnos al problema de la evaluación de pruebas, puedan constituirse en una herramienta útil por su capacidad predictiva de criterios de análisis. Es decir, al iniciarse un acercamiento a exámenes de evaluación sistemática de E/LE en el contexto actual, esperamos que la herramienta se convierta en un mecanismo con poder taxonómico, dentro de los límites naturales de este trabajo y bajo la premisa de que es un aporte inicial, una vez termine su proceso de validación actual.

## Conclusiones

Las posibles conclusiones que se derivarían del presente trabajo investigativo y que, a pesar de los esfuerzos, en esta primera fase, de alguna manera son provisionales, nos llevan a afirmar que dentro de ciertos límites uno de los retos planteados se resolvió: viabilizar de manera sistemática los rasgos presentes en las pruebas de lengua que, a la luz de los supuestos teóricos y conceptuales vigentes en los que se circunscribe la didáctica podrían ser particularmente rentables y próximos a la evaluación.

Los resultados de dicho estudio fueron los siguientes. Sí fue posible consolidar una metodología de análisis que ayudara a orientar a docentes e investigadores interesados en reflexionar, revisar y, si es plausible, afinar la calidad de sus criterios de evaluación, procesos evaluativos e instrumentos en E/LE aplicando un modelo de validación común. Segundo: sí fue factible evidenciar a través de un modelo coherente, sistemático y común el paradigma en relación con la lengua, el aprendizaje y la enseñanza que subyace a instrumentos estándar de evaluación en diferentes contextos. Los resultados arrojados y el posterior análisis nos llevaron a evidenciar un posible paradigma de lengua-aprendizaje-enseñanza que enmarca las pruebas analizadas en un 75% en una perspectiva de lengua basada en la forma. Tercero: las categorías de análisis propuestas en el modelo sí permitieron evidenciar criterios taxonómicos característicos distintivos para cada una de las pruebas. No obstante, la taxonomía resultante general manifestó un

patrón en el cual la gramática (estructuralista) parecía ser el hilo conductor de la metodología que se seguía en los cursos de E/LE. Las muestras indicarían que se evaluaba lo que se enseñaba: conocimientos gramaticales a nivel oracional. En síntesis, según los resultados, se detectó que se buscó que el estudiante diera cuenta de las formas que componen el sistema de la lengua, es decir lo que es gramaticalmente correcto.

En cuanto a las dificultades sorteadas en el acceso a la información (acceso a las pruebas e inexistencia de exámenes de certificación), ellas nos llevaron recopilar una muestra baja en número, aún cuando, considerando el contexto, se presume son representativas. En suma, podríamos afirmar que, desde los objetivos propuestos y las respuestas a las preguntas, todo lo pudimos abordar dentro de los límites propios del estudio: se construyó el marco teórico en relación con los criterios que podrían subyacer a una prueba de lengua de diferente naturaleza, dentro de un enfoque comunicativo y a partir de allí se diseñó un modelo de análisis por categorías que se implementó para dibujar una taxonomía de rasgos en dichas pruebas, es decir se confirmó la viabilidad de la herramienta en su tarea de escáner.

No podemos hablar de la comprobación de la hipótesis de la investigación pues en su carácter de exploratoria no sabíamos lo que podíamos encontrar en el contexto local. En efecto, las variables ya expuestas al principio determinaron el curso de la investigación. Otros aspectos que pueden generar inquietud corresponden a la fase 2 del estudio donde se incluye entre otros: la ampliación del marco teórico con relación a la construcción y desarrollo de pruebas, el análisis de estudios en los que se comparan exámenes de diversas lenguas para posibles fines de homologación, y la implementación de programas estadísticos e interpretación de datos para dar cuenta de aspectos técnicos de la evaluación certificativa, por ejemplo.

En el plano de las relaciones didáctica-evaluación se podría inferir que estamos ante fenómenos de “eclecticismo a priori” que de manera quizás dispersas “mezclan” bases teóricas y prácticas diferentes que tienden a confundirse por la diversidad de fuentes paradigmáticas que las sustentan. De ahí que, aunque se anuncian tareas y competencias comunicativas, se insiste en desarrollos estructuralistas, sintaxis oracionales fuera de contexto y léxicos denotativos aislados, entre otros, en últimas, se dice estar en el plano de la comunicación, pero bajo la tutela de la metalingüística.

No obstante, puede ser que los análisis hechos se perciban como un estudio sincrónico y que actualmente los exámenes de las instituciones facilitadoras se operativicen de manera totalmente diferente. Se requeriría, por lo tanto, recoger otras muestras y hacer investigación colegiada para analizar qué aspectos desde la evaluación y la lingüística aplicada se pudieron haber incorporado y qué relación se sugiere entre la didáctica y la evaluación. Esto enriquecería el debate actual ante un tema tan complejo y sensible como el de la evaluación y brindaría otras percepciones frente a los instrumentos de evaluación y a la misma evaluación certificativa.

## Referencias

Association of Language Testers in Europe [ALTE]. (2005). *Materials for the guidance*

*Revista Electrónica Matices en Lenguas Extranjeras*, número 8. ISSN 2011-1177. Páginas 129-154.  
Universidad Nacional de Colombia - Facultad de Ciencias Humanas - Departamento de Lenguas Extranjeras. Bogotá. <http://revistas.unal.edu.co/index.php/male>

*of test item writers* [en línea]. Disponible en <http://www.alte.org/about/index.php> (Consultado el 27 de abril de 2015).

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford Applied Linguistics. Oxford: Oxford University Press.

Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge Language Assessment Series. Cambridge: Cambridge University Press.

Bachman, L. F., y Palmer A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford Applied Linguistics. Oxford: Oxford University Press.

Bachman, L. F., y Palmer A. S. (2010). *Language assessment in practice*. Oxford Applied Linguistics. Oxford: Oxford University Press.

Bachman, L. F., y Kunnan A. (2005). *Statistical analyses for language assessment workbook and CD ROM*. Cambridge Language Assessment Series. Cambridge: Cambridge University Press.

Bachman, L. F., y Palmer A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests* (Oxford Applied Linguistics). Oxford: Oxford University Press.

Bloomfield, L. (1933). *Language*. Nueva York: Henry Holt.

Bordón, T. (2006). *La evaluación de la lengua en el marco de E/L2: Bases y procedimientos*. Madrid: editorial ARCO / LIBROS, S.L.

Breen, M. P. (1987). Contemporary paradigms in syllabus design, parte 1. *Language Teaching* 20 (2), 81-92. Cambridge University Press. Department of Linguistics and Modern English Language.

Buck, G. (2001). *Assessing listening*. 1ª edición. Cambridge: Cambridge University Press. [Cambridge Language Assessment Series. 1].

Bygate, M. Skehan, P. and Swain, M. (2001). *Researching pedagogic tasks second language learning, teaching and testing*. Applied linguistics and language study. General Editor. C.N. Candlin. London: Pearson Education Limited.

Canale, M., y M. Swain (1980). *Theoretical bases of communicative approaches to second language teaching and testing*. Applied Linguistics, Vol. 1.

Cárdenas Martínez, A. (marzo, 2007). *La fiabilidad de los DELE y los candidatos japoneses: ¿un motive de preocupación?* Cuadernos Canela, Vol. 18, 36-76.

- Carrillo, A. (2002). *La evaluación: un proceso en permanente construcción*. En: Actualidades pedagógicas N° 41-junio de 2002. Universidad de la Salle. 4-5.
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. En: E. Alcón Soler y M. P. Safont Jordà (Eds.), *Intercultural language use and language learning*. Springer: Netherlands.
- Chapelle, C., Grabe, W., y M. Berns (1997). *Communicative language proficiency: Definition and implications for TOEFL 2000*. Princeton, N.J.: Educational Testing Service.
- Chomsky, N. (1957). *Syntactic structures*. Berlin: Walter de Gruyter GmbH. (Reimpresión, 1999, Buenos Aires, Siglo XXI).
- Consejo de Europa. (2002). *Marco Común Europeo de Referencia para Las Lenguas: aprendizaje, enseñanza, evaluación*. Cambridge University Press.
- Crookes, G. (1986). *Task classification: A cross-disciplinary review (Technical Report no 4)*. Center for Second Language Classroom Research, Social Science Research Institute, University of Hawaii at Manoa.
- Da Silva Gomes, H. y Dorcasberro, A. (2005). *Temas sobre la adquisición de una segunda lengua*. Trillas. México.
- Ellis, R. (1998). *The evaluation of communicative tasks*. En: B. Tomlinson, ed. *Materials development in language teaching*. La edición. Cambridge: Cambridge University Press.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Ezeiza, J. (2006). *Analizar y comprender la topografía configuracional de los materiales de enseñanza de lenguas en una perspectiva de síntesis. Aplicación a los manuales para la enseñanza general de E/LE a adultos*. Tesis doctoral. Universidad Antonio de Nebrija. Facultad de lenguas aplicadas y humanidades.
- Figueras, N. (2008). El MCER, más allá de la polémica. *Monográficos marco ELE*, n. 7, 26-35. Disponible en <http://marcoele.com/el-mcer-mas-alla-de-la-polemica/>
- Hughes, A. (2008). *Testing for language teachers* (2a ed.). Cambridge: Cambridge University Press.
- Hymes, D.H., Hornberger, N.H., Canale, M., Widowson, H.G., Cots, J.M., Bachman, L., Spol sky, B. Llobera, M. (1995). *Competencia comunicativa documentos básicos en la enseñanza de lenguas extranjeras*. Edelsa. Madrid.
- Jurado Salinas, M., y otros (2012). *Examen de Español como Lengua Extranjera para*

*el Ámbito Académico (EXELEAA). Lineamientos generales para la elaboración de reactivos y consignas.* Documento de trabajo en construcción facilitado para fines investigativos. Sin información posterior.

- Larsen-Freeman, D. y Long, M. (1994). *Introducción al estudio de la adquisición de segundas lenguas.* Madrid: Gredos.
- Long, Michael H. (1985). A role for instruction in second language acquisition theory. En: S. Grass y C. Madden (Eds), *Input and second language acquisition theory.* Rowley Mass.: Newbury House.
- Llobera, M., Hymes, D.H., Hornberger, N.H., Canale, M., Widowson, H.G., Cots, J.M., Bachman, L., Spolsky, B. (1995-b). *Una perspectiva sobre la competencia comunicativa y la didáctica de lenguas.* En: M. Llobera, coord. 1995. *Competencia comunicativa. Documentos básicos en la enseñanza de lenguas extranjeras.* 1a edición. Madrid: Edelsa.
- Martín Peris, E. (1995). Las tareas en la enseñanza de e/le a adultos. En: Siguan, M. (1995). *La enseñanza de la lengua por tareas. XVIII Seminario sobre lengua y educación.* Barcelona: Editorial HORSORI.
- Mejía, A.M. y otros. (2009). *Los exámenes de evaluación en Colombia.* Instituto Caro y Cuervo - Universidad de los Andes. Investigación. Disponible en: [http://www.lenguasdecolombia.gov.co/sites/lenguasdecolombia.gov.co/files/informe\\_icc\\_andes.pdf](http://www.lenguasdecolombia.gov.co/sites/lenguasdecolombia.gov.co/files/informe_icc_andes.pdf) / (Consultado el 7 de mayo de 2015).
- Milanovic, M. (1998). *Multilingual glossary of language testing terms. Studies in language testing 6.* Cambridge: Cambridge University Press. Prepared by ALTE Members. 206.
- Nation, P. (2001). *Learning vocabulary in another language.* Cambridge Applied Linguistics Series. Cambridge: Cambridge University Press.
- Nunan, D. (1989). *Designing Tasks for the Communicative Classroom.* Cambridge: Cambridge University Press.
- Nunan, D. (2004). *Task-Based Language Teaching. A comprehensively revised edition of designing tasks for the communicative classroom.* 5th printing 2009. CUP. Cambridge Language Teaching Library.
- Parrondo, J. R. (2004). Aspectos éticos de la evaluación: impacto de la actividad evaluadora y códigos deontológicos de los examinadores. *Carabela 55: La evaluación en la enseñanza de Español como Segunda Lengua / Lengua Extranjera,* 31-41. Sociedad General Española de Librería, Madrid.
- Pérez Ruiz, J. (2013). Análisis de la percepción de los exámenes DELE A1 realizada

por estudiantes taiwaneses de ELE. *Revista Nebrija de Lingüística Aplicada*, 7 (2).

Prabhu, N.S. (1987). *Second Language Pedagogy*. Oxford: Oxford University Press.

Prati, S. (2007). *La evaluación en Español como Lengua Extranjera: Elaboración de exámenes*. Buenos Aires: Libros de la Araucaria.

Purpura, J. (2004). *Assessing grammar*. 1st ed. Cambridge: Cambridge University Press. [Cambridge Language Assessment Series].

Richards, J., y T. Rodgers (1986). *Approaches and methods in language teaching: A description and analysis*. Cambridge University Press.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P. (2001). Tasks and language performance assessment. En: M. Bygate, P. Skehan y M. Swain (eds.), *Researching pedagogic tasks: Second language learning, teaching and testing*. Applied Linguistics and Language Study. Pearson.

Skinner, B.F. (1957). *Verbal Behavior*. Prentice Hall.

Siguan, M. (Coord.). (1995). *La enseñanza de las lengua por tareas. XVIII. Seminario sobre "lengua y evaluación"*. Barcelona: Horsori.

Sistema Internacional de Certificación de Español como Lengua Extranjera [SICELE] (s.f.). Portal web institucional. Disponible en <https://sites.google.com/a/sicele.org/sicele/> (Consultado el 1° de junio de 2015).

Willis, J. (1996). *A framework for task-based learning*. Addison Wesley Longman Limited. Longman Handbook for Language Teachers. England.

Yagüe Barredo, A. (2010). Examinando exámenes: Consideraciones acerca de las pruebas de medición en ELE. *Marco ELE Revista de Didáctica Español Lengua Extranjera*, n. 10, 1-53. Disponible en <http://marcoele.com/numeros/> (Consultado el 30 de enero de 2014).

### **Siglas usadas**

ALTE                      Association of Language Testers in Europe  
Asociación de Evaluadores de Lengua en Europa

CELU                      Certificado de Español Lengua y Uso

COE [Model]	Committee of Examiners TOEFL Model
DELE	Diploma de Español como Lengua Extranjera
E/LE	Español como Lengua Extranjera
EALTA	European Association for Language Testing and Assessment Asociación Europea de Evaluación y Pruebas de Lengua
EXELEAA	Examen de español para extranjeros para propósitos académicos
ILTA	International Testing Association Asociación Internacional de Pruebas de Lengua
SICELE	Sistema de Certificación de Español como Lengua Extranjera
SIELE	Servicio Internacional de Evaluación de la Lengua Española
TOEFL	Test of English as a Foreign Language
UNAM	Universidad Nacional Autónoma de México