

Dimensiones ética y técnica en la literacidad de evaluación de lenguas^{1, 2}

Hugo A. Andrade Mayer³, Universidad Nacional Autónoma de México, Toluca, Estado de México, México

Resumen

La literacidad en evaluación de lengua es el conjunto de competencias que permiten a un individuo entender, evaluar y diseñar instrumentos de evaluación de lengua (Pill y Harding, 2013). Se compone de conocimiento sobre la lengua, de habilidades técnicas de diseño de instrumentos de evaluación y de principios éticos (Davies, 2008). A las últimas dos se les ha denominado dimensiones técnica y ética de la literacidad en evaluación, respectivamente. En el ámbito educativo, los profesores de segunda lengua son quienes generalmente están a cargo de diseñar y aplicar instrumentos de evaluación. Para este estudio, se entrevistó a dos profesoras de lengua inglesa en nivel superior para conocer acerca de su literacidad en evaluación de lengua. Con base en un análisis de su discurso fue posible identificar sus prácticas, así como la manera en que se estructuran las dimensiones técnica y ética y la relación entre sus componentes. Como resultado, ha sido posible tener un acercamiento más humanista al trabajo del profesor de lengua como evaluador. El estudio se llevó a cabo desde la perspectiva del pensamiento complejo (Morin, 1990), considerando que la evaluación es un sistema integrado de varios componentes que interactúan entre sí, creando a su vez un nuevo sistema complejo. El estudio aporta a la documentación de la práctica profesional del profesor de lengua en el papel de evaluador, así como a establecer una base de análisis para la toma de decisiones en varios ámbitos, como el desarrollo del profesor de lengua, la identificación de necesidades de literacidad y los procesos de evaluación de lengua.

Abstract

Language assessment literacy is a stock of competencies that enable an individual to understand, evaluate and create language assessment instruments (Pill y Harding, 2013). It consists of language knowledge, language assessment technical skills, and ethical principles (Davies, 2008). The latter two are the technical and the ethical dimension, respectively. In education, language teachers are generally in charge of developing and administering language assessment instruments. In this study, two higher education language teachers were interviewed to find out about their language assessment literacy. Some of their assessment practices were identified using discourse analysis, as well as the way in which the technical and ethical dimensions are structured and how their components interact. Consequently, it was possible to approach the assessor role of a language teacher more humanistically. The study was carried out under the perspective of complex thinking (Morin, 1990), considering language assessment as a system that consists of many different components interacting and developing into another new and complex system. The research contributes to documenting actual language teachers' practices and their role as assessors and sets the ground for analysis and decision making in different areas such as language teacher development, language assessment literacy needs analysis, and assessment processes.

Introducción

La evaluación del dominio de una segunda lengua (L2) requiere de una competencia particular, la literacidad en evaluación (Taylor, 2009; Inbar-Lourie, 2016), que consiste de tres componentes. El primero lo constituyen las habilidades técnicas necesarias para proporcionar validez, confiabilidad y practicidad a los instrumentos, El segundo es el conocimiento lingüístico, específicamente de la competencia comunicativa, que es el constructo a evaluar. El tercero consta de los principios éticos que guían y regulan la conducta del evaluador.

Los numerosos aspectos de estos componentes hacen de la tarea de evaluar el aprendizaje de una segunda lengua un proceso que exige al evaluador poseer una serie de conocimientos y habilidades que se esperan del especialista en evaluación. La diversidad de situaciones en las que se llevan a cabo procesos evaluativos del aprendizaje de lengua requiere que, para que éstos sean comprendidos, primero sean ubicados contextualmente (Fulcher, 2012). Por tal motivo, este estudio se centra en la evaluación del aprendizaje de

1 This is a refereed paper. Received: 11 June, 2019. Accepted: 9 September, 2019.

2 Ethical and Technical Dimensions of Language Assessment Literacy

³ handradem@gmail.com

lengua, específicamente en educación superior, donde el evaluador de L2 es un profesor de lengua y es, generalmente, quien está a cargo de la tarea de diseñar instrumentos de evaluación, aplicarlos, calificarlos y reportar resultados.

Es así que se han considerado dos de los componentes de la literacidad en evaluación de lengua, denominados dimensión técnica y dimensión ética, para analizarse en el discurso de dos profesoras de lengua, ubicadas en el contexto mencionado. Observadas a la luz de la complejidad (Morin, 1990), las participantes nos proporcionan una mirada humanista e introspectiva a la práctica evaluativa y diseño de exámenes del profesor de lengua y a la vez un entendimiento de la literacidad en evaluación de estos profesores.

Edgar Morin (1990) plantea que la ciencia analiza, descompone en partes, sin embargo, pierde de vista las relaciones entre los componentes de los sistemas. Considerando el proceso de diseño de exámenes de lengua como un sistema, analizar los componentes de la literacidad en evaluación de lengua y cómo son integrados por los docentes nos lleva a un mejor entendimiento del proceso evaluativo. La base epistemológica para la investigación fue la teoría de la complejidad (Morin, 1990; Munné, 2005; Castro Sáez, 2011), bajo la cual se intenta comprender la evaluación del aprendizaje de lengua, el diseño de exámenes y la individualidad de las participantes como sistemas que se componen de manera independiente, pero que se entrelazan creando un nuevo sistema, con numerosos componentes que deben observarse desde una perspectiva completa, en su papel dentro del sistema complejo y no de manera aislada.

Marco Teórico

Como marco de referencia teórica para la investigación, se presenta lo que se entiende por literacidad en evaluación de lengua, en este caso específico la evaluación del aprendizaje de segunda lengua y sus componentes. En seguida se presentarán con mayor profundidad cada una de las dos dimensiones de estudio de la investigación: la dimensión técnica y la dimensión ética de la evaluación del aprendizaje de segunda lengua. Además se incluirán los aspectos de importancia de la teoría de la complejidad para completar el cuadro de conceptos.

Literacidad en evaluación de segunda lengua

El término literacidad se utiliza para referirse a "los conocimientos y actitudes necesarios para el uso eficaz en una comunidad de los géneros escritos" (Cassany, 2005, p.1). Caro y Arbeláez (2009) la definen como "el conjunto de competencias que hacen hábil a una persona para producir, gestionar, editar, recibir ya analizar información en un determinado contexto para transformarla en conocimiento" (p. 5). Montes Silva y López Bonilla (2017) explican que existe un sinnúmero de literacidades, entre ellas la disciplinar, que es un tipo de literacidad académica y que implica un alto grado de especialización.

Como literacidad disciplinar en el ámbito de enseñanza de segunda lengua y específicamente en evaluación de segunda lengua, el término literacidad en evaluación de segunda lengua (LEL) se ha utilizado para referirse a los conocimientos, habilidades y principios que todo individuo relacionado con la práctica evaluativa debe contar para llevar a cabo las actividades correspondientes (Inbar-Lourie, 2016).

La investigación y desarrollo de la LEL se ha dado principalmente en este siglo en tres principales vertientes. En primer lugar, estudios que buscan la definición del concepto y construcción de un marco teórico en que se definan sus componentes (Davies, 2008; Taylor, 2009 y 2013; Fulcher, 2012). En segundo término, los estudios que buscan identificar necesidades de conocimiento de parte de los involucrados, léase administradores, usuarios, candidatos, alumnos y desarrollo profesional de evaluadores y profesores (Taylor, 2013; Inbar-Lourie, 2016). Como tercera tendencia en estudios de LEL, aquellos que definen las implicaciones que este tipo de literacidad tiene para profesores de lengua (Fulcher, 2012; Giraldo, 2018; Yastıbaş y Takkaç, 2018, Giraldo, 2019), quienes generalmente cargan con la mayor parte del peso y de las responsabilidades del diseño, pilotaje, aplicación, calificación y reporte de instrumentos de evaluación.

La composición de la LEL en conocimiento, habilidades y principios, sugerida por Davies (2008), parece ser un consenso en el estudio del concepto, es decir, el conocimiento sobre cómo medir la lengua, las habilidades para el diseño de instrumentos de evaluación o exámenes y principios sobre el uso adecuado y justo de instrumentos. Fulcher (2012) propone un concepto operacional más detallado y enfocado en los profesores de lengua, que se compone por prácticas, principios y contextos. Las prácticas y los principios equivalen a

las habilidades y principios mencionadas por Davies (2008). La diferencia más importante se da en la contextualización planteada por Fulcher, que integra marcos de referencia históricos, sociales, políticos y filosóficos que serían utilizados para explicar el impacto de los instrumentos.

Considerando las coincidencias entre los componentes propuestos por Davies y Fulcher, las habilidades (Davies, 2008) y las prácticas (Fulcher, 2012) conformarían lo que denominaremos dimensión técnica de la literacidad. Los principios que guían la práctica evaluativa, como ética, equidad y justicia, integrarán la dimensión ética. En seguida, se hace una propuesta de componentes de cada una de estas dimensiones, lo que permitirá tener una visión más completa de la literacidad en evaluación.

Dimensión técnica de la evaluación del aprendizaje de segunda lengua

La dimensión técnica de la evaluación del aprendizaje de segundas lenguas consiste en una serie de conceptos técnicos de diseño de instrumentos de evaluación, que se encuentran en parte del discurso de los profesores de lengua analizado en este artículo y tienen su base en la literatura sobre el proceso de diseño de exámenes de lengua de Bachman y Palmer (1996), Hughes (2003), Fulcher y Davidson (2007) y Brown y Abeywickrama (2010). Entre dichos conceptos componentes de la dimensión técnica se considera la validez como un eje principal, que a su vez está fundamentado en una serie de cualidades que aportan a la validez de un instrumento, como validez del constructo, validez de contenido y confiabilidad. Asimismo, otro componente de la dimensión técnica es la practicidad, propiedad de los exámenes que, si bien no es antagonista, es con frecuencia una limitante de la validez. Estos componentes conforman las categorías de análisis de la literacidad en evaluación de lengua encontradas en el discurso de las participantes en el estudio; se muestran en la Figura 1 y se definen en los párrafos siguientes.

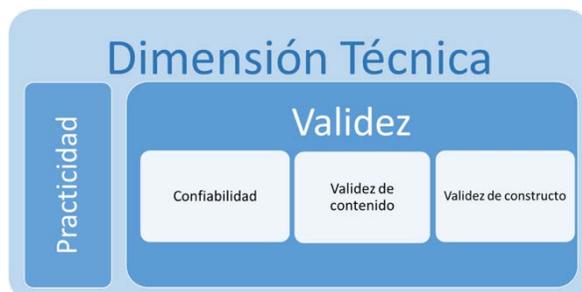


Figura 1: Dimensión técnica de la literacidad en evaluación de L2

La validez de constructo es el grado en que es posible interpretar el resultado de un examen como indicador de la habilidad o constructo que se quiere medir (Bachman y Palmer, 1996). El constructo a medir en los exámenes de lengua debiera ser la competencia comunicativa, que tiene como manifestaciones las habilidades como producción y comprensión oral y escrita. Un examen de lengua será válido en la medida en que el dominio de estas habilidades, en mayor o menor grado, sea demostrado por el resultado en el examen.

La validez de contenido consiste en que el examen constituya una muestra representativa de la competencia o del logro a ser medidos (Fulcher y Davidson, 2007; Brown y Abeywickrama, 2010). El profesor de L2 debe asegurarse que el examen refleje el logro de los objetivos del curso (Brown y Abeywickrama, 2010), lo que se cumple al evaluar de manera directa contenido gramatical, léxico o desempeño en alguna de las habilidades estudiadas o ejercitadas.

Otro de los componentes de la dimensión técnica es la confiabilidad, que es la consistencia en los resultados de un examen (Brown y Abeywickrama, 2010). Se logra con base en tres aspectos principales (Genesee y Upshur, 1996): primero, confiabilidad del instrumento, que implica desambigüedad en instrucciones, reactivos, opciones y respuestas; segundo, confiabilidad del evaluador, que se refiere al entrenamiento, unificación de criterios al evaluar, uso de rúbricas, claves de respuesta, etc.; tercero, confiabilidad de la aplicación, que supone una aplicación de examen o evaluación con las condiciones necesarias para que factores ajenos al instrumento, como ruido, espacio, ambiente, luz, posibilidades de trampa, que afecten el resultado del examen.

La practicidad es la cualidad que consiste en que los recursos necesarios para el diseño, aplicación y calificación de un instrumento de evaluación no excedan los recursos disponibles (Bachman y Palmer, 1996).

Los tipos de recursos a los que nos referimos son humanos y financieros, incluyendo el tiempo que dediquen las personas que intervengan en alguna de las etapas, y materiales.

Dimensión ética en la evaluación del aprendizaje de segunda lengua

Para conformar la dimensión ética de la evaluación del aprendizaje, fue necesario establecer un marco de referencia. Con base en la revisión de la literatura sobre ética en la evaluación de lenguas y una visión desde la perspectiva del pensamiento complejo, en particular la ética de Edgar Morin (2006), fue posible construir dicho marco, el cual posteriormente daría pie a las categorías de análisis resultantes del análisis del discurso de las participantes. En seguida, y de manera breve, se presenta el marco de referencia planteado

Los estudios sobre ética y diseño de exámenes de lengua en los últimos años se han enfocado principalmente en el aspecto social y en la responsabilidad de los evaluadores (Shohamy, 2000; McNamara y Roever, 2006), en el uso apropiado de los instrumentos (Bachman y Palmer, 2010) y en la equidad y justicia o *fairness* (Kunnan, 2000 y 2004), conceptos que conformarían el aspecto ético del profesor de lengua. Sin embargo, el estudio de la ética del profesor de lengua en su papel como evaluador requiere de trabajo de investigación con mayor profundidad.

Todos los aspectos éticos anteriores están estrechamente relacionados con la validez de constructo, elemento clave en la dimensión técnica. Actuar de manera ética en evaluación y diseño de exámenes de segunda lengua implica necesariamente la búsqueda de la validez de los instrumentos. El discurso de los profesores, donde revelan su práctica evaluativa, nos permitirá explorar esa interacción, observar cómo se construye la literacidad en evaluación y cómo se conforma en un sistema de toma de decisiones y lo que el profesor busca conseguir en ese aspecto particular de su práctica docente.

Adicionalmente a los conceptos de responsabilidad, justicia y equidad, para comprender la dimensión ética de la LEL se consideró necesario ubicar el aspecto ético desde una perspectiva ética-humanista, mas no moralista. Es así que se utilizaron los principios imperativos antagonistas de la ética compleja de Morin (2006), quien explica que el ser humano actúa con base en su moral, pero que con frecuencia se encuentra con disyuntivas que presentan acciones antagonistas, incongruentes entre sí y que son, ambas, un imperativo.

Los continuos imperativos antagonistas se presentan en pares como esencia y urgencia, inclusión y exclusión, egoísmo y altruismo, por mencionar algunos ejemplos. Urgencia y esencia implican acciones que se enfocan en lo urgente, que requieren una solución o decisión inmediata, donde esto tendrá más valía que lo importante o esencial. En ocasiones será necesario posponer lo importante y actuar con urgencia. La inclusión implica una manifestación de pertenencia a un grupo, una enunciación del "nosotros" como unidad, mientras que la exclusión involucra una diferenciación entre el "yo" y el "ustedes" como elementos ajenos.

Estos imperativos juegan un papel relevante en la ética compleja de Morin (2006), que se basa en el entendimiento de la trascendencia de la triada individuo-sociedad-especie en las acciones del ser. En esta relación de tres elementos, actuar con ética implica que las acciones realizadas por el ser humano sirvan para vincular y lograr el bienestar de los tres componentes de dicha triada, del individuo, de la sociedad y de la especie. Se considera que sumar esta postura con respecto a la ética del ser humano a la del profesional de lengua ayuda a explicar y relacionar comportamientos, acciones y decisiones con un marco de referencia.

La figura siguiente (Figura 2) muestra cómo se integra la dimensión ética de la literacidad en evaluación de L2:

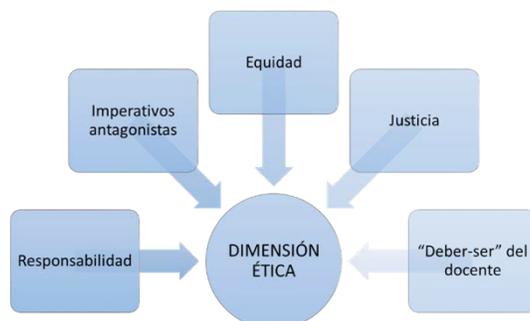


Figura 2: Dimensión ética de la literacidad en evaluación de L2

Metodología

El estudio que aquí se presenta busca identificar en el discurso de profesores de lengua extranjera (inglés) en educación superior, en su papel como evaluadores, cómo manifiestan las dimensiones técnica y ética de la literacidad en evaluación de lengua. Para lograr ello, se plantean como objetivos: primero, identificar y ejemplificar la presencia de dichas dimensiones en el discurso de los profesores; y segundo, mostrar la valoración de los profesores sobre hechos y pronunciamientos relativos al diseño de exámenes de lengua y las dimensiones técnica y ética de la literacidad para construir una base para un modelo de entendimiento del proceso.

El estudio se realizó bajo la perspectiva de la complejidad (Morin, 1990), en la cual se busca el entendimiento del fenómeno a través de encontrar interrelaciones entre sus componentes, así que en lugar analizarlos por separado se pretende integrarlos de manera que pueda observarse el conjunto, el sistema integrado, complejo por el número de componentes y su interacción. En este caso se estudian las dimensiones técnica y ética, componentes indispensables de la práctica evaluativa, y se presentan de manera interrelacionada, actuando en conjunto, uno en consecuencia del otro, pero a la vez de manera independiente. Para ello se ha realizado un estudio de corte cualitativo, a través del cual se busca conocer de manera introspectiva las dimensiones técnica y ética de los profesores de L2, en específico de los profesores entrevistados.

El estudio se llevó a cabo en las siguientes etapas:

- Elaboración de guion de entrevista
- Recolección de datos por medio de entrevistas
- Análisis de información

Para conducir la entrevista se elaboró un guion (Anexo A) y se eligió llevarla a cabo de manera semiestructurada, dando libertad al investigador de profundizar en los aspectos que considerara pertinente. Las preguntas para la entrevista fueron resultado de una serie de entrevistas informales con profesionales de la docencia de lengua y experimentados formadores de docentes de lengua, quienes aportaron ideas con respecto al diseño de exámenes en la enseñanza de L2. Las ideas proporcionadas por dichos profesionales se agruparon, de manera que, junto con la literatura revisada, se obtuvieron siete preguntas que aglutinan los temas más representativos mencionados.

Se entrevistó a dos profesoras de lengua inglesa de una universidad pública en el Estado de México, quienes accedieron de manera voluntaria a participar en el estudio. Las dos tienen licenciatura afín a la enseñanza de lengua inglesa; una de ellas con estudios de maestría y actualmente laboran en diferentes espacios universitarios. Las participantes se diferencian en dos aspectos principalmente. El primero es la experiencia; una de ellas, Ruth, tiene experiencia de 16 años como docente de lengua; la otra, Kathy, tiene cuatro años impartiendo clases de lengua y solamente dos en el nivel superior. El segundo aspecto son sus antecedentes académicos; Ruth cuenta con grado de Licenciatura en Lengua Inglesa y estudios de Maestría en Enseñanza del Inglés concluidos; Kathy es Licenciada en Lenguas con énfasis en traducción de la lengua inglesa pero cursó algunas materias relacionadas con la docencia. Fueron seleccionadas de entre 8 profesores que accedieron a ser entrevistados debido a que eran las de mayor y menor experiencia en el grupo. Esto representó una oportunidad de explorar una amplia gama de la literacidad en evaluación en dos puntos contrastantes en las carreras de las participantes. Con el fin de conservar el anonimato de las participantes, se les asignaron pseudónimos.

Las entrevistas se llevaron a cabo de manera individual y fueron grabadas en audio y transcritas para su análisis, para el cual se adoptó un enfoque cualitativo, utilizando el análisis del discurso como herramienta. Dentro de la teoría de análisis del discurso Phillips y Hardy (2002) han identificado cuatro perspectivas: el análisis sociolingüístico, estructuralismo interpretativo, análisis crítico y análisis crítico lingüístico. Este estudio se ubica en la tradición del estructuralismo interpretativo, donde el enfoque del análisis es en entender cómo se estructura el mundo con base en el texto y el contexto del hablante y cómo éste construye su realidad. En seguida se presentan los resultados de los análisis realizados.

Resultados

Se analizó cada entrevista con los profesores para identificar los elementos de las dimensiones de la responsabilidad profesional y describir cómo se refieren a ellos para conformar su realidad como profesores de lengua ejecutando el rol de evaluadores. Se analizaron los marcadores de discurso que pudiesen

indicarnos la valoración y postura del profesor con respecto a una enunciación, *modus* y *dictum* (Fuentes Rodríguez, 1991). *Modus* y *dictum* son términos acuñados por Charles Bally (en Hermoso, 1994) para analizar la relación entre el sujeto enunciador y su propio enunciado. El *dictum* es el contenido del enunciado, indica el saber del enunciador y una valoración sobre la realidad (Fuentes Rodríguez (1991); Hermoso, 1994). Este análisis permitirá tener una visión amplia de las prácticas del profesor, las posibles razones para ellas y su valoración, además de cómo construye su responsabilidad profesional a través de la exposición de su práctica evaluativa y de diseño de exámenes.

En el análisis del discurso que se presenta en este estudio, dentro de cada *dictum* las referencias a medición de las habilidades o a los componentes de la competencia comunicativa, como competencia léxica y sintáctica son consideradas como implicaturas de la validez del constructo, dado que son ejemplos de los constructos que son medidos en los exámenes de lengua. La validez es el eje alrededor del cual las dimensiones técnica y ética se construyen. Así que se han identificado los *dicta* que expresan una acción que implica un efecto en la validez de los exámenes y conforman dichas dimensiones. Los *dicta* que implican confiabilidad y practicidad complementan la dimensión técnica y las referencias o implicaturas a aspectos como el cuidado del alumno, justicia y equidad son parte de la dimensión ética.

A continuación se presenta el análisis del discurso de las profesoras acerca de sus prácticas evaluativas. El análisis se estructura iniciando con una contextualización del extracto de la entrevista a analizar, seguido de la transcripción; ésta muestra algunas partes subrayadas, que son elementos del discurso que pudiesen aportar al entendimiento de los argumentos de los participantes en cuanto a su práctica en evaluación. El análisis de cada fragmento se presenta en seguida, indicando los componentes de responsabilidad profesional identificados y cómo son valorados o modulados por medio de su *modus* y *dictum*.

Análisis de entrevista Participante 1: Ruth.

Ruth: Validez, practicidad y altruismo

En su discurso, Ruth comenta sobre aspectos de su trabajo como docente, a través de los cuales pueden notarse aspectos de la dimensión técnica, como practicidad y validez y un aspecto de la dimensión técnica que es el altruismo.

Ruth trabaja en un par de instituciones y comenta que en una de ellas se elaboran exámenes departamentales y en las otras tiene que diseñar todos los exámenes de sus grupos:

[R1]...ahí (en la Facultad de Derecho) la academia hace que todos los maestros hagan varios exámenes, no sé por ejemplo pide a dos o tres maestros que hagan el primer parcial otros dos o tres maestros que hagan el segundo y así para que este pues verifique que todos estemos viendo lo mismo entonces creo que sí es más fácil porque nada más hago uno, en otros lados siempre tengo que hacer yo los míos para el nivel que esté trabajando el primer parcial el segundo el final el extra y el título. (Ruth: Entrevista)

Ruth expresa que (el trabajo) es "*más fácil*" en la Facultad de Derecho debido a que entre todos los profesores elaboran exámenes departamentales y de esta manera solo realiza un examen, lo que implica practicidad. Cabe hacer notar la modalidad empleada por Ruth al expresarse acerca de esta conveniencia "...creo que sí es más fácil...". La expresión "*creo que*" no parece denotar incertidumbre sino que, en este caso, podría estar posicionando al hablante moderadamente distante con respecto a una situación que pudiese percibirse como negativa.

En el siguiente extracto se puede apreciar un ejemplo de la manera en Ruth vincula la validez con altruismo. En la entrevista Ruth declara que realiza ciertas acciones para que sea más fácil para ella evaluar; al preguntársele porqué ella respondió:

[R2] Porque me gusta me gusta ver si o comprobar encontrar algo con lo que yo pueda ver que mis alumnos de verdad están aprendiendo... (Ruth: Entrevista)

El *modus* "*me gusta*" denota un interés, una pretensión personal o profesional que en este caso se refiere al *dictum* "...comprobar (o) encontrar algo...", que significa medir el constructo. Es así que la maestra manifiesta un deseo de que su evaluación sea válida ("... ver que mis alumnos de verdad están aprendiendo..."), que sirva para demostrar el aprendizaje de sus alumnos. Ruth expresa un bienestar para los estudiantes y se puede percibir en el *modus* utilizado ("...me gusta ...que yo pueda..."). El uso del subjuntivo en el verbo *poder*, en este caso manifiesta un deseo de posibilidad.

Ruth: La validez del instrumento como obligación del profesional

Para Ruth, la validez es un tema de relevancia en cuanto a exámenes de lenguas refiere, ya sea validez como término en general, validez de constructo o validez de contenido. Esto puede observarse en el siguiente ejemplo, donde ella habla de su práctica en diseño de exámenes en sus inicios como docente y explica las dificultades que tenía para ajustar sus instrumentos al nivel adecuado y al contrastar aquella práctica con su proceder actual, ella concluye:

[R3] *...ya puedo ver si estoy calificando lo que debo calificar ni de más ni de menos sino lo que deben saber.* (Ruth: Entrevista)

La validez de un instrumento es calificar lo que se busca calificar, así lo expresa Ruth, a lo que además agrega "...debo..." lo que implica una obligación. Este extracto también es una muestra de la justicia o equidad manifestada por la maestra, con la expresión "...ni de más ni de menos sino lo que deben saber."

Ruth: Urgencia y esencia

Morin (2006) declara que el humano en ocasiones actúa con urgencia y olvida lo esencial de las cosas. Este postulado nos ayuda a entender el actuar de Ruth. Actuar con urgencia es un factor de decisión en su práctica. En su discurso puede notarse cómo debido a ello, a pesar de mostrar conciencia sobre la esencia de la práctica, esta se ve modificada por el poco tiempo disponible.

Al responder a la pregunta sobre si en alguna vez había dudado de alguno de los instrumentos realizados por ella, contestó lo siguiente:

[R4] *...sí creo que sí, sobre todo cuando no tengo mucho tiempo para revisarlo. un examen de inglés a mi me toma más o menos en horas de trabajo seguidas como dos días pero nunca lo hago en dos días, siempre empiezo un fin de semana hago una parte la reviso y como que hago un frame así de que aquí voy a poner este ejercicio aquí de este...* (Ruth: Entrevista)

La urgencia se expresa con el enunciado "(dudo del examen) *cuando no tengo mucho tiempo para revisarlo ...*"; este *dictum* recibe el *modus* "...creo que sí (he dudado de un examen)..." Ruth parece reflexionar y este *modus* parece expresar una aceptación. En el resto de la respuesta describe cómo hace un examen cuando tiene el tiempo suficiente, sin embargo parece reconocer que cuando no lo tiene y el examen no es revisado como la maestra acostumbra surgen dudas sobre su confiabilidad. Por otro lado, Ruth expresa conocimiento acerca del desarrollo de un examen y conciencia de lo que implica trabajar en la esencia, en los aspectos relevantes. No hay duda si se tiene el tiempo para enfocarse en lo esencial.

Ruth continúa la respuesta a la pregunta, describe los momentos en que improvisa un examen, cuando no tiene tiempo suficiente para preparar reactivos o como ella manifiesta "revisarlo", "de planearlo":

[R5] *...cuando lo he hecho de que mañana es el examen yo yo oh Dios mío mañana de esos exámenes es de los que más desconfío porque no los planeo, los tomo de un ejercicio de aquí de acá del mismo libro con el que estoy trabajando pero siento que repito algunos temas como que no lo puedo planear igual, y de los resultados de esos exámenes si dudo, aunque trato que eso no me pase porque si no después ya no puedo medir...* (Ruth: Entrevista)

Ruth expone "...(cuando no planeo el examen) *repito algunos temas...*" lo cual no es una práctica deseable, ya que reduce la confiabilidad del test y esto se confirma en la declaración "...de esos exámenes es de los que más desconfío porque no los planeo..." Ruth parece aceptar que realiza esta práctica con el *modus* del enunciado "...siento que repito algunos temas..." y más adelante en la oración confirma que esa práctica es la razón que genera duda en su propio instrumento "...y de los resultados de ese examen si dudo...".

Ruth demuestra conciencia de que las características principales del test, validez y confiabilidad, son afectadas por la falta de planeación del instrumento. Se ha analizado como se expresa Ruth acerca de la confiabilidad. Por otro lado sobre cómo se ve afectada la validez del instrumento Ruth se expresa de la siguiente manera: "...después (de elaborar un examen no planeado) *ya no puedo medir...*", el impacto de un instrumento improvisado se expresa con el *modus* "no puedo", la imposibilidad de llevar a cabo la medición adecuada que implica una reducción de la validez del instrumento.

Para Ruth, diseñar exámenes con tiempo, un examen planeado es esencial en una buena práctica, mientras que la urgencia lleva a una improvisación que tienen efectos en el desempeño profesional.

*Análisis de entrevista Participante 2: Kathy*Kathy: Practicidad y confiabilidad

En el extracto que se muestra a continuación, se preguntó a Kathy la manera en que diseña los exámenes que aplica. Cabe mencionar que en la institución donde labora, se trabaja con un libro de texto, al que Kathy se refiere como "el libro" o "el método". El proceso de diseño de exámenes de lengua que sigue Kathy, en sus propias palabras, es el siguiente:

[K1] *depende qué curso esté dando, este entonces me baso generalmente en el libro que me den, es muy sencillo hacer esos exámenes, entonces tomo lo que estuve viendo en clase, lo que considero que es necesario que los chavos sepan entonces es así como copio y pego el examen como venga en el formato del disco del método...* (Kathy: Entrevista)

En el discurso de Kathy, se pueden inferir algunos aspectos del diseño de exámenes que valora, en principio la practicidad; Kathy manifiesta sobre el diseño de exámenes que "...*es muy sencillo hacer esos exámenes...*" declaración, sin modulación, que se contrapone con lo mencionado en la literatura especializada que especifica las complejidades de diseñar exámenes. Otro indicador de la búsqueda de practicidad por parte de Kathy está puede notarse al momento en que expresa "...*copio y pego el examen como venga en el formato del disco del método...*" aunque esta acción podría afectar la confiabilidad del test, ya que los exámenes que menciona la maestra son accesibles a muchas personas.

Kathy: Validez de constructo, de contenido y exclusión

Este pasaje [K1] también es revelador en cuanto a la concepción de validez que tiene Kathy, quien expresa que cuando diseña exámenes, refiriéndose a las acciones que lleva a cabo: "...*tomo lo que estuve viendo en clase...*" enunciado que implica validez de contenido, característica imprescindible en evaluación en el aula de lengua. Kathy continúa con el enunciado "... (tomo) *lo que considero que es necesario que los chavos sepan...*"; el marcador "lo que considero" manifiesta alta subjetividad, ya que el constructo a evaluar no es establecido a priori, sino que es resultado de un razonamiento intrínseco, excluyente con respecto al alumnado o a otros participantes del proceso evaluativo, como colegas profesores.

Kathy: Practicidad, validez e inclusión

Kathy estudió una licenciatura en lenguas, que ofrece en su currículo materias de diseño de exámenes y evaluación. En el momento de la entrevista trabajaba en la institución donde realizó sus estudios de licenciatura; y donde los grupos tienen, generalmente, entre 20 y 25 alumnos. Durante la entrevista se le preguntó qué características debía tener un examen y en su respuesta considera sus estudios y el número de estudiantes.

[K2] *Pues, no me acuerdo de los términos ya porque tiene tiempo que estuve en esa materia pero creo que tienen que ser, no tan fácil para el alumno pero tampoco tan difícil, o sea tienen que como que, ¿cómo se dice challenging en español? ¿retarlo? este para que no sienta como que no vale la pena hacerlo porque está muy sencillo y tampoco se frustra porque está muy difícil y como para nosotros como maestros pues que no sea tan difícil calificar porque la verdad tener tantos alumnos claves de grupos de veintitantos y tener que calificar exámenes muy extensos se me hace así como que un trabajo muy pesado y al final pues muy tedioso, entonces en general pues creo que tienen que retar a los alumnos pero también tiene que ayudar al profesor porque no tiene que ser un dolor de cabeza para ninguno de los dos.* (Kathy: Entrevista)

Kathy manifiesta no recordar los términos técnicos para referirse a las características que debe tener un examen de lengua "...*no me acuerdo de los términos ya porque tiene tiempo que estuve en esa materia...*" Ella tiene formación como profesora de lengua y se refiere a una materia sobre diseño de exámenes que cursó en sus estudios de licenciatura. En este mismo extracto, sobre las características necesarias en un examen, Kathy utiliza en repetidas ocasiones el marcador "tiene que" o "no tiene que" lo cual indica una obligación, bien como una obligación impuesta por un superior o como una característica indispensable, sin la cual el instrumento no sería un examen. No obstante, en el enunciado "...*tener que calificar exámenes muy extensos...*", ese marcador podría estar siendo utilizado como un imperativo que no necesariamente implica actuar con corrección, sino como una actividad no deseada e ineludible. El hecho de "tener que" calificar exámenes tan extensos, como declara Kathy, es una expresión que denota practicidad, en perjuicio de la validez. Entre más fáciles de calificar sean los exámenes será menos probable que sean auténticos y cuenten con alto grado de validez del constructo.

En este mismo pasaje se puede percibir un marcador de inclusión, "nosotros", con el que Kathy expresa un sentido de pertenencia a una comunidad de maestros: "... (el examen) *para nosotros como maestros pues que no sea tan difícil calificar...*", enunciado que también deja ver un deseo por que el examen sea práctico

"*tan difícil de calificar*", marcado por el subjuntivo "...que sea...". Su proclividad hacia la practicidad de los exámenes es evidente en la oración "... (calificar exámenes) *se me hace así como que un trabajo muy pesado...*".

Kathy: Confiabilidad, altruismo, inclusión-exclusión

Se le preguntó si alguna vez había dudado de algún examen, a lo que respondió afirmativamente y dio un ejemplo. Kathy declaró que del examen en sí no había dudado, pero sí de una técnica, a la cual ella denomina ejercicio:

[K3] *Pues en general el examen no a lo mejor el ejercicio, porque una vez me pasó que no vimos el formato de ese ejercicio y les fue muy mal entonces dije pues creo que ahí metí la pata yo porque no tuvimos práctica de ese ejercicio en particular entonces por eso creo que les fue tan mal... a la mayoría les fue súper súper mal, entonces lo que hice fue descartar ese ejercicio por completo porque tampoco se me hacía justo que les fuera a calificar algo que ni siquiera habíamos visto bien y no sé como que esa parte de la justicia la tengo muy metida.* (Kathy: Entrevista)

Kathy declara que "...a lo mejor (duda de) el ejercicio...", donde el modus "a lo mejor" establece o concede una posibilidad de que haya tenido esa duda, lo cual contrasta con el enunciado previo donde asevera no haber tenido duda sobre el examen, en una modalidad declarativa. Más adelante, con el uso del marcador "creo" y la expresión idiomática "metí la pata" admite la posibilidad de haber cometido un error "... creo que ahí metí la pata yo...".

En este enunciado se puede apreciar un "yo" excluyente, en contraposición con otros enunciados que muestran un "nosotros" incluyente que implica que Kathy considera que está del mismo lado que sus alumnos, como clase: "...no vimos el formato de ese ejercicio..." y "...no tuvimos práctica de ese ejercicio en particular...". Entonces, se muestra excluyente al asumir una responsabilidad pero incluyente al referirse a sí misma y su trabajo con los alumnos. En seguida del enunciado anterior, Kathy expresa que no haber practicado ese tipo de técnicas es una posible razón para el resultado obtenido: "...entonces por eso (no haber practicado ese tipo de ejercicio) creo que les fue tan mal...". Esto queda solo como una posibilidad, como denota el marcador "creo que".

El pasaje de la entrevista concluye con dos enunciados para el análisis. Kathy comenta que decidió cancelar la técnica en cuestión "...entonces lo que hice fue descartar ese ejercicio por completo...", mostrando una actitud altruista en pro del alumno y de la confiabilidad del resultado, y concluye: "...tampoco se me hacía justo que les fuera a calificar algo que ni siquiera habíamos visto bien...". Este enunciado muestra exclusión e inclusión a la vez; en primer lugar se ubica como la profesora excluyendo a los estudiantes cuando se expresa acerca de calificar ("que (yo) les fuera a calificar"), pero a la vez incluyente cuando habla de lo que no se estudió en el curso (...algo que ni siquiera habíamos visto...). Asimismo, en este enunciado, Kathy sugiere con el uso del modus "se me hacía", que es una persona que obra con justicia, al igual que en la oración final donde expone "... no sé como que esa parte de la justicia la tengo muy metida...". La modalidad del marcador "no sé" indica falta de certeza o bien podría expresar modestia, ya que en el dictum expresa que es una mujer que tiene la justicia muy arraigada y no quiera ser presuntuosa en ese sentido, y por eso la modulación.

En el extracto hay una cualidad de la dimensión técnica implícita; como puede verse en este caso, la maestra está demostrando un sentido de confiabilidad al cancelar la técnica del examen que desembocó en bajos resultados. Si los resultados obtenidos no son representativos de la competencia del alumno, el test no es confiable.

Discusión

El trabajo que se presenta se suma a los estudios sobre la literacidad en evaluación de lengua, en búsqueda de una definición operacional para su estudio. En este sentido la aportación es en los componentes de lo que Davies (2008) y Fulcher (2012) llaman principios y que hemos denominado dimensión ética; además se proponen unidades de análisis, surgidas del análisis del discurso de las participantes.

En el análisis del discurso que se presenta en la sección anterior se percibe la presencia de las dimensiones técnica y ética en la literacidad en evaluación de lenguas. En ambas entrevistas, la validez, la confiabilidad y la practicidad (Bachman y Palmer, 1996; Brown y Abeywickrama, 2010) se presentan como elementos recurrentes. Otros elementos de la dimensión técnica. como la validez de contenido y la autenticidad, que son considerados tipos de validez, ocurren de manera más esporádica. La dimensión ética se presenta en

expresiones de justicia, altruismo, inclusión, exclusión, urgencia y esencia. No obstante, las dimensiones no se exhiben de manera aislada, particular o explícita, sino que se entrelazan de manera que cada parte del discurso nos permite observar muestras de las dos dimensiones de estudio de la literacidad en evaluación de lengua.

El discurso de las profesoras sobre sus prácticas evaluativas se construye con base en la dimensión técnica, o habilidades, como lo plantea Davies (2008). Cada uno de los temas, cómo son tratados éstos y cómo se construyen los argumentos de las respuestas de las profesoras convergen en un aspecto de la dimensión técnica de la literacidad en evaluación, ya sea la validez o alguno de los conceptos relacionados a ella, o la confiabilidad o la practicidad; aunque también en algunos casos se mostró la necesidad de desarrollar estas características en las prácticas.

La dimensión ética, por otro lado, acompaña a la dimensión técnica en el discurso de las participantes. Para cada práctica evaluativa descrita que refleja algún aspecto de la dimensión técnica hay evidencia de una perspectiva profundamente personal de la participante, reflejo de su ética profesional y personal (si es posible separarlas). Es así que, a través de los marcadores de discurso ejemplificados en el análisis de los resultados pueden percibirse la justicia, el altruismo, la urgencia o la esencia en las prácticas evaluativas de las profesoras.

En el caso de Ruth, la literacidad en evaluación se refleja en su discurso, en primera instancia con una constante implicatura a la validez, como el principio de "medir exactamente lo que se propone medir" (Brown y Abeywickrama, 2010, p. 30) como en los enunciados "...*me gusta ver si o comprobar encontrar algo con lo que yo pueda ver que mis alumnos de verdad están aprendiendo*", "no medir variables irrelevantes o contaminantes" (p. 30) y "...*calificar ni de más ni de menos...*" (Ruth: Extracto 3). Ruth busca con sus instrumentos medir correctamente, con precisión, de manera justa. Reconoce que actuar con urgencia, improvisando en el diseño de exámenes, tiene un efecto que reduce la confiabilidad y la validez del examen "...*de esos exámenes es de los que más desconfío porque no los planeo... de los resultados de esos exámenes si dudo...*" (Ruth: Extracto 5).

Kathy, con cuatro años de experiencia en docencia de lengua, manifiesta explícitamente su necesidad de justicia en su práctica evaluativa, en el sentido de ser justa con los estudiantes "...*tampoco se me hacía justo que les fuera a calificar algo que ni siquiera habíamos visto... como que esa parte de la justicia la tengo muy metida*" (Kathy: Extracto 3), justicia que está estrechamente relacionada con la confiabilidad de los instrumentos. Con base en este mismo pronunciamiento de la profesora, es posible notar que la validez de contenido parece ser un elemento indispensable en la aplicación de instrumentos justos, como lo plantea Kunnan (2004), siendo estos dos elementos, justicia y validez, ejes de las dimensiones ética y técnica de la literacidad en evaluación de Kathy. Al proponer un marco de referencia para el estudio de la justicia (*fairness*) en los exámenes de lengua, Kunnan (2004) menciona como componentes indispensables la validez y la confiabilidad, y es posible observarlo en el discurso de la profesora.

En los casos descritos en este estudio, tanto Ruth como Kathy muestran una necesidad de desarrollo de su literacidad en evaluación. A pesar de que Ruth tiene claridad con respecto al propósito de los exámenes que diseña y aplica, que es lograr la mayor validez posible en sus instrumentos "...*me gusta...ver que mis alumnos de verdad están aprendiendo...*" (Ruth: Extracto 2), la terminología relativa al diseño de exámenes y evaluaciones de lengua no se presenta en su discurso, lo cual demuestra una formación empírica más que teórica, en cuanto a evaluación se refiere. En el caso de Kathy, existe de forma manifiesta el elemento de justicia como un principio base componente de la literacidad en evaluación en su discurso "...*esa parte de la justicia la tengo muy metida...*" (Kathy: Extracto 3). Sin embargo, la ausencia de elementos técnicos tanto en la práctica descrita, como en la enunciación y uso de la jerga de evaluación de lengua denota una necesidad de aspectos elementales de literacidad, como expresa en el enunciado "...*no me acuerdo de los términos...*" (Kathy: Extracto 2). El análisis del discurso de las participantes nos ha permitido tener una idea clara de sus necesidades de literacidad hablando de principios y habilidades técnicas (Davies, 2008; Fulcher 2012), contribuyendo a los trabajos realizados por Fulcher (2012), Inbar-Lourie (2016) y Giraldo (2018), quienes de igual manera, identifican necesidades de capacitación o desarrollo de la literacidad en evaluación; aunque en este estudio se tienen como base una metodología de carácter cualitativo.

Un objetivo del estudio fue encontrar el posicionamiento del individuo con respecto a sus prácticas evaluativas. La valoración sobre los *dictum* emitidos por las profesoras acerca de sus prácticas evaluativas

muestra su postura con respecto a la literacidad en evaluación. Ruth considera la validez como una obligación "...lo que debo..." (Ruth: Extracto 3); admite la posibilidad de duda con respecto a sus instrumentos, especialmente cuando la urgencia prevalece sobre la esencia "...de esos exámenes es de los que más desconfío..." (Ruth: Extracto 4).

Para Kathy la practicidad se antepone a la validez como implica el enunciado "...que (el examen) no sea tan difícil de calificar...". El uso del subjuntivo en este enunciado puede significar obligación o necesidad. En este sentido, se requiere más información para determinar qué significa que un examen "no sea tan difícil de calificar" y así poder determinar cómo se da la relación y balance entre practicidad y validez, ya que la primera podría darse en detrimento de la segunda, o en otro escenario, la falta de practicidad reduciría su validez (Fulcher y Davidson, 2007). Por otro lado, el discurso de Kathy muestra una práctica de cuestionable confiabilidad, que no es modulada, sino expresada en indicativo, como un hecho que no da lugar a duda, y que es el caso de esta práctica "...copio y pego el examen como venga en el libro del método..." (Kathy: Extracto 1). Como lo apuntan Brown y Abeywickrama (2010), esto repercute en la falta de confiabilidad de un examen. No obstante, Kathy también indica valoración y consciencia de confiabilidad y validez de contenido, además de "...tampoco se me hacía justo que les fuera a calificar algo que no habíamos visto bien..." (Kathy: Extracto 3).

El estudio ha revelado cómo, en los extractos presentados, las dimensiones técnica y ética se presentan en el discurso de las profesoras, de manera que los componentes de la dimensión ética, validez, confiabilidad, y practicidad tienen estrechas relaciones y entrelazamientos con los de la dimensión ética, como la justicia, el altruismo, la esencia. Esto establece un sistema complejo que, como sugiere Morin, (1990) asemeja a un tejido que ha de verse como constituyente de una micro-dimensión o individualidad, en este caso cada una de las profesoras y sus prácticas evaluativas y decisiones éticas. Estas a su vez conforman una macro-dimensión, que podría ser la literacidad en evaluación de lengua. Morin no toca en ningún momento algún tema relacionado con la evaluación de lengua, sin embargo, su visión de complejidad en los fenómenos se comparte y se trata de estudiar, aplicando este entendimiento a la problemática.

La literatura sobre evaluación de lengua establece procedimientos, prácticas deseables, estándares y generalizaciones que no observan la complejidad individual y contextual. Ésta, por encima de aquellos aspectos, influye en la literacidad en evaluación de lengua, es decir, en el logro de cualidades de los instrumentos de evaluación como validez o confiabilidad y en los principios éticos reguladores de la conducta del docente. El conocer las decisiones de los docentes de lengua de forma directa y con profundidad provee una visión más completa que pudiese contribuir en su desarrollo de la literacidad en evaluación de lengua.

Conclusiones

El estudio ha cumplido sus objetivos al identificar, en el discurso de las profesoras, las dimensiones técnica y ética de la literacidad en evaluación de lengua. A través de manifestaciones explícitas e implícitas se ha podido probar la presencia de las dimensiones en el discurso de las profesoras. Ha sido posible detectar la valoración hacia dichas dimensiones debido a la presencia de marcadores de modulación, *modus*, que nos ayudan a entender la postura de las profesoras hacia los diversos componentes de las dimensiones. Finalmente fue posible establecer una base o una dirección hacia el estudio de la literacidad en evaluación con un enfoque más humanista, más cercano al principal protagonista que es el profesor de lengua en uno de sus roles, que es como evaluador.

Los análisis de las entrevistas presentadas otorgan la oportunidad de realizar, en el futuro, estudios acerca del proceso seguido por los profesores en este contexto universitario y verificar las prácticas identificadas a través de su discurso y compararlas con las acciones tomadas en los hechos, lo cual sería el complemento para tener una visión más clara y completa de la literacidad en evaluación de las participantes.

De igual manera, fue posible determinar necesidades de literacidad en evaluación de las participantes. La elección de la tradición cualitativa para abordar la problemática ha permitido estos hallazgos desde una perspectiva más cercana a las participantes como seres humanos y profesionales, sin que se partiera de una generalización que podría o no cubrir esas necesidades particulares. Por supuesto, si bien este enfoque de estudio nos permite ese acercamiento, el modelo no permite una aplicación a contextos de mayor envergadura, como institucionales o regionales, que exigirían un enfoque cuantitativo. Sin embargo, el modelo de análisis presentado en este estudio podría ser sujeto de adaptación en este sentido, tomando

como base hallazgos de la tradición cualitativa en la elaboración de instrumentos aplicables a una población mayor en número, que requiera de otro tipo de decisiones.

Es preciso estudiar la literacidad en evaluación como un proceso complejo cuyas dimensiones técnica y ética se desarrollan en un contexto. Tal desarrollo combinado con la complejidad inherente al individuo, resulta en una particularidad que debe ser considerada en la toma de decisiones de carácter evaluativo y de desarrollo de la literacidad en evaluación. Se requiere de investigación que profundice en las conexiones entre los componentes de las categorías y de igual manera dé continuidad al trabajo desarrollado en la conceptualización y operacionalización del constructo "literacidad en evaluación de lengua". Dentro de la complejidad mencionada, será necesario explorar otros ámbitos ligados a estas dimensiones de la literacidad en evaluación y estudiarlos bajo esa perspectiva.

En el futuro será necesario indagar sobre el impacto de los contextos y antecedentes de los profesores en su literacidad. Explorar aspectos como la experiencia, formación y desarrollo profesional, actualización académica, así como su relación con las prácticas evaluativas y necesidades de desarrollo de literacidad en evaluación de lenguas representa oportunidades de investigación que podrán mejorar esta práctica en el salón de clase de lengua.

Referencias

- Bachman, L. F. y Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Brown, H. D. y Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. Pearson Education.
- Caro Salcedo, L. C. y Arbeláez Echeverri, N. C. (2009). Hipertextualidad, literacidad y discurso académico: conceptos para la gestión del conocimiento en la red. *Revista Virtual Universidad Católica del Norte*, 28, 1-23. <http://www.redalyc.org/articulo.oa?id=194214468007>
- Cassany, D. (2005) Investigaciones y propuestas sobre literacidad actual: Multiliteracidad, internet y criticidad. *Congreso Cátedra UNESCO para la lectura y escritura*. <http://www2.udec.cl/catedraunesco/05CASSANY.pdf>
- Castro Sáez, B. (2011). Aportes de Niklas Luhmann a la teoría de la complejidad. *POLIS, Revista Latinoamericana*, 29. <https://journals.openedition.org/polis/2017>
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347. <https://doi.org/10.1177%2F0265532208090156>
- Fuentes Rodríguez, C. (1991). Algunas reflexiones sobre el concepto de modalidad. *Revista Española de Lingüística Aplicada*, 7, 93-108. <https://idus.us.es/xmlui/handle/11441/33002>
- Fulcher, G. (2012) Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <https://doi.org/10.1080/15434303.2011.642041>
- Fulcher, G. y Davidson, F. (2007). *Language testing and assessment*. Routledge.
- Genesee, F. y Upshur, J. A. (1996). *Classroom based evaluation in second language education*. Cambridge University Press.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179-195. <http://dx.doi.org/10.15446/profile.v20n1.62089>
- Giraldo, F. (2019). Language assessment practices and beliefs: implications for language assessment literacy. *HOW: A Colombian Journal for Teachers of English*, 26(1), 35-61. <http://dx.doi.org/10.19183/how.26.1.481>
- Hermoso, A. (1994). Análisis de la modalidad en el discurso. En J. F. Corcuera, M. Dijan, y A. Gaspar (Eds.). *La lingüística francesa. Situación y perspectivas a finales del siglo XX* (pp. 185-195). Universidad de Zaragoza. <https://idus.us.es/xmlui/handle/11441/17996>
- Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press.
- Inbar-Lourie, O. (2016). Language assessment literacy. En Shohamy, E. Or, I y May, S. (Eds). *Language testing and assessment. Encyclopedia of Language and Education* (3rd ed.). Springer. https://doi.org/10.1007/978-3-319-02326-7_19-1
- Kunnan, A. J. (2000). Fairness and justice for all. En A. J. Kunnan (Ed.), *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida, Studies in Language Testing*, 9 (pp. 1-14) Cambridge University Press.
- Kunnan, A. J. (2004). Test Fairness. En M. Milanovic y C. J. Weir (Eds.), *European Language testing in a global context: Proceedings of the ALTE Barcelona Conference, July, 2001, Studies in language testing*, 18 (pp. 27-48). Cambridge University Press.
- McNamara, T. y Roever, C. (2006). *Language testing: The social dimension*. Blackwell Publishing Ltd.
- Montes Silva, M. E. y López Bonilla, G. (2017). Literacidad y alfabetización disciplinar: enfoques teóricos y propuestas pedagógicas. *Perfiles Educativos*, 39(155). http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982017000100162&lng=es&nrm=iso&tlng=es
- Morin, E. (1990). *Introducción al pensamiento complejo*. Editorial GEDISA.
- Morin, E. (2006). *El método 6: Ética*. Ediciones Cátedra.
- Munné, F. (2005). *¿Qué es la complejidad?* <https://teoriadacomplejidade.com.br/wp-content/uploads/2017/01/QueEsLaComplejidad.pdf>
- Phillips, N. y Hardy, C. (2002). *Discourse analysis: Investigating processes of social construction*. Sage.

- Pill, J. y Harding, L. (2013). Defining the language assessment literacy gap. *Language Testing*, 30(3), 381-402. <https://doi.org/10.1177%2F0265532213480337>
- Shohamy, E. (2000). The relationship between language testing and second language acquisition, revisited. *System*, 28(4), 541-553. [https://doi.org/10.1016/S0346-251X\(00\)00037-3](https://doi.org/10.1016/S0346-251X(00)00037-3)
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36. <https://doi.org/10.1017/S0267190509090035>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. <https://doi.org/10.1177%2F0265532213480338>
- Yastibas, A. E., y Takkac, M. (2018). Understanding language assessment literacy: Developing language assessment. *Journal of Language and Linguistic Studies*, 14(1), 178-193. <https://www.jlls.org/index.php/jlls/article/view/861>

ANEXO A

GUIÓN DE ENTREVISTA

1. ¿Evalúa a sus alumnos con exámenes? ¿Qué piensa de evaluar a los alumnos de lenguas con un examen?
2. Podría describir, paso a paso ¿cómo diseña sus exámenes de lengua?
 - a. ¿Pilotea sus exámenes?
 - b. ¿Evalúa las cuatro habilidades?
 - c. ¿Evalúa gramática y vocabulario?
3. ¿Qué características debe tener un examen?
4. ¿Alguna vez ha dudado de un examen?
5. ¿Qué piensa de sus propios exámenes?
6. ¿Ha cambiado la forma en que diseña exámenes de la primera vez que lo hiciste a la fecha?
7. ¿Qué nos dicen tus exámenes de tus alumnos?