

Evaluación de los aprendizajes en lenguas extranjeras: hacia prácticas justas y democráticas¹

Clara Inés Arias

Liliana Maturana

María Isabel Restrepo

Universidad de Antioquia
Fundación universitaria Luis Amigó
Medellín, Colombia

La falta de coherencia encontrada entre la evaluación y la promoción en lenguas extranjeras se abordó en esta investigación-acción interinstitucional con la consolidación de un Sistema consensuado de Evaluación. La implementación de dicho Sistema en cursos de inglés para principiantes en tres programas de extensión para adultos, demostró que la variedad de tipos y formas de evaluación, la rigurosidad y la sistematicidad, y el diseño meticuloso de instrumentos, formatos y tareas de evaluación, transforman las prácticas evaluativas en justas y democráticas, y benefician a estudiantes, profesores e instituciones.

Palabras clave: evaluación de los aprendizajes en lengua extranjera, evaluación formativa, evaluación sumativa, democracia, justicia.

Evaluation in foreign language learning: towards fair and democratic practices

The lack of congruence found between evaluation and pass/fail decisions in foreign languages was addressed in this interinstitutional action research project through the consolidation of a collaboratively designed evaluation system. The implementation of the system in English beginner courses of three adult extension programs revealed that a variety of evaluation types and forms; rigor and systematicity; and the meticulous design of evaluation instruments, forms, and tasks made evaluation practices fair and democratic and benefited students, teachers, and institutions.

¹ Este artículo se deriva de la investigación Prácticas evaluativas en la adquisición de Lenguas Extranjeras. Fase II: consolidación de un sistema de evaluación consensuado, realizada por Arias, Areiza, Estrada, Ferreira, Marín, Maturana, Restrepo y Restrepo y los estudiantes Marín y Mejía, financiada por las universidades de Antioquia, Nacional de Colombia - sede Medellín y Fundación Universitaria Luis Amigó y se llevó a cabo de marzo 2007 a septiembre 2009.

Key words: evaluation in foreign language learning, formative evaluation, summative evaluation, democracy, fairness

L'Évaluation des apprentissages en langues étrangères : vers des pratiques justes et démocratiques

Nous avons abordé le manque de cohérence entre l'évaluation et la promotion dans certains cours de langues étrangères lors de cette recherche-action interinstitutionnelle à travers un Système d'Évaluation consensuel. La mise en place de ce Système dans des cours d'anglais appartenant à trois programmes d'éducation non-formelle, pour un public adulte débutant, a démontré que les types et les formes variés ainsi que la conception d'instruments, de formats et de tâches diverses, conçus dans le cadre d'un processus systématique et rigoureux, rend les pratiques d'évaluation plus justes et démocratiques et favorise les étudiants, les professeurs et les institutions.

Mots clés : évaluation des apprentissages en langue étrangère, évaluation formative, évaluation sommative, démocratie, justice

INTRODUCCIÓN

En Colombia, los sistemas educativos y los programas de lenguas extranjeras, apoyados en los proyectos para el fomento de estas lenguas², renuevan y evalúan metodologías y recursos con alguna regularidad; paradójicamente, la evaluación de los aprendizajes no ha recibido igual atención (López & Bernal, 2009). Sin embargo, sobre la evaluación de los aprendizajes en lenguas extranjeras se han realizado estudios de corte cuantitativo (Muñoz & Álvarez, 2009; Muñoz & Álvarez, 2008), así como otros de corte cualitativo (López, 2010; Muñoz, 2009). Nuestro estudio, de corte cualitativo, es una investigación-acción (IA) en la que se concibe la evaluación como una construcción rigurosa y sistemática, democrática y justa, caracterizada por la participación de los actores que la realizan cotidianamente en el aula de clase: los profesores evaluadores y los evaluados. Nuestro equipo de investigación abordó las carencias y las necesidades de evaluación surgidas de una primera investigación³ en la que se encontró que la promoción de los estudiantes

2 Por ejemplo: Plan Nacional de Bilingüismo (www.mineduacion.gov.co), Plan Decenal 2006-2016, Plan Nacional de Desarrollo 2007-2010.

3 Prácticas evaluativas en la adquisición de lenguas extranjeras (Arias, C.I., Frodden, C., Vallejo, M., Usma, J.A., Murillo, G. J. & Molina, J. M 2003) que tuvo como propósito evidenciar las prácticas evaluativas de los profesores de lenguas extranjeras y su relación con la promoción de los estudiantes al siguiente nivel.

no correspondía a los logros alcanzados o requeridos. Entre los factores de esta falta de coherencia están la poca claridad en las directrices de los programas, la falta de consensos entre profesores y administradores, la escasa presencia de planeación y registros formales, las diferencias en el concepto de lengua, la presencia insuficiente de las cualidades de la evaluación (Frodden, Restrepo & Maturana, 2004), y la confusión en los tipos de evaluación (Arias & Maturana, 2005). Estos hallazgos mostraron la necesidad de lineamientos generales para prácticas evaluativas más confiables y congruentes que se materializaron en un Sistema de Evaluación en Lenguas Extranjeras (Arias, Estrada, Areiza & Restrepo, 2009) construido de manera colectiva y democrática, favoreciendo el empoderamiento de los profesores, actores educativos históricamente marginados de muchas decisiones en la educación y en la evaluación (Shohamy, 2001). Para dicha construcción, el equipo de investigación del primer estudio se enriqueció con estudiantes en formación, y profesores-investigadores que implementaron el Sistema entre 2007 y 2008 en cursos iniciales de los programas participantes.

La pregunta de investigación fue: **¿Cómo un Sistema consensuado de Evaluación en lenguas extranjeras genera coherencia entre la competencia comunicativa propuesta para un nivel y el desempeño de los estudiantes al final del mismo?** Las preguntas guía fueron: a) ¿Cuál es el papel de la competencia estratégica en el desempeño lingüístico del estudiante en situación de evaluación?, b) ¿Cuál es el nivel de apropiación del Sistema en términos de conceptos y procedimientos de evaluación por parte del equipo?⁴, c) ¿Cómo contribuyen las cualidades de las pruebas a la coherencia entre la evaluación del desempeño lingüístico de los estudiantes y los logros planteados en los cursos? y d) ¿Cómo evidenciar la rigurosidad de las pruebas alternativas?

Este artículo presenta, en primer lugar, los referentes teóricos; luego, la metodología; posteriormente, los resultados; y finalmente, las conclusiones.

⁴ La reflexión que surge de esta investigación sobre el empoderamiento de los profesores investigadores en sus prácticas evaluativas se encuentra en Arias y Restrepo (2009a & 2009b).

REFERENTES TEÓRICOS

Los referentes teóricos que guiaron este estudio fueron en gran medida contruidos por el equipo en las discusiones y consensos durante la investigación, lo cual se materializó en el diseño conjunto del Sistema de Evaluación en Lenguas Extranjeras (Arias, Estrada, Areiza & Restrepo, 2009). Presentamos a continuación la definición y las características del Sistema seguidas de los conceptos de evaluación allí contenidos.

Sistema de evaluación en lenguas extranjeras

Definido como “un conjunto integrado de partes interdependientes relacionadas con la evaluación del desempeño de los estudiantes de lenguas extranjeras.... Este Sistema es una propuesta articuladora de las prácticas evaluativas... que permite generar consensos entre docentes y administradores” (p. iii) y propiciar coherencia al ofrecer criterios comunes y un mismo lenguaje sobre evaluación. Las características del Sistema son: a) generalidad, pues aunque se diseñó para tres programas de lenguas extranjeras con metodología, objetivos, contenidos, población estudiantil y perfil de los profesores similares, puede ser acogido por otros programas de lenguas extranjeras; b) flexibilidad, porque es adaptable a las especificidades de los programas; c) rigurosidad, ya que sus principios, normas y procedimientos deben seguirse de manera consciente y sistemática; d) continuidad, pues promueve regularidad en los eventos evaluativos y oportuna retroalimentación a los estudiantes (p. 1). De esta manera, el Sistema tiene en cuenta la universalidad de los constructos teóricos y la particularidad del contexto local y nacional.

Habilidad lingüística

La concepción de lengua que el equipo acoge es la de “un sistema socialmente construido cuyos componentes posibilitan al usuario tanto la expresión y la interpretación de significados como la creación de relaciones consigo mismo, con los demás y con el mundo” (Arias et al., 2009, p. 5). La capacidad del usuario en el empleo de la lengua es lo que Bachman y Palmer (1996) denominan habilidad lingüística. Dicha habilidad involucra interacciones múltiples y complejas, por un lado, entre las características de los usuarios de la lengua (conocimiento

lingüístico, esquemas afectivos, conocimiento temático, características personales, competencia estratégica) y, por el otro, entre éstas y aquéllas de la situación en la cual se emplea (participantes y su relación, contexto e intención). Esta habilidad está conformada por el conocimiento lingüístico y la competencia estratégica. El conocimiento lingüístico, denominado competencia comunicativa por el equipo investigador, incluye las áreas del conocimiento organizativo (gramatical y textual) y del pragmático (funcional y sociolingüístico), y se activa gracias a la competencia estratégica. La competencia estratégica, más que un conjunto de componentes metacognitivos (Bachman & Palmer, 1996), es la capacidad para movilizar y activar conocimientos y destrezas necesarios para cumplir con las exigencias de las situaciones comunicativas dentro y fuera del aula (Celce-Murcia, Dörnyei & Thurrell, 1995, en Pawlikowska-Smith, (2002).

Cualidades de las pruebas y principios de las prácticas evaluativas

Para asegurar la calidad, la pertinencia y la rigurosidad de la evaluación de los aprendizajes en lenguas extranjeras en el aula de clase, en esta investigación adaptamos cualidades planteadas por distintos autores, varias de ellas aplicadas a pruebas estandarizadas (*tests*). Por consiguiente, de Bachman y Palmer (1996) se tomaron la *validez de constructo*, la *interactividad* y la *autenticidad* para el diseño de las tareas evaluativas; de los mismos autores y de Brown y Hudson (1998), la *confiabilidad* en las mediciones, y de Bachman y Palmer y de Shohamy (2001), la *viabilidad* con respecto al diseño, la administración y la calificación de las pruebas, así como el *impacto*, relacionado con el uso de los resultados. De Fredriksen y Collins (1989, en Shohamy, 2001) y de Bustos (1997) se adaptó la *validez sistémica*; de Brown (2004), la *validez de contenido*, y de Carroll (1993), la *transparencia*. Finalmente, como principios de las prácticas evaluativas, asumimos de Shohamy (2001) y Bustos (1997), la *justicia*; la *justicia procedimental* de Maiese (2003), y la *democracia* de Shohamy (2001). Las siguientes definiciones fueron utilizadas por el equipo y consignadas en el Sistema.

- *Validez de constructo*: pertinencia de las interpretaciones o inferencias que el evaluador hace de la competencia comunicativa del estudiante a partir de los resultados de las evaluaciones. Esta pertinencia requiere de un constructo

definido en términos funcionales en situaciones comunicativas concretas –con intenciones claramente especificadas–, materializadas en tareas evaluativas que permitan el despliegue del desempeño lingüístico de los estudiantes.

- *Interactividad*: capacidad de la prueba de involucrar al evaluado comprometiendo su habilidad lingüística, conocimiento temático y esquemas afectivos.
- *Autenticidad*: relación entre las características de la tarea evaluativa y las situaciones cotidianas de comunicación.
- *Confiabilidad (Trustworthiness)*: consistencia en los resultados de una prueba administrada en diferentes momentos a un mismo estudiante o grupo en condiciones similares, o valorada por distintos evaluadores.
- *Viabilidad*: relación entre los recursos disponibles (talento humano, espacios, materiales y tiempo) y los necesarios para el diseño, la administración y la calificación de una prueba.
- *Impacto*: repercusión de los resultados de una prueba en el ámbito educativo y personal del estudiante, y en las instituciones.
- *Validez sistémica*: relación de la prueba con el concepto de lengua inscrito en la propuesta pedagógica de una institución.
- *Validez de contenido*: representación de los contenidos y procedimientos del curso en la evaluación.
- *Transparencia*: grado de claridad y precisión en la información de la prueba acerca de procedimientos, propósito, ponderación, calificación, tiempo, entre otros.
- *Justicia*: principio cuya acepción más específica (*Fairness*) se ocupa del bienestar del otro y se da en un contexto determinado; por lo tanto, busca procedimientos adecuados a los participantes, o sea, *justicia procedimental*. Sus bases son: a) la *imparcialidad*, entendida como la objetividad y honestidad en la toma de decisiones a partir de información apropiada; b) la *consistencia*, referida al establecimiento de criterios comunes en términos del constructo y del nivel de paso, c) el *reconocimiento* o la valoración de la voz propia y la del otro como miembros de un colectivo social; y d) la *confianza*, o sea, la certeza del evaluado de ser tratado con consideración de su bienestar y sus necesidades en la evaluación. Estas dos últimas, al tener en cuenta a los involucrados en la evaluación, generan *democracia*.

- *Democracia*: como principio, es un modo de vida en proyección colectiva basado en el respeto a la dignidad humana, a la libertad y a los derechos de todos los miembros de la comunidad. En la evaluación, la *democracia* exige diversos momentos de evaluación, variedad de tipos y formas de evaluación y negociación.

Instrumento, formato, tarea y rúbrica

Para cumplir con las cualidades de las pruebas, todas las evaluaciones (en sus diversos tipos y formas) deben contar con un instrumento de evaluación, definido como el dispositivo que contiene de manera exhaustiva toda la información de la rúbrica. En la evaluación tradicional el desempeño del estudiante ha quedado consignado en el mismo instrumento de evaluación (examen, *quiz* o taller); en tanto que en las pruebas alternativas, especialmente las orales, el desempeño del estudiante no siempre queda registrado de manera directa (video o audio grabación). Por esta carencia tan importante, proponemos entonces que para las pruebas alternativas orales se acompañe el instrumento del formato de evaluación, definido como el dispositivo diseñado por el profesor para tener constancia del desempeño del estudiante (Maturana, Restrepo & Ferreira, 2009). Las tareas, a su vez, son “actividades o situaciones propuestas en un momento de evaluación que plantean al estudiante oportunidades para demostrar su competencia comunicativa” (Arias et al., 2009, p. 15). La definición de rúbrica ofrecida por Bachman y Palmer fue ampliada por el equipo para referirse no sólo a la especificación de cómo se evaluará una prueba (criterios, ponderación, descriptores, escalas y nivel de paso), sino al dispositivo para formalizarla que además incluya la información institucional y el conjunto de indicaciones relacionadas con el constructo, la estructura y las características de la prueba, la forma y el tipo de evaluación, la duración total, y las instrucciones para realizar cada tarea que la componga.

Tipos de evaluación

Los tipos son evaluación sumativa y evaluación formativa. El concepto de evaluación sumativa fue construido teniendo en cuenta lo propuesto por Hadji (1997) como “aquella que se lleva a cabo una vez

culminados determinados momentos de instrucción y aprendizaje, y su propósito es verificar, tanto de manera parcial como definitiva, los logros alcanzados por los estudiantes” (Maturana et al., 2009, p. 24); en esta investigación, la competencia comunicativa es objeto de evaluación sumativa. La evaluación formativa se concibió como “un proceso continuo de acopio de información sobre el alcance del aprendizaje, así como sobre los puntos fuertes y débiles, que el profesor puede introducir en su planificación del curso...” (Instituto Cervantes, 2002, p. 181). De la primera investigación concluimos que estos propósitos se alcanzan cuando, de manera sistemática y rigurosa, se ponen en práctica los tres ejes fundamentales que caracterizan la evaluación formativa: la evaluación continua, la autoevaluación y la realimentación (Arias & Maturana, 2005). La evaluación continua está integrada al proceso de aprendizaje por medio de tareas evaluativas realizadas a lo largo del curso cuyo interés es remediar dificultades, afianzar aprendizajes y preparar al estudiante para cuando requiera desplegar su competencia comunicativa. Se realiza de manera regular durante el proceso de aprendizaje y, no obstante ser de carácter formativo en su implementación, adquiere carácter de sumativa en el momento en que dichas tareas evaluativas se tienen en cuenta en las decisiones de promoción. La autoevaluación, cuyo objeto es la habilidad lingüística, representa la valoración que hace el estudiante de sus conocimientos, destrezas, valores y actitudes mediante un proceso de introspección. Finalmente, la retroalimentación es un diálogo entre profesor y estudiante a partir de la autoevaluación y/o la evaluación continua sobre las visiones de ambos con respecto al aprendizaje; estas pueden ser de carácter cognitivo, metacognitivo y afectivo por lo que la habilidad lingüística y las características del estudiante son su insumo.

La realimentación y la autoevaluación son esenciales ya que a partir de éstas, por un lado, el estudiante toma decisiones informadas sobre las acciones para autorregularse, o sea, mejorar sus fortalezas y superar sus dificultades y, por el otro, el profesor se regula con acciones pedagógicas para superar dificultades y reforzar éxitos en el aprendizaje (Jorba & Casellas, 1997; Hadji, 1997). Shohamy (2001) se refiere a todo este proceso como *assessment for learning*. La investigación anterior indicó que estas regulaciones, así como la complementariedad de la evaluación sumativa y la formativa (Estrada & Vallejo, 2007), deben darse a lo largo

del proceso educativo y basarse tanto en la habilidad lingüística como en las características personales del estudiante.

Formas de evaluación

Según la intención con relación a la respuesta esperada (Brown & Hudson, 1998) y al grado de control sobre ésta, la evaluación de los aprendizajes adquiere diversas formas, tales como la tradicional y la alternativa. En un continuo entre mayor y menor control de la respuesta esperada, a mayor control, la prueba tenderá a ser más tradicional. Esto se da en las pruebas de respuesta seleccionada (falso y verdadero, pareamiento y selección) en las cuales el estudiante no tiene que usar el lenguaje de manera creativa. Las pruebas que requieren del uso del lenguaje para construir la respuesta tienden a ser más tradicionales cuando se ejerce mayor control sobre ella (tareas de complementación, transformación y respuesta corta) o más alternativas, como en el caso de las pruebas de desempeño donde el control es mínimo y el estudiante se enfrenta con tareas similares a las de la vida real (ensayos, juegos de roles, solución de problemas). Finalmente, están las tareas de respuesta personal en las que los estudiantes deben producir, o sea que hay un grado mínimo de control, cuando lo hay. Como son de proceso, el foco es la habilidad lingüística; la conferencia (entre profesor y estudiante), el portafolio y la autoevaluación son ejemplos de esta categoría; sin embargo, el equipo propone que se utilicen con fines formativos y no de promoción.

METODOLOGÍA

La metodología empleada fue la investigación-acción colaborativa y participativa. La colaboración se dio gracias a la vinculación de profesores, investigadores y estudiantes en formación investigativa. La participación y el empoderamiento de los miembros del equipo se generaron en todas las fases de la investigación debido a los tres principios que guiaron nuestro trabajo: la reflexividad, la crítica dialéctica y la colaboración (Arias & Restrepo, 2009a y 2009b).

Contexto⁵

El Sistema fue implementado por cinco docentes en tres programas de enseñanza de lenguas extranjeras de dos universidades públicas de la ciudad de Medellín que responden a políticas de internacionalización. En uno, los cursos no son parte del pregrado, sin embargo los estudiantes deben pasar un examen de comprensión de lectura para obtener su título. En otro, se requiere aprobar mínimo cuatro semestres de una lengua extranjera para recibir el diploma. El tercero ofrece cursos de extensión a personas de 16 años en adelante de la comunidad externa. En los tres programas se evalúa con aprobó o reprobó. Todos los cursos son de corta duración, razón por la cual los grupos fueron diferentes, situación que contribuyó a la característica de generalidad del Sistema consolidado.

Participantes

Cuatro investigadoras del primer estudio (investigadoras iniciales), cinco profesores (nuevos investigadores)⁶ y dos estudiantes conformaron el equipo. Las investigadoras iniciales estuvieron a cargo de facilitar el trabajo colaborativo entre los participantes (Chisholm & Elden, 1993), acompañar la implementación del Sistema apoyando el proceso de diseño de tareas evaluativas, dinamizar la reflexión sobre el desarrollo profesional relacionado con la evaluación de los aprendizajes y la IA, actuar como amigas críticas (Burns, 1999) y cultivar en los nuevos investigadores destrezas de recolección y análisis de datos. Los cinco profesores –encargados de poner en práctica el Sistema en cursos de inglés del nivel 1 o 2– son licenciados en idiomas o profesionales en áreas afines y tienen experiencia en la enseñanza de lenguas extranjeras a diferentes poblaciones etarias, a nivel de pregrado y postgrado, y su contratación es de cátedra o por prestación de servicios. Entre sus responsabilidades investigativas estaban: diseñar tareas evaluativas acordes con sus cursos y con el Sistema, fomentar en sus estudiantes la autoevaluación, mantener registros organizados de evaluación, evaluar periódica y sistemáticamente sus prácticas evaluativas y el Sistema, y reflexionar sobre su desarrollo en evaluación y en IA. Finalmente, los

⁵ Esta era la situación de los programas en el momento del estudio; los cambios que se hayan generado no están contemplados en este artículo.

⁶ Algunos profesores ingresaron en distintos momentos de la investigación, por esta razón no se contó con los datos de los cinco profesores en cada uno de los momentos de recolección de datos.

estudiantes en formación investigativa participaron en la observación de un evento evaluativo y en el análisis de datos. (Ver Tabla 1.)

Tabla 1. Participantes de la investigación y sus funciones

Participantes	Funciones
4 investigadoras iniciales	-facilitar el trabajo colaborativo -acompañar la implementación del Sistema -apoyar el diseño de tareas evaluativas -dinamizar la reflexión sobre el desarrollo profesional -actuar como amigas críticas -cultivar destrezas investigativas en los participantes
5 nuevos investigadores o profesores-investigadores	-poner en práctica el Sistema -diseñar tareas evaluativas acordes con sus cursos y el Sistema -fomentar la autoevaluación en sus estudiantes -mantener registros organizados de evaluación -evaluar sus prácticas evaluativas y el Sistema -reflexionar sobre su desarrollo en evaluación y en IA
2 estudiantes en formación investigativa	-observar un evento evaluativo -contribuir al análisis de datos

Ciclos de la investigación

La consolidación del Sistema de Evaluación marcó el derrotero de esta investigación-acción (Burns, 1999). Por su carácter colaborativo y participativo, la conformación de la comunidad investigativa se inició en un seminario-taller⁷; por su carácter formativo, el equipo definió reuniones y talleres⁸ como dinámica de trabajo. Siguiendo la metodología de la investigación-acción, durante el proyecto se implementaron dos acciones. La primera consistió en una prueba alternativa oral y su análisis e interpretación condujo a la planeación e implementación de una evaluación cooperada como segunda acción para validar hallazgos y dar continuidad al ciclo de la investigación-acción. La Tabla 2 resume las técnicas, los tipos y el número de instrumentos de recolección y los tipos de análisis.

⁷ Este seminario-taller perseguía tres propósitos: validar hallazgos de la primera investigación, actualizar a los profesores en evaluación e invitarlos a participar en el segundo proyecto.

⁸ Las reuniones se destinaron a la logística; los talleres, al diseño y análisis de instrumentos de evaluación y al análisis de datos.

Tabla 2: Técnicas, tipos y número de instrumentos de recolección y tipos de análisis

Acción no. 1: Prueba alternativa oral		
Técnicas	Instrumentos	Análisis
Observación no participante	Formatos guía (2)	Deductivo e inductivo
Auto observación	Escritos (3)	
Instrumentos (3) y formatos (3) de evaluación		
Análisis de documentos	Actas (4)	Inductivo y deductivo
Descripción individual	Reflexiones escritas de la competencia estratégica (3)	
Acción no. 2: Evaluación-observación cooperada		
Técnicas	Instrumentos	Análisis
Observación-evaluación cooperada	Formatos de evaluación de la competencia comunicativa (8) Notas de campo sobre la competencia estratégica (8)	Deductivo e inductivo
Descripción individual	Escritos acerca del seguimiento (4)	Deductivo

La primera acción consistió en la aplicación de una evaluación alternativa oral⁹ cuyos objetivos fueron recoger información a) sobre la competencia comunicativa de los estudiantes y las cualidades de las pruebas, y b) sobre la percepción de los profesores-investigadores acerca de la competencia estratégica. Las técnicas y los instrumentos utilizados para el primer objetivo fueron: observación no participante y auto-observación, instrumentos y formatos de evaluación, y actas de talleres. La observación no participante se registró en formatos-guía y la auto-observación de los profesores, en escritos; ambas técnicas se usaron con el fin de identificar la competencia comunicativa de los estudiantes. La auto-observación permitió encontrar semejanzas y divergencias con la observación no participante, pues el tener estas dos perspectivas imprimió un grado de confiabilidad a los datos. Adicionalmente, los instrumentos y formatos de evaluación diseñados por cada profesor-investigador teniendo en cuenta las características (rigurosidad, generalidad, continuidad, flexibilidad) y las cualidades propuestas en el Sistema tuvieron como propósito determinar la presencia de ellas y su relación con la competencia comunicativa evaluada en la prueba.

⁹ Dos profesores-investigadores emplearon una evaluación formativa y el otro, una sumativa.

Finalmente, las actas de talleres con las discusiones relacionadas con el diseño y el análisis de los instrumentos y formatos de evaluación empleados, y con el análisis de los informes de observación y auto-observación sirvieron para posterior triangulación. Para el segundo objetivo los profesores-investigadores escribieron sus reflexiones acerca de los aspectos de la competencia estratégica que consideraban incidían en el desempeño lingüístico de los estudiantes al ser evaluados.

La segunda acción fue una evaluación-observación cooperada¹⁰ concebida con dos objetivos: a) recolectar información sobre el uso y la efectividad de la competencia estratégica por parte de los estudiantes en situación de examen y b) analizar factores que inciden en la *confiabilidad* de las pruebas. El primer objetivo se debió a que en la primera acción se estudió la incidencia de dicha competencia en el desempeño solamente desde la percepción de los profesores-investigadores. El segundo objetivo surgió de: los resultados de una primera investigación que revelaron poca inclusión de la *confiabilidad* en pruebas alternativas y la necesidad de contar con instrumentos y formatos para aumentarla en estas pruebas, la alta recurrencia de esta cualidad en los resultados de la primera acción y el interés del equipo por buscar niveles de coherencia en la calificación por dos evaluadores de la competencia comunicativa en un mismo evento. En esta acción se dio especial énfasis a dos elementos desde los cuales se puede inferir el mayor o menor grado de *confiabilidad* en las pruebas: el diseño de un instrumento y de un formato, y la presencia de un evaluador externo.

Durante la evaluación-observación cooperada tanto el profesor del grupo como el evaluador cooperador consignaron información que posteriormente fue utilizada para calificar la competencia comunicativa de los evaluados e igualmente como datos para la investigación. Los datos fueron consignados en dos tipos de instrumentos con un propósito cada uno. Por un lado, un formato de evaluación que sirvió para recolectar datos relacionados con el desempeño lingüístico y su instrumento correspondiente para establecer la presencia o no de las cualidades en el diseño de las pruebas y su posterior influencia en dicho desempeño. Por otro lado, las notas de campo sirvieron para registrar el tipo de estrategia utilizada y hacer una corta descripción de ella.

¹⁰ Consistió en una evaluación sumativa alternativa de respuesta construida, a saber, un juego de roles.

En último lugar, tras el análisis colaborativo de los datos de la prueba cooperada, el equipo determinó incluir una fuente más de datos para poder hacer inferencias sobre la coherencia entre los resultados del proceso evaluativo y los resultados de esta prueba sumativa. Con este fin, cada uno de los cuatro profesores-investigadores escribió una descripción detallada sobre las formas de evaluación (tradicionales o alternativas) utilizadas en la evaluación continua o seguimiento, así como un análisis descriptivo de los instrumentos y formatos usados por cada uno durante el proceso evaluativo.

El análisis de los datos fue de tres tipos: a) deductivo e inductivo, b) inductivo y deductivo y c) deductivo. En el primero se partió de categorías preestablecidas en el Sistema (primera y segunda acción) y en Pawlikowska-Smith (2002) (segunda acción) y se complementó con categorías emergentes. En el segundo, se inició con las categorías emergentes y se enriqueció con las preestablecidas teniendo como referente las categorías para la competencia estratégica presentadas en Pawlikowska-Smith (marzo 2002). En el tipo de análisis deductivo se tuvieron en cuenta solamente las categorías previamente establecidas en el Sistema.

En términos de validación, la triangulación de técnicas e instrumentos así como la triangulación entre investigadores permitió que las interpretaciones fueran fiables y que la metodología de la investigación revistiera rigurosidad y validez (Burns, 1999).

RESULTADOS Y DISCUSIÓN

En esta sección se presentan los resultados de la triangulación de los datos obtenidos a partir de los instrumentos de recolección (ver tabla 2) que dan respuesta a la pregunta de investigación y a aquellas que direccionaron el estudio. Inicialmente, se discute el papel de la competencia estratégica y de algunos factores personales en el desempeño lingüístico del estudiante en situación de evaluación. A continuación se abordan el nivel de apropiación del Sistema por parte del equipo, la contribución de las cualidades a la coherencia entre la evaluación del desempeño lingüístico de los estudiantes y los logros planteados en los cursos, y la rigurosidad de las pruebas alternativas. Esto se hace a partir de la discusión de la competencia comunicativa y sus constructos; de

la importancia del diseño de las tareas evaluativas en lo que respecta a propósitos, cualidades, formatos e instrumentos y de la manera como la diversidad de prácticas y la interrelación de las cualidades favorecen las prácticas evaluativas justas y democráticas.

Con respecto a la competencia estratégica, los datos mostraron cuatro circunstancias: a) el uso adecuado de estrategias para favorecer la comunicación, b) la débil relación entre el número de estrategias empleadas y el buen desempeño en la prueba, c) la participación de ciertos factores que obstaculizaron la culminación de las tareas, y d) la relación entre una forma alternativa y determinadas estrategias de interacción.

Para la primera circunstancia, en algunos casos los evaluados tuvieron un buen desempeño en la prueba (juego de roles) al mantener la comunicación haciendo uso adecuado de estrategias de interacción (e.g., negociación de significado y solicitud de aclaración y ayuda en la lengua extranjera o en la materna) y de compensación (e.g., traducción literal, cambio de código lingüístico, empleo de gestos) (Pawlikowska-Smith, marzo 2002).

En la segunda, hubo evaluados que, aunque emplearon un gran número de estrategias, no fueron exitosos en la prueba al no lograr mantener la comunicación; es el caso de los estudiantes Odiseo, Telémaco y Perséfone, quienes usaron entre cinco y nueve veces diferentes estrategias durante la prueba y no lograron pasarla. Entre las estrategias que estos estudiantes emplearon se observaron algunas de las mencionadas líneas arriba; sin embargo, también se observaron estrategias de evasión como el abandono del mensaje, la evasiva de la respuesta o el reemplazo de ésta por otra (Evaluación cooperada, profesores 1, 3, 4). De estos ejemplos se concluyó que la importancia de la competencia estratégica radica en el uso apropiado e intencional que se haga de las estrategias para solucionar un problema de comunicación o impedir que ésta se pierda, y no en la cantidad de estrategias empleadas.

En la tercera, algunos casos mostraron que no fue el uso, el mal uso o el no uso de las estrategias lo que evitó llevar a cabo la tarea comunicativa, sino la presencia de factores como el conocimiento temático, los esquemas afectivos y las características personales. Debido a la influencia de algunos de estos factores, la estrategia categorizada como de evasión no fue usada para evitar un tema proponiendo otro, sino para abandonar la tarea comunicativa, obstruyendo así el flujo lingüístico de los evaluados. La

siguiente evidencia ejemplifica esta circunstancia: “*el tema de evaluación [la familia] parece no ser un buen referente afectivo para ella; ... no vive con su familia desde que era adolescente, y desconoce... acerca de la rutina de sus vidas, no muestra afecto por ellos...*” (Análisis colaborativo de evaluación cooperativa, profesor 3, septiembre 2008).

El caso anterior ejemplifica la cuarta circunstancia en la que una prueba alternativa (juego de roles) puede perder *interactividad* por factores externos a la prueba, pero inherentes al evaluado, que lo llevan a no usar la competencia estratégica para solucionar el problema de comunicación, sino para abstraerse y abstenerse de participar en ella. Se concluyó que este tipo de factores, además de poder afectar el desempeño lingüístico de los estudiantes, atraviesa la competencia estratégica que les ayudaría a llevar a feliz término la tarea comunicativa. En las reflexiones, los profesores-investigadores se refirieron a estos factores (‘el temor’, ‘la ansiedad’, ‘la capacidad de relacionarse con los demás’, y las ‘estrategias de comunicación como parafrasear, usar sinónimos y completar significados con gestos’), y consideraron que no debían ser objeto de calificación cuando se aplica una prueba, mas sí de evaluación susceptible de refuerzo, mejoramiento o cambio. En este sentido, el lugar que ocupan en la evaluación la competencia estratégica y los factores mencionados se sitúa en la evaluación formativa, especialmente en la realimentación y en la autoevaluación. Finalmente, se puede aseverar que el uso de las estrategias depende de la situación comunicativa en la cual se enmarca la tarea, de la intención de ésta, así como de las características personales de los evaluados. Los datos mostraron que tareas alternativas como el juego de roles tienden a favorecer el empleo de estrategias de interacción que se darán siempre y cuando la tarea tenga un buen grado de *interactividad* y se minimice la incidencia negativa de factores como los mencionados.

Con respecto a la competencia comunicativa, se encontraron definiciones atinadas del constructo en términos funcionales, así como otras con problemas. Ejemplo de la primera es “*dar y solicitar información personal acerca de sí mismo y otros*” (Descripción del seguimiento, profesor 4, junio 2008; Instrumento evaluativo, profesor 1, mayo 2008); evidencia de la segunda es “*el constructo... resultó un poco confuso, pues no todo se explicitó en términos de funciones comunicativas*” (Auto observación, profesor 3, marzo 2007). Aun cuando hubo casos en que los criterios se plantearon desde el conocimiento pragmático, “*Introducing myself to an*

e-pal" (Instrumento evaluativo, profesor 5, noviembre 2008) y "*greetings*" (Auto observaciones y formatos de evaluación, 2007), también hubo otros en que se propusieron desde el conocimiento organizativo, "*verb to be in affirmative and negative sentences*", "*... dificultades en el uso de los plurales*". Cuando no se produjo la respuesta esperada, el evaluador nuevamente recurrió al área organizativa, por ejemplo, la formulación de oraciones con verbos diferentes al verbo ser o la formulación de preguntas (Instrumento evaluativo y auto observación, profesor 2, marzo 2007). Algunos profesores, entonces, operacionalizaron el desempeño del estudiante a partir del conocimiento organizativo, eclipsando las otras áreas de la competencia comunicativa y afectando la *validez de constructo*.

La competencia comunicativa se presentó, pues, como *leitmotiv* en la evaluación de los aprendizajes a partir de tareas comunicativas que posibilitaron a los estudiantes el desplegarla de manera contextualizada e intencionada. Las prácticas evaluativas de los profesores-investigadores se tornaron cada vez más democráticas y justas al contar con una variedad de tipos y formas de evaluación, con instrumentos y formatos, y con un diseño cuidadoso en términos de las cualidades de las pruebas y de los momentos de su desarrollo.

En lo concerniente al diseño, reconocimos tres momentos: el diseño de la prueba, su implementación y su calificación¹¹. Si bien no todas las cualidades pudieron observarse en todos los momentos, sí pudieron predecirse en gran medida desde el diseño, momento fundamental cuando se definen el tipo y la forma de evaluación, el constructo por evaluar y se planea la tarea. Para el segundo momento se necesitó una rúbrica clara y completa en los instrumentos y formatos para el registro del desempeño del estudiante. En el tercero, se confirmó cómo el desempeño del evaluado en la tarea evidenciaba el nivel de logro alcanzado con fines de promoción.

La diversidad de prácticas se vio en los tipos y formas de evaluación favorecidos por el equipo. En lo referente a los tipos, los datos mostraron tanto la evaluación formativa y la sumativa como su complementariedad

11 Shohamy (1998) se refiere a nueve etapas: determinación del propósito, definición de conocimiento lingüístico, selección de procedimientos, diseño de ítems y tareas, administración, determinación de los criterios de calidad de la lengua, evaluación de los procedimientos de evaluación, interpretación de resultados, calificación e informe. Nuestro equipo incluyó las primeras cuatro y la sexta en el primer momento; no hay diferencia en la administración y las etapas ocho y nueve corresponden al que denominamos el tercer momento. La séptima etapa, la evaluación de la evaluación, se dio en cada momento.

al proveer información extensa y suficiente para los propósitos de los cursos. Además de confirmar el potencial de la evaluación formativa para fines sumativos, cual es el caso de la evaluación continua, se descubrió una zona de intersección que le imprimía a la evaluación sumativa un carácter más amplio y tolerante con respecto al uso de los resultados. Aparte de ser usada para tomar decisiones (Arias et al., 2009; Maturana et al., 2009; Hadji, 1997), en las evaluaciones parciales la evaluación sumativa era susceptible de adquirir características de la evaluación formativa. Precisamente, fue utilizada para retroalimentación y autoevaluación, lo cual permitió que estudiante y profesor propusieran acciones remediales con seguimiento: *“La realimentación... les ayudó a ser más conscientes del uso de la estructura de la oración. En... la enseñanza, me llevó... a regular... metodología, contenidos a reforzar y... selección de temas.”* (Descripción del seguimiento, profesor 3, abril 2009). *“Las cuatro pruebas tienen un impacto en la promoción y... la posibilidad que tuvieron [en] las dos primeras de remediar,... tuvo un impacto en los estudiantes... en algunos los puso a moverse hacia la autorregulación...”* (Descripción del seguimiento, profesor 1, junio 2008). De esta manera, estas instancias de retroalimentación y autoevaluación dadas a partir de una evaluación de propósito sumativo, difieren de las mismas que se pueden dar en las evaluaciones sumativas finales en las cuales no hay oportunidad para proponer y observar acciones remediales.

Con relación a las formas de evaluación, la investigación anterior mostró que aunque los profesores favorecieron las pruebas tradicionales también administraron múltiples pruebas alternativas que “al no contar con criterios, los resultados evidenciaban criterios... tan disímiles como el uso del idioma en las actividades de clase o en las entrevistas, además de variables afectivas como la motivación, el compromiso y el esfuerzo” (Frodden et al., 2004, p. 189). En esta segunda investigación las pruebas alternativas orales contaron con mayor rigurosidad y *confiabilidad* que las tradicionales. Algunos ejemplos de formas alternativas de evaluación en las pruebas cooperadas estudiadas fueron: vender/comprar una casa o apartamento (profesor 5, noviembre 2008) y hacer una encuesta demográfica como agentes del gobierno solicitando información personal y de la familia (profesor 4, junio 2008). Debido a que dichas pruebas acusaron un alto grado de *interactividad*, se dedujo que para favorecer de manera deliberada la *interactividad*, la forma de evaluación alternativa era la más apropiada. Sin embargo, no fue la prueba alternativa en sí la

que tenía *interactividad*; fue la relación que se pretendía establecer y que de hecho se estableció entre el estudiante y la prueba lo que finalmente determinó el grado de *interactividad*. El conocimiento que los profesores tenían de sus estudiantes en lo referido al tema por evaluar, los esquemas afectivos y sus características personales (Bachman & Palmer, 1996) y la selección de una prueba alternativa devinieron elementos fundamentales para diseñar la prueba y favorecer la *interactividad*. Se pudo inferir con anterioridad el grado de *interactividad* de la prueba; pese a ello, por ser cualidad relacional, sólo se pudo verificar su grado al momento en que se estableció la relación entre el estudiante y la prueba.

Varias evaluaciones alternativas revelaron cierta *autenticidad* en las tareas propuestas al ser posibles en la vida real y, adicionalmente, ser de respuesta construida o personal como “*escribir un e-mail a los padres para contarles acerca de las cosas que hace o no hace bien la persona que han dejado encargada de la casa*” (Descripción del seguimiento, profesor 4, junio 2008) o “*describir la casa o apartamento nuevo a un amigo por e-mail*” (Descripción del seguimiento, profesor 5, noviembre 2008). Sin embargo, tampoco fue garantía que de suyo una prueba alternativa fuese auténtica pues se dio el caso de un juego de roles que no era auténtico debido a que la tarea no daba cuenta de su factibilidad en la vida real. Más exactamente, en una evaluación parcial los estudiantes se entrevistaban repetidamente para obtener información personal, luego entrevistaban a otro acerca de dicha información y finalmente escribían una composición sobre la tercera persona acerca de quien se había recibido información (Descripción del seguimiento, profesor 3, abril 2009).

Finalmente, las cualidades de las pruebas demostraron estar explícitamente relacionadas con la coherencia investigada. Primero, la *validez de contenido* mostró un alto grado de recurrencia y claridad. El conocimiento y la comprensión que los profesores tenían del programa y de los temas contribuyeron sustancialmente a que esta cualidad hubiese estado presente en las pruebas diseñadas, lo cual garantizó coherencia entre lo enseñado y lo evaluado. Aunado a lo anterior, y aunque no fue directa ni deliberadamente expresado por los participantes en el diseño de las pruebas ni al describir su proceso evaluativo, se encontró en la mayoría de los casos coherencia entre las pruebas y la propuesta pedagógica institucional. Lo anterior demuestra que, al igual que en la *validez de contenido*, los profesores-investigadores comprendían el concepto

de lengua planteado por el Sistema –coherente con el concepto planteado por las instituciones– lo cual contribuyó a la inclusión de la *validez sistémica* en el diseño de las pruebas. La *validez de constructo*, por su parte, se explicitó en pocas pruebas; sin embargo, en las descripciones hechas por los profesores se daba por sentado que esta *validez* estaba inscrita en la tarea comunicativa. Para la definición del constructo por evaluar y el planteamiento de la tarea, los profesores-investigadores se basaron en la *validez sistémica* y en la de *contenido*. No obstante, ello no siempre conllevó a que se diera automáticamente la *validez de constructo*, como ocurrió con el *e-mail* para contarles a los padres sobre las actividades de la persona a cargo de la casa, pues quedó clara la tarea, mas no la función comunicativa o la intención en la cual podría inscribirse, a saber, quejarse, sugerir, describir. Llamó la atención que tanto la *validez de contenido* como la *sistémica* fuesen claras para los profesores-investigadores, pero que algunos asumiesen que la *validez de constructo* se infiriera de suyo de la tarea comunicativa. La interpretación con respecto a los tres tipos de validez demostró que si bien son tres conceptos definidos separadamente, sólo adquieren significado y trascendencia en evaluación cuando se ve la interrelación que se establece entre ellos.

La *confiabilidad* que desde la primera investigación surgió como una de las cualidades con mayor recurrencia en las pruebas tradicionales mas no en las pruebas alternativas estudiadas arrojó esta vez hallazgos importantes en dos aspectos: a) la *confiabilidad* entre evaluadores y b) las relaciones entre *confiabilidad*, *transparencia* y *viabilidad*. La *confiabilidad* entre evaluadores se dio gracias a la existencia de instrumentos y formatos con una rúbrica completa, clara y explícita en términos de criterios, instrucciones y constructo, entre otros (Instrumentos de evaluación cooperada, profesores 1, 3, 4 y 5). En cada una de las cuatro pruebas, los dos evaluadores dieron calificaciones similares y por ello se asevera que se correlacionaron en el sentido de que se ubicaron dentro del mismo rango/nivel y en la mayoría de los casos coincidieron en la decisión de dar nota aprobatoria o reprobatoria. Lo anterior se evidencia en “...hay un alto grado de proximidad en los resultados de los estudiantes... el formato bien diseñado... garantiza la confiabilidad. La interpretación de los resultados... fue consistente entre los dos evaluadores...” (Análisis colaborativo de evaluación cooperada, profesor 5). Sin embargo, hubo pequeñas pero significativas variaciones en la calificación, como en los

casos de Penélope y Helena (del profesor 4) y de Perseo y Jasón (del profesor 1). A las dos primeras, el evaluador cooperador otorgó 3.4 y 4.3 y el profesor titular, 4.0 y 4.6, respectivamente. En el segundo caso, Perseo obtuvo 3.0 del evaluador cooperador y 4.3 del profesor titular, en tanto que Jasón obtuvo 3.6 y 4.2, respectivamente. En los cuatro casos la nota más alta fue la asignada por el profesor titular y la más baja, por el evaluador cooperador. Estas diferencias pueden explicarse de dos maneras: por los roles que juega el profesor titular durante la ejecución de la evaluación y por el mayor conocimiento que él tiene de los evaluados. En cuanto a los roles, en la mayoría de las pruebas cooperadas, el profesor titular actuó simultáneamente de evaluador, observador y profesor y, en un caso, también participó como interlocutor en la tarea evaluativa. Esta diversidad de roles afectó el registro del desempeño de los estudiantes. Así mismo, el mayor conocimiento de sus estudiantes lo llevaron a ser más flexible que el evaluador cooperador. Lo anterior demuestra que en las pruebas alternativas de respuesta construida y de respuesta personal siempre habrá un grado de subjetividad por parte del profesor titular. Sin embargo, el tener instrumentos y formatos con las características explicadas arriba produciría, en otra situación similar de prueba, resultados semejantes aunque no se cuente con la presencia de otro evaluador. En contextos educativos como los nuestros en los que se hace difícil, si no imposible, llevar a cabo evaluaciones cooperadas, el contar con instrumentos y formatos con la rúbrica completa, explícita y clara indudablemente aumentará el grado de *confiabilidad* de la prueba.

En cuanto a las relaciones entre *confiabilidad*, *transparencia* y *viabilidad*, los instrumentos y los formatos devinieron elementos clave para fortalecer la rigurosidad y la sistematicidad¹² en la evaluación de la competencia comunicativa. Adicionalmente, la inclusión consciente e intencionada de una rúbrica completa en tales instrumentos y formatos imprimió a éstos un alto grado de *transparencia* que a su vez favoreció la *viabilidad* y la rigurosidad que confluyeron en la *confiabilidad*. En los documentos analizados se observó que la presencia del constructo, la claridad en su definición y la presencia de criterios de evaluación claros y suficientes fueron factores que afectaron tanto la *transparencia* como la *viabilidad*. Para los profesores-investigadores ‘la información clara’ y ‘completa’, ‘los valores’ y las ‘instrucciones’ fueron elementos fundamentales en

¹² Sistematicidad entendida como la observancia de la continuidad, característica del Sistema.

la *transparencia* de las pruebas. Los elementos señalados, parte integral de nuestra definición de rúbrica, mostraron que ésta, además de ser un dispositivo de formalización de las tareas evaluativas, también hizo posible tres de las cualidades de las pruebas: la *confiabilidad*, la *transparencia* y la *viabilidad*.

Con respecto a la rigurosidad, puede decirse que entre ésta y la *confiabilidad*, la *transparencia* y la *viabilidad*, se estableció una relación explícita y multidireccional que se observó en los datos. Por ejemplo, los profesores-investigadores establecieron como esenciales la presencia de 'instrumentos', 'formato', 'rúbrica', 'instrucciones' y 'la descripción clara de la tarea' y la 'información previa' y 'oral' sobre la prueba como elementos coadyuvantes para la *confiabilidad* y la *transparencia*. Se concluye que si las prácticas evaluativas buscan ser confiables, serán transparentes, tendrán un alto grado de *viabilidad* y serán rigurosas.

Esta relación entre *confiabilidad*, *transparencia*, *viabilidad* y rigurosidad llevó al equipo a aclarar que la *transparencia* y la *viabilidad* no eran las que posibilitaban la *confiabilidad*, sino que es esta última la que revierte en *transparencia* y *viabilidad*. Así pues, la *confiabilidad*, la *viabilidad*, la *transparencia* y la rigurosidad confluyen en los instrumentos, los formatos y en la rúbrica.

En definitiva, cuando sistemáticamente las prácticas evaluativas cuentan con instrumento, formato y rúbrica que favorecen la *transparencia* y la *confiabilidad*, se puede decir que son prácticas rigurosas; cuando sistemáticamente favorecen tipos y formas de evaluación que permitan al estudiante desplegar la competencia comunicativa, se puede decir que hay diversidad de prácticas, y cuando éstas sistemáticamente acusan un diseño cuidadoso y consciente en términos de la tarea, de las cualidades y de la rigurosidad, se está hablando de procedimientos articulados que favorecen la justicia procedimental (Maiese, 2003) conduciendo a unas prácticas justas y democráticas.

Es importante anotar que la Justicia y la Democracia inicialmente fueron consideradas cualidades de una prueba; no obstante, durante el proceso de análisis e interpretación, el equipo encontró que por ser más abstractas que las demás cualidades, eran menos operacionalizables en una prueba; por ello se concluyó que eran principios de las prácticas evaluativas. No es pues una prueba la que acusa justicia y democracia, sino las prácticas evaluativas en su conjunto las que son justas y democráticas.

En términos concretos y observables, no obstante los datos no ofrecer mención explícita y directa de la Justicia y de la Democracia, hay instancias en las cuales se pueden inferir a partir de las cuatro bases de la justicia procedimental. Para empezar, la *imparcialidad* puede verse reflejada cuando el profesor 5 plantea: *“el recibir instrucciones... orales... por escrito y... en los dos idiomas y tener una rúbrica con información suficiente para el desempeño..., elevó el grado de transparencia... y... favoreció su desempeño.”* (Descripción del seguimiento, noviembre 2008). Así mismo, la *consistencia* se infiere en: *“las tareas evaluativas alternativas permitieron que la respuesta producida... correspondiera al constructo..., [lo que] favoreció... la interpretación de la producción de los estudiantes. Adicionalmente, había correspondencia entre la respuesta esperada y los criterios de evaluación consensuados con los estudiantes....”* (Evaluación cooperada, profesor 5, noviembre 2008). El *reconocimiento*, por su parte, se evidencia cuando el profesor 3 expresa *“...[los estudiantes] han manifestado sumo interés al verse involucrados como individuos propositivos, quienes pueden entrar a negociar con el docente las condiciones y metodologías de evaluación”* (Reflexión sobre competencia estratégica, marzo 2007). Finalmente, la *confianza* se evidencia en: *“Los estudiantes conocían los contenidos sobre los cuales se les evaluaría, los criterios, la forma..., y el valor porcentual que la prueba representaba en el proceso evaluativo....”* (Análisis colaborativo de evaluación cooperada, profesor 3, abril 2009). Además del *reconocimiento* y la *confianza* que favorecen la Democracia, en las prácticas evaluativas de los profesores-investigadores, ésta se evidenció al: a) acusar interés en que las pruebas tuviesen *validez de contenido*, *“El final lo prepararé ..., pensando en ese grupo ... y en pruebas que permitieran que los estudiantes exhibieran todo su potencial y que... recogieran todo lo visto”* (Descripción del seguimiento, profesor 1, junio 2008); b) dar a los estudiantes la posibilidad de participar en el diseño de las pruebas, *“... los estudiantes tomaron parte en la definición de criterios a evaluar, el baremo... y el nivel de paso”* (Descripción del seguimiento, profesor 5, noviembre 2008); y c) tener diversidad en cuanto a formas y momentos de evaluación, *“de las tres tareas evaluativas sumativas realizadas en el seguimiento dos de ellas son alternativas y la tercera combina las dos formas de evaluación, tradicional y alternativa”* (Descripción del seguimiento, profesor 5, noviembre 2008).

Para finalizar, en términos de la coherencia entre la evaluación de la competencia comunicativa del estudiante y la promoción de éste, se encontró que existen tres niveles de coherencia: conceptual, práctica y

procedimental, que a su vez interactúan entre sí. La coherencia conceptual se evidenció en el discurso académico de los profesores participantes al construir y apropiarse de la terminología planteada en el sistema demostrando así un cambio consciente de un lenguaje *emic* a uno *etic*. La coherencia práctica se evidenció en la manera como la evaluación comenzó a permear la enseñanza y viceversa. La coherencia procedimental se estableció en el diseño de las pruebas, las cuales acusaban relación bidireccional entre el constructo por evaluar y la tarea. Concretamente, caminando desde la definición de la competencia comunicativa y sus áreas, descubrimos la inmensa y fructuosa diversidad en la evaluación y nos fuimos percatando del rigor y de la sistematicidad necesarios en una evaluación responsable. Al hacerlo, pasamos de una idea fraccionada de las cualidades de las pruebas a una concepción sistémica de éstas, ineludibles en cualquier proceso evaluativo. Esta búsqueda de la congruencia entre lo requerido por los programas, lo evaluado y el desempeño del estudiante, nos llevó a ubicar a la Justicia y a la Democracia como brújula orientadora de nuestras prácticas evaluativas.

CONCLUSIONES

Es muy satisfactorio concluir este artículo resaltando la manera como la metodología empleada y el compromiso de los participantes hicieron que este estudio fuera enriquecedor y congruente en su tema, su desarrollo y su impacto. Por ello, aseveramos que la investigación-acción participativa es terreno fértil para la construcción de conocimiento –en este caso sobre evaluación– a partir de la literatura, de la experiencia profesional y personal de los investigadores, y de los ciclos de aplicación, reflexión y discusión. Además, el trabajo colaborativo y dialógico propio de este tipo de investigación coadyuva al establecimiento de relaciones cordiales de respeto y convivencia, de sentido de pertenencia y colegaje, esenciales para dicha construcción colectiva de conocimiento y su aplicación.

La justicia y la democracia, como hallazgos sustanciales de este estudio, son así mismo principios de la metodología de investigación-acción debido a que con ella se busca la justicia y la democracia en las prácticas de los participantes. Este estudio demostró que en la educación estos dos principios incorporados al HACER pedagógico revierten

en el SER del maestro. En el **hacer**, estos principios se construyen a medida que a) se establece un lenguaje común y *etic* en términos de la evaluación de los aprendizajes, y b) se van refinando la variedad de tipos y formas de evaluación, los instrumentos y formatos, y el diseño. La relación interdependiente entre estos tres últimos componentes del **hacer** contribuye a que la evaluación sea cada vez más rigurosa y sistemática. En el **ser**, la democracia y la justicia se revelan en a) la preocupación por el bienestar de los evaluados, b) el interés por conocer los programas y contenidos y c) el deseo de alcanzar claridades esenciales para unas prácticas evaluativas éticas. Por las consecuencias sociales de la evaluación, se hacen indispensables la ética y la responsabilidad social de los evaluadores. La primera se visibiliza en la claridad en las intenciones, la implementación y los usos de la evaluación; la segunda, en la pertinencia, la justicia y la democracia en las prácticas evaluativas.

La promoción de estudiantes a niveles superiores sin contar con la competencia comunicativa requerida fue la motivación del primer estudio. La propuesta de un sistema consensuado para generar coherencia entre la competencia comunicativa del estudiante y su promoción a niveles superiores se constituyó en la motivación de este segundo estudio. Este artículo presentó cómo los principios, conceptos y procedimientos evaluativos plasmados en el Sistema consensuado de evaluación son garantes de coherencia entre la competencia comunicativa del estudiante y su promoción. Esta coherencia se alcanza en prácticas evaluativas cuando la competencia comunicativa por evaluar y su materialización en las tareas están alineadas con la validez sistémica, la de contenido y la de constructo –de donde se infiere la competencia comunicativa– y con las demás cualidades –observables en los instrumentos y formatos–.

La competencia estratégica indiscutiblemente juega un papel vital no sólo en la evaluación de los aprendizajes sino en cualquier otra instancia de aprendizaje. Con respecto a la evaluación, concluimos y aseguramos que el lugar de la competencia estratégica está en la evaluación formativa, específicamente en la autoevaluación y la realimentación. Igualmente, nos atrevemos a afirmar que la competencia estratégica está mediada por otros factores como los personales y otros conocimientos como el temático; no obstante, se requiere más investigación.

Las políticas lingüísticas gubernamentales no eran objetivo de esta investigación, pero tanto la dinámica de la investigación–acción como

la temática estudiada nos permitieron comprobar que la evaluación de los aprendizajes permea el acto pedagógico. Por tanto aseveramos que la propuesta de prácticas evaluativas justas y democráticas emanada de la aplicación del Sistema de evaluación en este estudio deviene un primer paso hacia el empoderamiento de los maestros “históricamente excluidos”, empoderamiento necesario para incidir en las políticas lingüísticas del país desde una mirada crítica a sí mismos y al sistema educativo.

Cada avance trae consigo nuevas preguntas, por ello quedan por explorar diversos temas. Uno de ellos es el efecto de arrastre de las prácticas evaluativas justas y democráticas en la práctica docente y en el aprendizaje. Otro sería la repercusión de la implementación del Sistema con los demás profesores de los programas participantes, o su adaptación en otros. Adicionalmente, podría considerarse la contribución de la evaluación formativa con sus tres condiciones al aprendizaje de los estudiantes y al desarrollo de su autonomía.

REFERENCIAS

- Arias, C. I.; Frodden, C.; Vallejo, M.; Usma, J. A.; Murillo, G. J & Molina, J. M. (2003). Prácticas evaluativas en la adquisición de lenguas extranjeras. Manuscrito del proyecto presentado al CODI, Universidad de Antioquia, Medellín, Colombia.
- Arias, C. I. y Maturana, L. M. (2005). Evaluación en lenguas extranjeras: discursos y prácticas. *IKALA, Revista de Lenguaje y Cultura*, 10 (16), 63-91.
- Arias, C. I.; Estrada, L.E., Areiza, H. N. & Restrepo, E. A. (2009). *Sistema de evaluación en lenguas extranjeras*. Grupo de investigación acción y evaluación en lenguas extranjeras (GIAE), Escuela de Idiomas, Universidad de Antioquia. Medellín, Colombia: Reimpresos, duplicación de textos y documentos académicos de la Universidad de Antioquia.
- Arias, C. I. & Restrepo, M. I. (2009a). La investigación acción en educación: un camino hacia el desarrollo profesional y la autonomía. *IKALA, Revista de Lenguaje y Cultura*, 14 (22), 109-122.
- Arias, C.I. & Restrepo, M. I. (2009b). Working towards consensus in action research. En M. L. Cárdenas Beltrán (Ed.), *Investigación en el aula en L1 y L2 – estudios, experiencias y reflexiones* (pp. 161-172). Bogotá: Universidad Nacional de Colombia, Facultad de Ciencias Humanas.
- Bachman, L. & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

- Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32 (4), 653-728.
- Brown, H. D. (2004). *Language assessment. Principles and classroom practices*. New York: Longman.
- Burns, A. (1999). *Collaborative action research for English language teachers*. Cambridge: Cambridge University Press.
- Bustos, F. (1997). Mutaciones en evaluación y lineamientos del ministerio: análisis y crítica (documento incompleto).
- Carroll, B. J. (1993). Typologie des tests de langue. *Le Français dans le Monde, numéro spécial: Évaluation et certifications en langue étrangère*.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6 (2), 5-35. University of California: Regents.
- Chisholm, R. F. & Elden, M. (1993). Features of emerging action research. *Human Relations*, 46 (2), 275-298.
- Estrada, L. & Vallejo, M. (2007). Réussir une évaluation sommative en classe de langues grâce à une évaluation formative performante. *Synergies Venezuela : Revue de Didactologie des Langues-Cultures*, 2.
- Frodden, C., Restrepo, M. I. & Maturana, L. M. (2004). Analysis of assessment instruments used in foreign language teaching. *ÍKALA, Revista de Lenguaje y Cultura*, 9 (15), 171-201.
- Hadji, C. (1997). *L'évaluation démystifiée. Collection pratiques & enjeux pédagogiques*. Paris: ESF Editeur.
- Instituto Cervantes. (2002). *Marco de referencia europeo para el aprendizaje, la enseñanza y la evaluación de lenguas*. (Traducción al español de Alejandro Valero Fernández). Consultado el 13 de junio de 2007 en <http://cvc.cervantes.es/obref/marco/>
- Jorba, J. & Casellas, E. (Eds.). (1997). *La regulación y la autorregulación de los aprendizajes*. Madrid: Editorial Síntesis.
- López, A. & Bernal, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile*, 11 (2), 55-70.
- López, A. (2010). La evaluación formativa en la enseñanza y aprendizaje del inglés. *Voces y Silencios: Revista Latinoamericana de Educación*, 1(2), 111-124.
- Maiese, M. (2003). Principles of justice and fairness. Consultado el 20 de mayo de 2010 en http://www.beyondintractability.org/essay/principles_of_justice/
- Maturana, L. M., Restrepo, M. I. & Ferreira, M. P. (2009). *Módulo-guía en evaluación de los aprendizajes en lenguas extranjeras*. Medellín, Colombia: Reimpresos, duplicación de textos y documentos académicos de la Universidad de Antioquia.
- Muñoz, A. & Alvarez, M. E. (2008). A study of the effect of a writing assessment system on instruction. *MEXTESOL Journal*, 32 (2), 55-70.

- Muñoz, A. & Álvarez, M. (2009). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27 (1), 1-17.
- Muñoz, J. H. (2009). Exploring teachers' practices for assessing reading comprehension abilities in English as a foreign language. *Profile*, 11(2), 71-84.
- Pawlikowska-Smith, G. (2002). *Canadian Language Benchmarks 2000: Theoretical framework*. Centre for Canadian Language Benchmarks. Consultado el 27 de mayo de 2008 en <http://www.language.ca>
- Shohamy, E. (1998). Evaluation of learning outcomes in second language acquisition: A multiplism perspective. En H. Byrnes (Ed.), *Learning foreign and second languages* (pp. 238-261). NY: The Modern Language Association of America.
- Shohamy, E. (2001). *The power of tests – A critical perspective on the uses of language tests*. London: Longman.

SOBRE LAS AUTORAS

Clara Inés Arias

Magíster en inglés como segunda lengua (Universidad de Hawai'i); Profesora Titular Escuela de Idiomas, Universidad de Antioquia, en investigación, evaluación y fonética; investigadora de temas de evaluación del Grupo de Investigación-Acción y Evaluación en Lenguas Extranjeras (GIAE, Categoría C, Colciencias).

Correo electrónico: clariaspeale@gmail.com.

Liliana María Maturana

Especialista en Lenguas extranjeras (Universidad de Antioquia); docente auxiliar de tiempo completo de adquisición de lengua, escritura, historia y cultura inglesa y estadounidense, e investigación educativa, de la Fundación Universitaria Luis Amigó; profesora cátedra de investigación, práctica académica y trabajo de grado de la Universidad de Antioquia. Investigadora de temas de evaluación de los aprendizajes, procesos de adquisición y mediaciones tecnológicas del grupo Construcciones Investigativas en Lenguas Extranjeras (CILEX-FUNLAM).

Correo electrónico: lili.balk@gmail.com.

María Isabel Restrepo

Licenciada en Lenguas extranjeras (Universidad de Antioquia); docente de tiempo completo de adquisición de lengua de la Fundación Universitaria Luis Amigó. Investigadora de temas de evaluación de los aprendizajes, procesos de adquisición y mediaciones tecnológicas del grupo Construcciones Investigativas en Lenguas Extranjeras (CILEX-FUNLAM).

Correo electrónico: m.isabelrpo@gmail.com.

Fecha de recepción: 13-07-2011

Fecha de aprobación: 16-04-2012