Documento elaborado por Pilar Franco Naranjo/Pedro Pascual Zahinos/Antonio Romero Rodríguez y basado en la documentación recibida en el curso Language Testing at Lancaster (2015) por la Dr. Green y demás profesores/as del curso y el curso Testing, Evaluation and Assessment en Norwich (2015).

Curso financiado por la acción KA1 Erasmus+

**CONCEPTOS CLAVE PARA ELABORACIÓN DE PRUEBAS DE EVALUACIÓN**

1. ***Principios básicos en la evaluación de una lengua extranjera:***
2. **Validez**

Un test es válido si hace lo que debe hacer, es decir, si mide lo que tiene que medir. Se han de tener en cuenta aspectos como:

* La claridad de los resultados.
* La facilidad con la que se pueden utilizar los resultados.
* Lo bien que esté relacionado el contenido con la población a la que se dirige el test.
* La validez de la teoría de la lengua o el concepto de las destrezas en la que se basa el test. Por ejemplo, para elaborar un test de comprensión auditiva habrá que basarse en un concepto que defina dicha destreza.
* Cómo se transforma la teoría de la lengua o el concepto de destreza en un test. Este aspecto está relacionado con el tipo de tareas que se eligen y lo que se requiere al examinando que haga con la lengua de estudio. En esta línea, los objetivos y las tareas del test se basarán en este concepto y es importante que el test mida el cumplimiento de estos objetivos que estarán basados bien en las microdestrezas o bien en los contenidos lingüísticos. De esta forma nos aseguraremos que el test mida lo que tiene que medir.
* Los aspectos que interfieren con el objetivo del test (por ejemplo, si para una pregunta de comprensión auditiva es necesario utilizar las matemáticas para resolver una cuenta.)
* Si el test representa bien la teoría de la lengua o el concepto de destreza y el nivel de la lengua que se quiere evaluar. Este aspecto está relacionado con el tipo de tareas que se eligen y lo que se requiere al examinando que haga con la lengua de estudio.
* Para saber si un test es válido o no se pueden recoger datos y analizarlos cuando se pilota y después de su aplicación. Además, se pueden pedir las especificaciones del test y comprobar si el test las cumple. En el documento de la bibliografía *Research notes Issue 59* (notas de investigación, ejemplar 59) viene un ejemplo en la página 5-7 (el artículo entero titulado Item analysis (análisis de datos) trata del tema). Si alguien está interesado en saber más se puede comentar en algunas de las sesiones en español, por ejemplo, en la del pilotaje.

1. **Fiabilidad**

Un test es fiable si mide algo de forma consistente para el mismo grupo de personas. Se han de tener en cuenta aspectos como:

* Si el resultado es un reflejo justo de la habilidad del examinando.
* Si el resultado refleja el nivel que se quiere examinar.
* Si el examinando mantendrá el nivel del que se examina en el futuro.
* Quien elabora el test tiene que procurar minimizar el error en el test (cuanto menos errores haya bien debidos al instrumento que se emplea, las instrucciones que se dan, el ambiente en el que se hace el test, etc. más fiable será el test).
* La fiabilidad está influida por:
  + El número de ítems: cuanto más largo sea el test más fiable será. Cuanto mayor sea el número de personas que hacen el test, más fiables serán los resultados que se puedan analizar para comprobar si es fiable.
  + Los ítems que discriminan: si son demasiado fáciles o difíciles, la fiabilidad será más baja. En el documento de la bibliografía *Research notes Issue 59* (notas de investigación, ejemplar 59) viene un ejemplo en la página 11-13 (el artículo entero titulado Test construction (elaboración de tests) trata del tema).
  + La homogeneidad de los ítems: en relación con el concepto de lengua o destreza en la que se basa el test.
  + La calidad de las tareas (concepto de lengua o destreza)
  + El número de tareas.
  + La elección de las tareas.
  + Cómo se administra el test.
  + La clave o la hoja de corrección de los examinadores.
  + La doble corrección.
  + El tiempo que se permite a los examinadores dedicar para la corrección.

1. **Practicabilidad**

Este punto tiene que ver con todos los aspectos prácticos del test en todas las etapas, desde su elaboración a su administración, corrección y retroalimentación. Algunos aspectos a tener en cuenta son:

* Los materiales (las fuentes de donde se toman los materiales, el nivel de dificultad, lo apropiado del tema, la tarea que se ha de realizar en relación con las destrezas, la velocidad del discurso en la audición, la longitud de palabras y/o minutos en la lectura o la audición, el número de versiones que se necesitan.
* La formación de todas las personas implicadas (redactores, interlocutores, correctores, administradores). Número de personas, cualificaciones requeridas, remuneración.
* Almacenamiento de los datos: en una plataforma etc. (seguridad, acceso, mantenimiento, materiales como tareas, pistas de audio, especificaciones etc.)
* Otros documentos del test (especificaciones, guías para las distintas personas implicadas, ejemplos de tareas con claves, justificaciones, transcripciones etc.
* Pilotaje (quién, cuando, cuántos, dónde, administradores, seguridad, puntuación, análisis estadísticos, revisión).
* Preparación del cuadernillo del test (selección de tareas, nivel de dificultad, método, número de ítems, en papel o en línea, diseño y estandarización de las instrucciones, numeración de páginas, visibilidad, color, grabación de audios, coste y distribución etc.
* Administración (lugar, acústica, luz, calor, ruido, distribución del aula, equipos de audio, seguridad)
* Retroalimentación: de los administradores del test, por escrito, oral; de las autoridades educativas, informes.
* Reclamaciones: procedimiento etc.
* Segunda convocatoria: número de test que se necesitan.
* Adaptación de los tests: personas con problemas de visión, dificultades de aprendizaje etc.

1. **Constructo o concepto:** los constructos son las habilidades de los alumnos que pensamos que subyacen en la realización del test, pero que no podemos observar directamente. Son conceptos que podemos identificar porque son normalmente nombres abstractos. Por ejemplo, la inteligencia, la actitud, la fluidez etc. Cuando se abstrae lo que se observa y se crean etiquetas, se está construyendo una teoría básica que explica fenómenos observables. En el caso de los tests para las lenguas extranjeras se utilizan modelos de la competencia comunicativa como el de Canale (1983) o el de Bachman (1990).