

ÓSCAR SOLER CANELA

Profesor colaborador en el Instituto Cervantes de Londres, ha sido coordinador de la Sección de Español y Portugués en el Departamento de Lenguas de Regent's College (Londres) y formador de profesores en Goldsmiths College (Londres). Es miembro de la Asociación Europea de Evaluación de Lenguas (EALTA) y la Asociación Internacional de Evaluación de Lenguas (ILTA). Máster de Enseñanza de Idiomas (especialidad en Lingüística Aplicada) por la Universidad de Lancaster, donde se especializó en evaluación de idiomas, y cuyo tema de memoria fue la vinculación de exámenes de idiomas al *MCER*

## Resumen:

Este artículo es una aproximación al uso de tareas en la evaluación de idiomas. En la primera parte el autor presenta los orígenes de la evaluación mediante tareas (EmT) como una evolución en los últimos años de la evaluación de la actuación lingüística, y la combinación de ésta última con el uso de tareas. El autor describe las diferentes propuestas que se han realizado en el campo de la evaluación de lenguas extranjeras como, por ejemplo, los enfoques *cognitivo / psicolingüístico* o los llamados enfoques *muestra de trabajo*. El autor expone algunas de las reacciones que han causado estos enfoques, y se hacen algunas consideraciones sobre estas propuestas.

**PALABRAS CLAVE:** Tarea, evaluación de lenguas, evaluación de la actuación

**TASK-BASED APPROACHES TO LANGUAGE ASSESSMENT**

**KEYWORDS:** Task, language assessment, language testing, performance assessment

## 1. INTRODUCCIÓN

El interés suscitado desde hace tiempo por la evaluación de la competencia lingüística, ha venido acompañado por un interés creciente en el uso de tareas para este propósito. En los últimos años se han realizado diferentes propuestas orientadas a la evaluación mediante tareas (EmT), propuestas que difieren en su grado de conceptualización, operativización, e investigación. Una de las dificultades que se puede encontrar alguien interesado en el tema es la escasez de referencias directamente relacionadas con la EmT. Una de las conclusiones que se pueden sacar al repasar la literatura existente es que los problemas planteados por las diferentes propuestas de EmT son de naturaleza muy amplia, y van más allá de consideraciones específicas, poniendo en cuestión, en mayor o menor medida, un número de supuestos teóricos fundamentales en el campo de la evaluación de lenguas. El objetivo de este artículo es dar una perspectiva general sobre el tema, presentar y comentar algunas de estas propuestas, y considerar algunos de los problemas y reacciones que las mismas han suscitado desde algunos sectores del ámbito de la evaluación de idiomas.

## 2. ORÍGENES DE LA EmT

Bachman (1990) describe la existencia de dos enfoques diferentes en relación a cómo se define el dominio lingüístico a lo largo de la historia de la evaluación de lenguas. El autor los denomina enfoque de *interacción / habilidad* (IH) y enfoque de *mundo real* (MR). En el primero de los enfoques, el IH, la definición del *dominio lingüístico* viene dada por la descripción de las competencias que lo componen. En el ámbito de la lingüística aplicada (en la enseñanza de lenguas, por ejemplo) son muy conocidas las descripciones de estas competencias, que se pueden encontrar en marcos como los propuestos por Canale y Swain (1980)<sup>68</sup>, Canale (1983)<sup>69</sup>, Bachman (1990)<sup>70</sup> y Bachman y Palmer (1996). Lo que el enfoque IH hace es definir los componentes de la habilidad lingüística de tal forma que los diferentes componentes se puedan distinguir unos de otros; al hacerlo, los constructos de los diferentes componentes se pueden especificar de forma clara.

En contraste con este tipo de enfoque, al enfoque MR no le interesa tanto definir lo que es el dominio lingüístico, sino identificar un ámbito de uso real del lenguaje que se considere característico de la actuación lingüística de los usuarios competentes. El enfoque MR se centra en tratar de desarrollar pruebas para evaluar de forma directa la actuación, de modo que se pueda predecir la actuación y generalizarla de un contexto a otro. Se concentra en crear métodos directos de recogida de datos, como, por ejemplo, entrevistas orales. Su objetivo es predecir la actuación lingüística en situaciones restringidas, de tal forma que los formatos de pruebas usados puedan simplemente simular las situaciones deseadas de forma bastante precisa y fiel, usando condiciones de actuación realistas así como material basado en contextos que sean lo más parecidos posible a aquel que se encontrará el candidato en el mundo real. Este tipo de enfoque ha sido muy popular en el ámbito de las lenguas con fines específicos, así como en la evaluación para la orientación profesional.

Bachman (1990) insiste en la necesidad de que en ambos enfoques se definan de la forma más clara posible los aspectos sobre los que se basan: los ámbitos de la *vida real*, en el caso del MR; o las habilidades o componentes que forman el dominio lingüístico, en el caso del IH.

## 2. 1. Evaluación de la actuación lingüística

Las pruebas basadas en la actuación de un candidato son aquellas en las que “se requiere el uso integrado de destrezas lingüísticas y pragmáticas en una porción de discurso”, como las define McNamara (1997: 131). Aunque este tipo de pruebas están estrechamente ligadas a la definición de EmT, se han venido usando, en mayor o menor medida, durante los últimos 50 años<sup>71</sup> (Ibíd.). En los años 70, con la aparición del movimiento comunicativo en la enseñanza de idiomas, la evaluación de la actuación encontró argumentos de base en la teoría de la competencia comunicativa, y al igual que para algunos la enseñanza de idiomas mediante tareas se puede ver como la continuación lógica de la enseñanza comunicativa de idiomas, la EmT se puede ver

---

<sup>68</sup> Competencia gramatical, sociolingüística y estratégica.

<sup>69</sup> Competencia gramatical, sociolingüística, estratégica y discursiva.

<sup>70</sup> Competencia organizativa (gramatical y textual) y pragmática (ilocucionaria y sociolingüística).

<sup>71</sup> Por ejemplo, la entrevista oral de dominio lingüístico del Foreign Service Institute de los EEUU tras la segunda guerra mundial.

como la evolución natural de la evaluación de la actuación, aunque eso no implica que sean exactamente lo mismo.

En la última década, especialmente en Norteamérica, el interés en la evaluación de la actuación se ha centrado, en mayor medida, en el uso de tareas como instrumentos para provocar la actuación, y en cómo los candidatos actúan en tareas escritas u orales. McNamara (1997) sugiere dos razones para este interés en la evaluación de la actuación, pero en mi opinión estas razones son también válidas para explicar el interés por la EmT. La primera es el interés en armonizar las prácticas en el ámbito de la evaluación con el desarrollo de la investigación tanto en adquisición de segundas lenguas (ASL), como en enseñanza de idiomas. La segunda es la necesidad de responder a los requerimientos de usuarios o interesados específicos (por ejemplo, universidades, empleadores, etc.) de forma que los resultados puedan ser medibles, y se desarrollen procedimientos para grupos específicos de aprendices de L2. Por ejemplo, una empresa puede estar interesada en saber si los candidatos a una plaza como recepcionista son capaces de responder y mantener conversaciones relevantes para su puesto de trabajo, en un idioma determinado. La EmT, en este sentido, se puede definir como una "evaluación que utiliza tareas holísticas que implican conductas reales (o lo más cercano a éstas), o ponen en marcha el tipo de procesamiento lingüístico que se daría en actividades del mundo real" (Ellis, 2003: 285).

## 2.2. Tareas

El uso de tareas en diferentes ámbitos de la lingüística aplicada, especialmente en aquellos relacionados con la enseñanza de segundas lenguas y lenguas extranjeras, ha aumentado en los últimos años. El término *mediante tareas* se usa para referirse a enfoques basados en el uso de tareas como unidad central de planificación e instrucción en la enseñanza de idiomas, pero también en el diseño de programas o en la investigación en campos como la adquisición de segundas lenguas. De la misma forma, también han ido apareciendo los enfoques que abogan específicamente por el uso de tareas en la evaluación de lenguas. Pero si las tareas se ven como un principio en estos enfoques, para operativizarlos, y diseñar y desarrollar pruebas mediante tareas, la primera pregunta que hay que plantearse es ¿qué una tarea?

En la literatura existente sobre la investigación de ASL o la pedagogía de idiomas es difícil encontrar una única definición sobre qué es exactamente una tarea. Se han propuesto muchas definiciones de tarea en los últimos 20 años al menos<sup>72</sup>. Algunas de estas definiciones son muy amplias, e incluso se pueden referir a casi cualquier actividad que puede ser realizada por una persona. Tal perspectiva está representada por la definición de Long (1985; citada en Ellis, 2003), en la que el autor pone el énfasis en la relación con el mundo real para calificar una actividad como tarea:

Tarea es un trabajo llevado a cabo para uno mismo o para otros, de forma recompensada o no. Así, ejemplos de tareas incluirían pintar una valla, vestir a un niño, rellenar un formulario, comprar un par de zapatos, reservar un billete de avión, sacar un libro de una biblioteca, hacer el examen de conducir, escribir a

---

<sup>72</sup> El lector interesado puede consultar Ellis (2003) para una lista detallada.

máquina una carta, pesar a un paciente, ordenar cartas, atender una reserva de hotel, escribir un cheque, encontrar una calle, y ayudar a alguien a cruzar la calle. Dicho de otra forma, por "tarea" se entiende las mil y una cosas que *hace* la gente en la vida diaria, en el trabajo y fuera de él. "Tareas" son las cosas que la gente dirá que hace si se les pregunta, y no son lingüistas.

Más recientemente, se han propuesto definiciones más específicas, como la de Bygate *et al.* (2001):

Tarea es toda actividad que requiere que los aprendices usen la lengua, poniendo énfasis en el significado, para obtener un objetivo.

Una de las características interesantes de la propuesta de estos autores es la de afirmar que "las definiciones de tarea tendrán que ser diferentes para los diferentes propósitos para los que se usan dichas tareas" (ibíd.). En este sentido, Bygate *et al.* completarían su definición básica de tarea para adaptarla a un contexto evaluativo de la siguiente manera:

Una tarea es una actividad que requiere el uso del lenguaje por parte de los aprendices, con el énfasis en el sentido, para alcanzar un objetivo, y que es seleccionada de tal forma que es muy probable que proporcione a los aprendices información que les ayude a evaluar su propio aprendizaje.

Se desprende de esta definición que los autores están pensando más en una evaluación de tipo formativo que de tipo sumativo. En el campo de la evaluación de lenguas, como recuerda Chalhoub-Deville (2001: 211), el término tarea se ha conectado con frecuencia a la noción de método de prueba, en el sentido de formatos de pruebas (por ejemplo, pruebas de compleción, de selección múltiple, entrevistas orales, etc.). Bachman y Palmer (1996:44) también ofrecen su propia definición de tarea, "es una actividad que implica a los individuos en el uso de la lengua con el objetivo de alcanzar una particular situación".

Basándose en una revisión de la investigación y la literatura existente sobre tareas, Ellis (2003) presenta varias características de una tarea; estas son algunas de ellas:

- La correcta realización de una tarea debe implicar el uso del lenguaje.
- La atención de los aprendices debe centrarse fundamentalmente en la transmisión de un mensaje.
- Los participantes en la tarea deben actuar como "usuarios del lenguaje", es decir, deben emplear los mismos procesos comunicativos que en actividades en el mundo real.
- La tarea tiene que corresponderse con alguna actividad del mundo real para obtener autenticidad, aunque, a veces, una conducta lingüística comunicativa puede aparecer en tareas que no se corresponden a las del mundo real.

- La tarea implica el uso de alguna de las cuatro destrezas, aunque en la literatura sobre enseñanza mediante tareas, se pone el énfasis en el uso de destrezas orales.

- La tarea suele implicar la obtención de algún producto final, al margen del uso de la lengua, que pueda ser juzgado en función de su contenido.

### 3. DIFERENTES ENFOQUES EN LA EmT

Durante los últimos años se han propuesto diversos enfoques que se basan en el uso de tareas en la evaluación de idiomas. Estos enfoques, a veces, difieren en sus perspectivas teóricas. McNamara (1996; 1997) distingue entre dos tipos de evaluación de la actuación lingüística, clasificación que, en mi opinión, también es aplicable a la EmT: un enfoque *suave* (o enfoque *cognitivo / psicolingüístico*), y un enfoque *duro* (o enfoque *muestra de trabajo*).

En el enfoque *suave*, las tareas no necesitan necesariamente ser tareas del mundo real, sino que pueden ser más artificiales, y la actuación en esas tareas se juzga en términos de la habilidad lingüística manifestada. En el enfoque *duro*, las tareas representan tareas del mundo real y la actuación se juzga con criterios del mundo real.

#### 3.1. Enfoques cognitivo / psicolingüístico

En este primer tipo, la perspectiva de dominio lingüístico que subyace a la pruebas comunicativas, se considera un constructo. Un ejemplo de este enfoque es la propuesta de Skehan (1998), que propone considerar una teoría dual de competencia lingüística. En esta teoría dual, Skehan afirma que la competencia lingüística se compone de (1) expresiones formulaicas con el fin de conseguir acceso fácil y rápido durante la actuación lingüística, y (2) reglas que permiten una mayor precisión y creatividad pero que entrañan un mayor esfuerzo de procesamiento. En este enfoque, es importante conocer cómo estos dos sistemas, o su combinación, están influenciados por el tipo de tareas que se realizan, y qué aspectos diferentes de la actuación influyen a los diferentes tipos de tareas. Skehan identifica estos componentes de la actuación como fluidez, complejidad y corrección.

Skehan aboga por fundamentar la evaluación en mecanismos y procesos para que, de esta forma, se puedan realizar, en base a las pruebas, afirmaciones adecuadas sobre la naturaleza de la actuación y sobre cómo se fundamenta en la competencia (Ibíd.). El autor argumenta que si queremos que las deducciones que hagamos sobre las habilidades lingüísticas de un candidato y que estén basadas en su actuación durante las pruebas, sirvan para predecir futuras actuaciones y generalizar a contextos diferentes, es necesario que lleguemos a establecer cómo los diferentes tipos de tareas influyen en las actuaciones de los candidatos.

#### 3.2. Enfoques muestra de trabajo

Un buen ejemplo de este tipo de enfoques es la propuesta de un grupo de investigadores de la Universidad de Hawái en Manoa (Norris *et al.*, 1998; Brown *et al.*, 2002). Esta propuesta, como señala Bachman (2002:454), es probablemente el ejemplo mejor conceptualizado, operativizado, e investigado de los enfoques

evaluativos mediante tareas. Los autores proponen una distinción entre los enfoques mediante tareas y la evaluación de la actuación lingüística. Otro tipo de evaluación de la actuación puede utilizar sencillamente tareas particulares como medio para suscitar la actuación lingüística, que es luego evaluada de acuerdo a criterios que pueden no estar directamente relacionados con las tareas en sí mismas, del que sería un ejemplo la propuesta de Skehan. Tales pruebas aportan los instrumentos mediante los que se provocan conductas relacionadas con la habilidad o el conocimiento lingüístico, de tal forma que se puedan realizar interpretaciones sobre tales habilidades o conocimientos. Según estos investigadores, este tipo de evaluación de la actuación no se debería considerar mediante tareas, ya que ni las tareas del mundo real, ni las condiciones situacionales y de interacción asociadas a ellas, sirven como base para la evaluación de la actuación (Brown *et al.*, 2002). Los mismos autores comparten la definición que Long y Norris (2000) dan sobre EmT:

La EmT usa la tarea en sí misma como unidad de análisis fundamental (...). La EmT no solo utiliza las tareas de mundo real como medio para provocar el uso de componentes específicos del sistema lingüístico que son evaluados o medidos, sino que el constructo que interesa es la realización de la tarea en sí misma.

La actuación lingüística que se deriva de la realización de una tarea es evaluada, pero dicha actuación es un medio para recoger información sobre las habilidades de los estudiantes para llevar a cabo una tarea específica o un tipo de tareas. Vemos, pues, cómo en el caso de la EmT el interés se desplaza de la evaluación de las habilidades lingüísticas a la de las habilidades para llevar a cabo tareas específicas o tipos de tareas en las que la comunicación en la lengua meta es esencial. Long y Norris argumentan que la EmT puede ser apropiada en contextos educativos con objetivos relacionados con la lengua con fines profesionales o académicos, la enseñanza mediante contenido<sup>73</sup>, y la enseñanza mediante tareas. Los tipos de pruebas mediante tareas propuestos por Long y Norris (2000) requieren que los estudiantes se involucren en algún tipo de conducta que simule, de la forma más fiel posible, un uso de la lengua meta orientado a un objetivo fuera de las situaciones de la prueba de idioma. Las actuaciones en estas tareas se evalúan de acuerdo a (1) elementos de criterio predeterminado y relacionados con el mundo real (procesos y productos de la tarea), y (2) niveles de criterio (estándares auténticos relacionados con el éxito de la tarea).

La EmT también pone el énfasis en que la evaluación de la actuación de la tarea tenga en cuenta si los estudiantes son capaces, o no, o hasta qué punto lo son, de conseguir la tarea de acuerdo a elementos y expectativas de criterio del mundo-real. Como vimos anteriormente, otros tipos de evaluación de la actuación no se centran necesariamente en el logro de tareas específicas o tipos de tareas. Estas consideraciones se podrían aplicar a este ejemplo de tarea usada en su investigación por estos autores:

#### TAREA A. 1: ELEGIR UNA PELÍCULA

#### INSTRUCCIONES:

---

<sup>73</sup> El término inglés es *Content-Based Instruction*, y se refiere a un enfoque en el que la enseñanza del idioma está organizada en base a los contenidos no lingüísticos que los estudiantes van a aprender.

Lee la nota de tu amigo donde dice cuándo puede ir al cine y qué tipo de película le gustaría ver. Luego, escucha la lista de películas de tres cines diferentes. Presta atención a los horarios de las sesiones y los argumentos de las películas. Apunta los títulos y los horarios que parecen más apropiados. Ahora empareja las preferencias y horarios preferidos de tu amigo con tantas películas como creas posible. Llama a tu amigo y déjale un mensaje en el contestador dándole la información necesaria sobre algunas de las películas adecuadas. Por último, dile una película que te gusta y explica por qué.

**MATERIALES:**

Nota del amigo; lista grabada de las películas y horarios, con vocabulario de cine y posibles horarios; teléfono, mensaje del contestador del amigo.  
(Brown *et al.*, 1998)

#### 4. ALGUNAS CONSIDERACIONES

Uno de los grandes problemas, quizás el mayor, que se plantean al considerar estas propuestas, es el hecho de que el centro de la evaluación no sea un constructo lingüístico, sino la propia tarea. Este es el caso en el enfoque muestra de trabajo de EmT. Dicho enfoque no especifica un constructo subyacente y, así, permite una muy limitada generalización de la actuación en la prueba, es decir, disminuye la capacidad para realizar predicciones válidas sobre futuras actuaciones lingüísticas de los candidatos basadas en los resultados de las pruebas. Esto se relaciona con la imposibilidad de realizar inferencias sobre las habilidades de los candidatos, o como lo expresa Messick (1994), realizar inferencias de conductas observadas a constructos tales como conocimiento y destrezas subyacentes a esas conductas. Como resultado, es imposible alegar replicabilidad y generalizabilidad, es decir, la amplitud de situaciones que la actuación generaliza.

De igual forma, la cobertura de contenidos se ve como un problema en la EmT. Si el constructo es la tarea, y la habilidad de llevarla a cabo con éxito es lo que se evalúa, la tarea debería cubrir tantas destrezas o estrategias como fuera posible, que se necesitaran para completar tareas parecidas dentro del ámbito cubierto. Pero si miramos el ejemplo anterior (la tarea: elegir una película), uno se podría preguntar si la tarea está midiendo la habilidad del candidato para comprender el argumento de una determinada película o de cualquier película en general, o de comprender las notas de su amigo respecto a sus gustos sobre cine, u otro tipo de notas, etc. Esto, que puede parecer irrelevante, puede ser visto como un serio problema cuando uno quiere entender lo que significan las puntuaciones dadas a una actuación. Bachman (1990: 330) afirma que este tipo de enfoque “no puede proporcionar demasiada información sobre las habilidades lingüísticas, y por eso no puede demostrar la validez de sus inferencias”.

Otro problema que podríamos encontrar estaría relacionado con la fiabilidad. Bachman y Palmer (1996: 19) definen la fiabilidad como la “consistencia de medición”. Cuando la medición se basa en la realización de una tarea, parece legítimo hacerse la siguiente pregunta: ¿qué significa llevar a cabo con éxito una tarea? Los criterios para

evaluar el éxito de una tarea deberían basarse en juicios sobre lo que significa en el mundo real llevar a cabo con éxito una tarea. Pero, ¿cómo se pueden establecer estos criterios? ¿De quién serán los criterios establecidos? Si en una tarea, por ejemplo, los candidatos tuvieran que elegir, usando la lengua meta, de entre varios currículos al mejor candidato a un puesto de trabajo, ¿qué sería llevar a cabo con éxito la tarea?

Algunos nos hemos encontrado en alguna ocasión en la situación de ver cómo se seleccionaban, para determinados puestos, candidatos que parecían adecuados pero que con el tiempo han resultado ser poco idóneos para el puesto. ¿Las personas encargadas fracasaron en dicha tarea? ¿Cuál sería el criterio aplicado? Todas estas preguntas plantean algunos problemas básicos relacionados con la fiabilidad de este enfoque. Asumiendo que estos criterios se hayan establecido, el siguiente problema, aunque obviamente no sólo relacionado con la EmT, es el de los juicios de los evaluadores. Este tema ha sido identificado como una fuente de inconsistencia en medición y la única forma para asegurar la fiabilidad en este aspecto es la formación de los evaluadores, aunque Brindley (1994) señala que, en ocasiones, a pesar de que exista esta formación, las diferencias de juicio parece que permanecen. Otro problema respecto a la fiabilidad, que también afectaría a la validez, sugerido por Fulcher (1996; citado en Norris *et al.*, 1998: 60) es “la potencial falta de generalizabilidad de la actuación de una tarea a otra”.

Para tratar de minimizar estos problemas de fiabilidad, se pueden adoptar algunas medidas. El uso de triangulación al establecer los criterios y selección de las tareas. También el uso de herramientas estadísticas para identificar problemas relacionados con las decisiones de los evaluadores parece bastante efectivo, pero estas herramientas no siempre están al alcance de los profesionales, e incluso de las instituciones.

Otras consideraciones menores estarían relacionadas con aspectos como la autenticidad, la interactividad, el impacto y la viabilidad. La EmT trata de operativizar el concepto de autenticidad tratando de replicar muestras reales de uso lingüístico en situaciones no de examen, para obtener validez aparente y utilidad predictiva. Aunque Bachman (1990) afirma que “definir autenticidad al usar muestras de uso lingüístico de la vida real no permite reconocer qué partes de uso lingüístico dependen, por naturaleza, del contexto y, de esta forma, son únicas”. También se podría argumentar que una situación de examen no es una situación auténtica, pero se podría considerar auténtica como situación de examen. Algo similar ocurre en la enseñanza de idiomas, donde la clase en sí misma es una situación de la vida real, y como tal auténtica. En cuanto a la interactividad, Bachman y Palmer (1996) afirman que ésta depende de “cómo definimos el constructo “habilidad lingüística” para una situación determinada de examen”. Para los autores, la interactividad está relacionada con el grado de implicación de las características del candidato<sup>74</sup> en el proceso de realización de la tarea. Candidatos diferentes pueden procesar la tarea de diferentes formas. Un problema evidente es que en la propuesta del grupo de investigadores de Hawái, la lengua se requiere para llevar a cabo las tareas, pero no se da una definición de habilidad lingüística.

---

<sup>74</sup> Bachman y Palmer identifican características como la habilidad lingüística, el conocimiento del mundo (*topical knowledge*) y los patrones afectivos (*affective schemata*).



Bachman (1990) sugiere que la aplicación de este tipo de evaluación “se limite a situaciones en las que el criterio de uso de la lengua es relativamente homogéneo e identificable”, y propone su uso en el aula u en otro tipo de situaciones de evaluación de bajo impacto. Parece sensato admitir que uno de los aspectos positivos de la EmT es que puede proporcionar un potente y positivo efecto de rebote<sup>75</sup>. Si pensamos en la conexión de las tareas con la enseñanza de segundas lenguas y en cómo promueven la interacción comunicativa en el aprendizaje de L2, su uso tendría un impacto positivo en contextos educativos. Aunque podrían surgir algunas dudas sobre su viabilidad a la hora de implementar este tipo de pruebas. Los recursos requeridos en el diseño, desarrollo y uso de pruebas mediante tareas parecen ser muchos y diversos. El análisis de las necesidades de la población para identificar el ámbito de uso de la lengua meta (ULM), y el proceso de identificación de las tareas relacionadas con ese ámbito de ULM y la creación de sus equivalentes en tareas para la pruebas parece, al menos, largo y complicado. La complejidad entonces para identificar criterios para la evaluación de las diferentes tareas, con el fin de establecer un criterio claro de éxito al realizar la tarea, hace que el proceso de desarrollo del examen consuma mucho tiempo y sea difícil de implementar.

## 5. CONCLUSIONES

El hecho de usar tareas para evaluar, en mi opinión, no es en sí mismo un problema para obtener evidencia de validez. El uso de tareas como mecanismos diseñados para provocar muestras de habilidad lingüística que puedan ser usadas para inferir el grado de dominio lingüístico por parte de los candidatos, presenta problemas parecidos a los del uso de otro tipo de actividades. Sin embargo, se podrían expresar algunas objeciones cuando la tarea desplaza al constructo lingüístico, y ocupa su lugar.

Es importante recordar la existencia de diferentes enfoques a la EmT, a la hora de implementar este tipo de evaluación, ya que las calificaciones obtenidas en exámenes mediante tareas pueden ser distintas y del uso que se hiciera de ellas se desprenderían diferentes implicaciones. En cualquier caso, el desarrollo de mediciones explícitas de actuaciones en tareas es un problema complejo que también afecta a la fiabilidad de las puntuaciones obtenidas de estas mediciones y a la validez del uso de estas puntuaciones. En mi opinión se necesita más investigación para obtener más argumentos empíricos a favor o en contra del uso de la EmT. Asimismo, sería necesario que estos estudios cubrieran los diferentes enfoques a la EmT. Como recuerda Briendly (1994) parece sensato, en estos momentos, ver la EmT “como una inversión más a largo plazo que a corto plazo”.

También me parece sensato, como señala el grupo de Hawái (Brown *et al.*, 2002), que la EmT pueda desempeñar un papel junto a otro tipo de evaluación, dada la importancia del uso de diferentes fuentes de información para tomar decisiones, para seleccionar estrategias de evaluación e interpretar sus resultados. Bachman (2002) señala que la complejidad de la EmT nos está haciendo repensar muchos de los asuntos importantes sobre la evaluación de idiomas (qué evaluar y cómo, y qué argumentos y

---

<sup>75</sup> *Washback*, en terminología inglesa.

pruebas aportar para justificar las inferencias y decisiones que se tomen). Y en mi opinión nunca están de más propuestas que suponen un reto a la teoría y a la práctica establecidas.

## 6. BIBLIOGRAFÍA

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

Bachman, L.F. (2002): "Some reflections on task-based language performance assessment", *Language Testing* 19: 453-76.

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

Brindley, G. (1994): "Task-centred assessment in language learning: the promise and the challenge". En N. Bird *et al.* (eds.) (1994): *Language and learning*. Hong Kong: Hong Kong Education Department.

Brown, J. D. *et al.* (2002): *An investigation of second language task-based performance assessments*. Honolulu: University of Hawaii Press.

Bygate, M. *et al.* (eds.) (2001): *Researching pedagogic tasks: second language learning, teaching and testing*. New York: Pearson Education.

Canale, M. (1983): "On some dimensions of language proficiency". En Oller, J.W. (ed.) (1983): *Issues in language testing research*. Rowley: Newbury House.

Canale, M. & Swain, M. (1980): "Theoretical bases of communicative approaches to second language teaching and testing", *Applied Linguistics* 1: 1-47.

Chalhoub-Deville, M. (2001): "Task-based assessments: characteristics and validity evidence". En M. Bygate *et al.* (eds.) (2001): *Researching pedagogic tasks: second language learning, teaching and testing*. New York: Pearson Education.

Ellis, R. (2003): *Task-based language learning and teaching*. Oxford: Oxford University Press.

Fulcher, G. (1996): "Testing tasks: issues in task design and the oral group", *Language Testing* 13: 23-51.

Long, M.H. & Norris, J.M. (2000): "Task-based language teaching and assessment". En Bayram, M. (ed.) (2000): *Encyclopedia of language teaching*. London: Routledge.

McNamara, T. (1996): *Measuring second language performance: a new era in language testing*. New York: Longman.

McNamara, T. (1997): "Performance assessment". En Clapham, C. y Corson, D. (eds.) (1997): *Encyclopedia of Language and Education, Volume 7: language testing and assessment*. Dordrecht: Kluwer.

Messick, S. (1994): "The interplay of evidence and consequences in the validation of performance assessments", *Educational Researcher* 23: 13-23.

Norris, J.M. *et al.* (1998): *Designing second language performance assessments*. Honolulu: University of Hawaii Press.

Skehan, P. (1998): *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2001): "Tasks and language performance assessment". En M. Bygate *et al.* (eds.) (2001): *Researching pedagogic tasks: second language learning, teaching and testing*. New York: Pearson Education.