



# Bases para el diseño de escalas de evaluación de producción, interacción y mediación oral y escrita

Juan Lago Leis  
Cristina Rodríguez

III Jornada GIELE

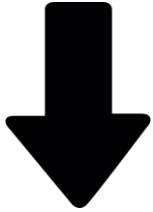
Donostia, 15 de noviembre de 2019

# El proceso de evaluación

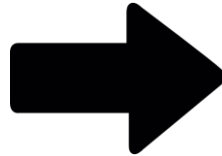


¿Qué son las escalas de evaluación?

**Herramientas**



**Criterios específicos**



**Categorías / parámetros**

# ¿Para qué sirven las escalas?

Nos permiten

**medir**

**producción/m  
ediación**



**recoger información**  
(intervención)



¿Qué aspectos son importantes?

**Fiabilidad**



**Precisión**

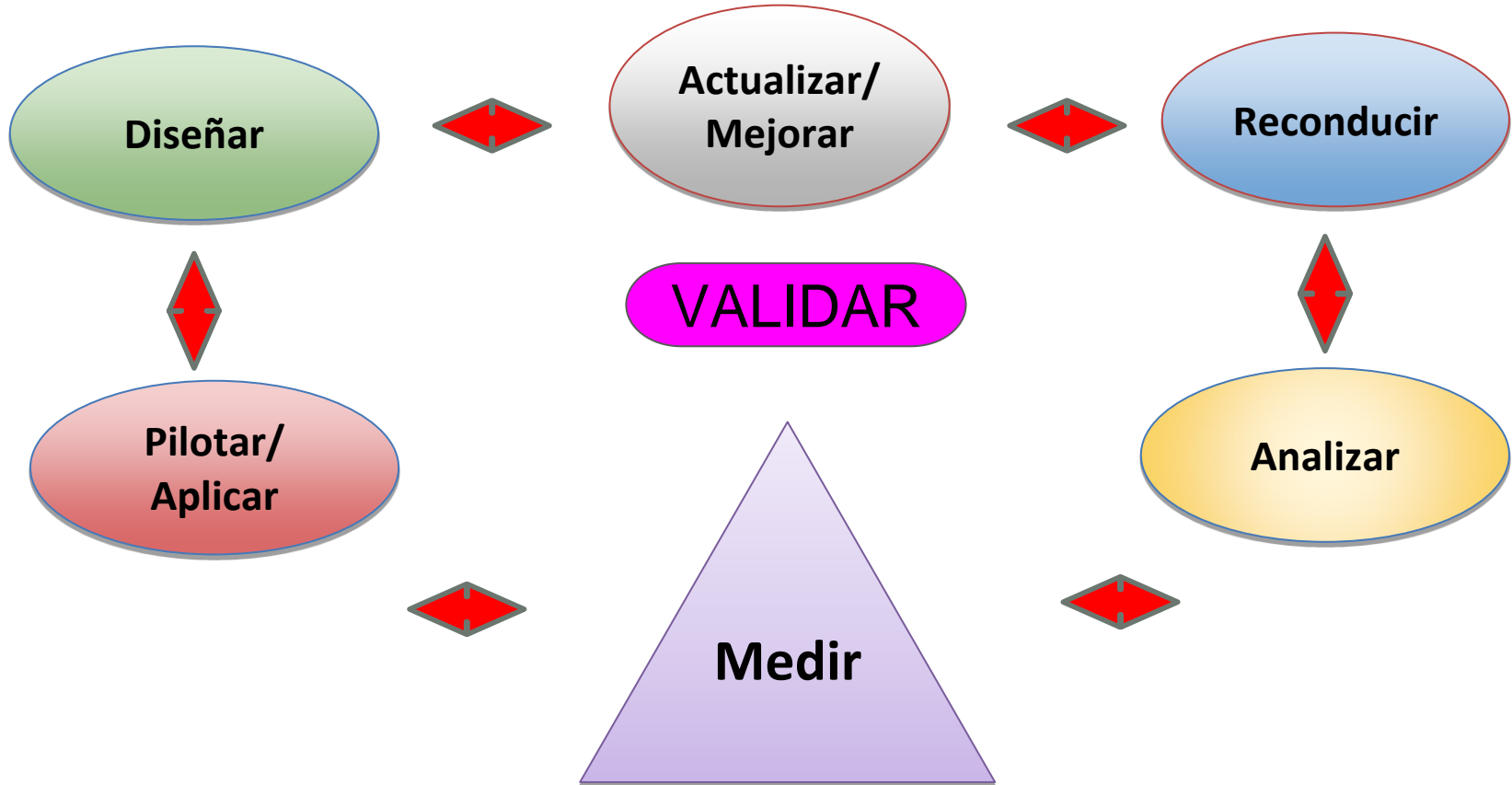
Calidad de los descriptores  
(**claros, precisos y observables**)



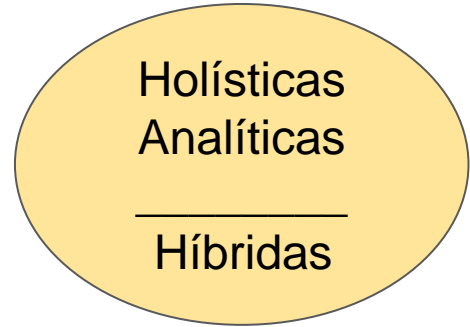
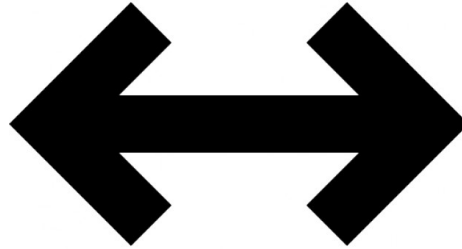
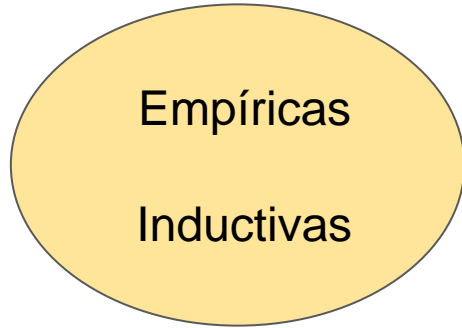
## ¿Por qué elaborar escalas de evaluación?

- Dotar al profesorado / evaluadores(-as) con unos **instrumentos fiables**.
- Crear herramientas para la **auto-evaluación**.
- Hacer un **seguimiento** pautado del proceso de enseñanza-aprendizaje.

# Documentos vivos



## Tipos de escalas









¿Qué metodología utilizar para elaborar escalas?

**Empírica**



**Inductiva/Intuitiva**

# Recomendaciones para elaborar escalas y redactar descriptores

-  Optar por un formato sencillo, claro y **fácil de usar** e interpretar.
-  Asegurarse de que los descriptores son claros, específicos y **observables**.
-  Las escalas son **documentos vivos**, han de estar actualizadas y al amparo de la normativa vigente.
-  Dejar **un espacio en blanco** donde el profesorado examinador pueda anotar comentarios y observaciones.

# Directrices para elaborar escalas y redactar descriptores



Cuanto más **detallado y preciso** sea el descriptor, más fiable y eficaz será la escala.

www.inec.es • 52019111



Una buena **especificación** del descriptor es, si cabe, más importante que la puntuación asignada a cada banda / categoría.



Es importante detallar el **rango de frecuencia** de los errores, las pausas, los deslices, las vacilaciones (esporádicos / sistemáticos, impiden o no la comunicación...).

www.inec.es • 52019111

# Directrices para elaborar escalas y redactar descriptores



www.inec.es + 32013111

Descriptores: organizados de una forma lógica y **gradual** en cada categoría / banda.



Las categorías tienen que estar **claramente** definidas (basadas en las tareas).



www.inec.es + 32013111

~~Ambigüedades~~ / Rasgos **subjetivos**



www.inec.es + 32013111

Errores / Deslices / Vacilaciones / Logros / Aciertos

+ / -

## Practical language testing de Glenn **Fulcher**

*The **purpose** is to define the quality of the language and its **communicative** effectiveness within the **boundaries** set out in the **specifications** (Fulcher, 2008a).*

# Criterios que recomienda el MCER para redactar descriptores



## - **Formulación positiva**

**El léxico es suficiente, pero limitado.**

**Tiene un repertorio básico de elementos lingüísticos.**

**Produce expresiones breves y habituales para satisfacer necesidades básicas concretas.**



## - **Precisión**

**Evitar expresiones del tipo: “una serie de”,**

**“unos pocos”, “estrategias adecuadas/apropiadas”.**

## Criterios que recomienda el MCER para redactar descriptores



### **Claridad**

**Sintaxis sencilla y estructura explícita y lógica**



### **Brevedad**

**Pronuncia de forma clara y natural. (B2)**



### **Independencia**

**Usados como criterios**

## Versiones de escalas

1. Escalas para evaluadores / profesorado  
(terminología específica y técnica / *jargon*)

1. Versión más sencilla para el alumnado



# Criterios que recomienda el MCER para redactar descriptores

## Cuestión de descripción

¿ Cómo redactarlos?

- **Libres de contexto** - permite un uso más generalizado (i)

*Una escala de uso común no debería crearse para un contexto muy específico (es de aplicación muy reducida)*

- Basada en **teorías / observación**
- Apropriados para la **función** que se usan en ese contexto (constructo/tareas) ¿Qué consigue **hacer**? ¿Qué debe **mejorar**?

Criterios que recomienda el MCER para redactar descriptores

## Cuestión de medición

¿Qué aspecto/s de la lengua pretenden medir?

Valores determinados **objetivamente**

## Más consejos para elaborar escalas de evaluación

3-5 bandas

Bandas intermedias

4-7 categorías

Primero definir la banda del **APTO** (nivel meta)

Agrupar las categorías

Usar las escalas para guiar y monitorizar el proceso de aprendizaje  
(el qué, el cómo, el dónde...)

¿Un checklist o una escala? ¿Por qué?

Un checklist: Sí/No. **Rápido** de aplicar pero no es fiable como una escala.

Analogía del interruptor

Las escalas son más **precisas**, indican el grado de dominio, la frecuencia, la competencia/subcompetencias, el repertorio, la corrección, c/c...  
(deben reflejar el constructo).

Sirven también para medir/evaluar las diferentes **fases** de aprendizaje y conocimiento.

# Fiabilidad de los descriptores

**Válidos para medir y evaluar** la calidad de una muestra

**Válidos** para sus respectivas **categorías**

**Válidos** para nuestro **constructo**

## Ejemplos:

Tiene pausas pero el discurso es fluido. (Fluidez)

Expone la información de forma efectiva/eficiente. (Cohesión)

Enlaza las ideas de forma lógica y eficaz. (Cohesión)

Expresa ideas y justifica opiniones de forma clara y comprensible, con detalles y ejemplos relevantes. (Eficacia comunicativa)

Usa el léxico seleccionado de forma consistente. (Mediación)

Utiliza un léxico sencillo para cumplir funciones habituales. (Corrección formal)

Se adapta a las necesidades de la persona destinataria. (Mediación)

# Escalas - holísticas y analíticas

- Dos tipos: holísticas y analíticas
- Las escalas reflejan la teoría de lengua que hay detrás de todo examen.  
(MCER - Constructo)

# Ventajas / Desventajas

<b>ESCALAS HOLÍSTICAS</b>	<b>ESCALAS ANALÍTICAS</b>
<ul style="list-style-type: none"><li>+ Más rápidas de usar</li><li>+ Más formación para evaluadores</li><li>+ Mejor fiabilidad inter-evaluador</li><li>- Efecto “halo”</li><li>- No hay información diagnóstica</li></ul>	<ul style="list-style-type: none"><li>- Más complejas de usar / requieren más tiempo</li><li>+ Más guiado para evaluadores, menos formación necesaria</li><li>+ Mejor fiabilidad intra-evaluador</li><li>+ Información diagnóstica para candidatos/as</li></ul>

# Categorías / Criterios

Las escalas analíticas se dividen en criterios, cada uno de los cuales puede agrupar varios aspectos similares.

¿Que categorías / criterios puede haber en una escala analítica?



# Categorías en la expresión escrita

Cumplimiento de la tarea: contenido, efecto comunicativo, registro, estilo, tipología de texto

Organización: estructura, cohesión, coherencia, párrafos, línea de argumentación, puntuación

Lengua: vocabulario, gramática, ortografía, corrección, riqueza / rango

Otros: longitud

# Categorías en la expresión oral

Cumplimiento de la tarea: contenido, efecto comunicativo, registro, adecuación

Organización: estructura, cohesión, coherencia, argumentación, desarrollo

Interacción: tomar la palabra, turnos de palabra, iniciar, mantener y acabar el discurso, gestión de la interacción

Lengua: vocabulario, gramática, corrección, riqueza / rango

Pronunciación, entonación, prosodia - (Nuevo descriptores Companion Volume)

Fluidez, cantidad de pausas, repeticiones, autocorrección, re-estructuración

# Escalas del MCER: Producción oral

## Escalas de producción oral:

- Expresión oral en general
- Monólogo sostenido: Descripción de experiencias
- Monólogo sostenido: Argumentación (por ej. en debate)
- Declaraciones públicas
- Hablar en público

## Estrategias de producción oral:

- Planificación
- Compensación
- Control y corrección

# Escalas del MCER: Interacción oral

## Escalas de interacción oral:

- Interacción oral en general
- Comprender a un interlocutor
- Conversación
- Conversación informal (con amigos)
- Conversación formal y reuniones de trabajo
- Colaborar para alcanzar un objetivo
- Interactuar para obtener bienes y servicios
- Intercambiar información
- Entrevistar y ser entrevistado

## Estrategias de interacción oral:

- Tomar la palabra (turnos de palabra)
- Cooperar
- Pedir aclaraciones

# Escalas del MCER: Competencias comunicativas

- Competencia lingüística general
  - Riqueza de vocabulario / Dominio del vocabulario
  - Corrección gramatical
  - Pronunciación
  - Competencia sociolingüística: Adecuación sociolingüística
- Competencia discursiva
  - Flexibilidad
  - Turnos de palabra
  - Desarrollo de descripciones y narraciones
  - Coherencia y cohesión
- Competencia funcional:
  - Fluidez
  - Precisión

# Decisiones

- ¿Qué criterios se seleccionan para cada nivel?
- ¿Se pueden agrupar criterios? ¿De qué forma?
- ¿Qué peso tendrá cada criterio en la puntuación final? ¿Igual o distinto peso?

# Baremos - Corrección gramatical

A2	B1	B2	C1
Utiliza algunas estructuras sencillas correctamente, pero sigue cometiendo errores básicos sistemáticamente; por ejemplo, suele confundir tiempos verbales y olvida mantener la concordancia; sin embargo, suele quedar claro lo que intenta decir.	Generalmente tiene un buen control gramatical, aunque con una influencia evidente de la lengua materna. Utiliza con razonable corrección un repertorio de «fórmulas» y estructuras habituales relacionadas con las situaciones más predecibles.	Buen control gramatical; todavía puede cometer «deslices» esporádicos, errores no sistemáticos y pequeños fallos en la estructura de la frase, pero son escasos y a menudo puede corregirlos retrospectivamente.	Mantiene un alto grado de corrección gramatical de modo consistente; los errores son escasos y apenas se notan.

# Baremos - Corrección

A2	B1	B2	C1
Generalmente queda claro lo que intenta decir.	Comete errores, pero queda claro lo que intenta expresar.	No comete errores que produzcan malentendidos.	Mantiene un alto grado de corrección gramatical de modo consistente; los errores son escasos y casi no se notan.

**Hay que valorar más el significado sobre la forma, y distinguir entre errores que impiden la comprensión de los que no la impiden.**



# Baremos – Turno de palabra

A2	B1	B2	C1
Es capaz de usar técnicas sencillas para iniciar, mantener o acabar una conversación breve.	Es capaz de iniciar, mantener y acabar conversaciones sencillas cara a cara sobre temas cotidianos o de interés personal.	Es capaz de iniciar el discurso, tomar el turno de palabra cuando es necesario y acabar la conversación cuando debe, aunque puede que no lo haga siempre con elegancia.	Es capaz de seleccionar la frase apropiada para iniciar sus comentarios con el fin de tomar la palabra o de ganar tiempo.

# Escala analítica

		Criterios / Categorías				
Bandas	Banda superior	Descriptoros	Descriptoros	Descriptoros	Descriptoros	Descriptoros
	Banda de apto	Descriptoros	Descriptoros	Descriptoros	Descriptoros	Descriptoros
	Banda de no apto	Descriptoros	Descriptoros	Descriptoros	Descriptoros	Descriptoros

# Escala analítica de nivel B1

		Criterios / Categorías				
Bandas	Banda superior 5	Descriptor B1+	Descriptor B1+	Descriptor B1+	Descriptor B1+	Descriptor B1+
	4					
	Banda de apto 3	Descriptor B1	Descriptor B1	Descriptor B1	Descriptor B1	Descriptor B1
	2					
	Banda de no apto 1	Descriptor A2	Descriptor A2	Descriptor A2	Descriptor A2	Descriptor A2

		Criterios / Categorías				
		CORRECCIÓN GRAMATICAL				
<b>Bandas</b>	Banda superior <b>5</b>	Se comunica con razonable corrección en situaciones cotidianas; generalmente tiene un buen control gramatical, aunque con una influencia evidente de la lengua materna. Comete errores, pero queda claro lo que intenta expresar.	Descriptor <b>B1+</b>	Descriptor <b>B1+</b>	Descriptor <b>B1+</b>	Descriptor <b>B1+</b>
	<b>4</b>					
	Banda de apto <b>3</b>	Utiliza con razonable corrección un repertorio de «fórmulas» y estructuras habituales relacionadas con las situaciones más predecibles.	Descriptor <b>B1</b>	Descriptor <b>B1</b>	Descriptor <b>B1</b>	Descriptor <b>B1</b>
	<b>2</b>					
	Banda de no apto <b>1</b>	Utiliza algunas estructuras sencillas correctamente, pero sigue cometiendo errores básicos sistemáticamente; por ejemplo, suele confundir tiempos verbales y olvida mantener la concordancia.	Descriptor <b>A2</b>	Descriptor <b>A2</b>	Descriptor <b>A2</b>	Descriptor <b>A2</b>

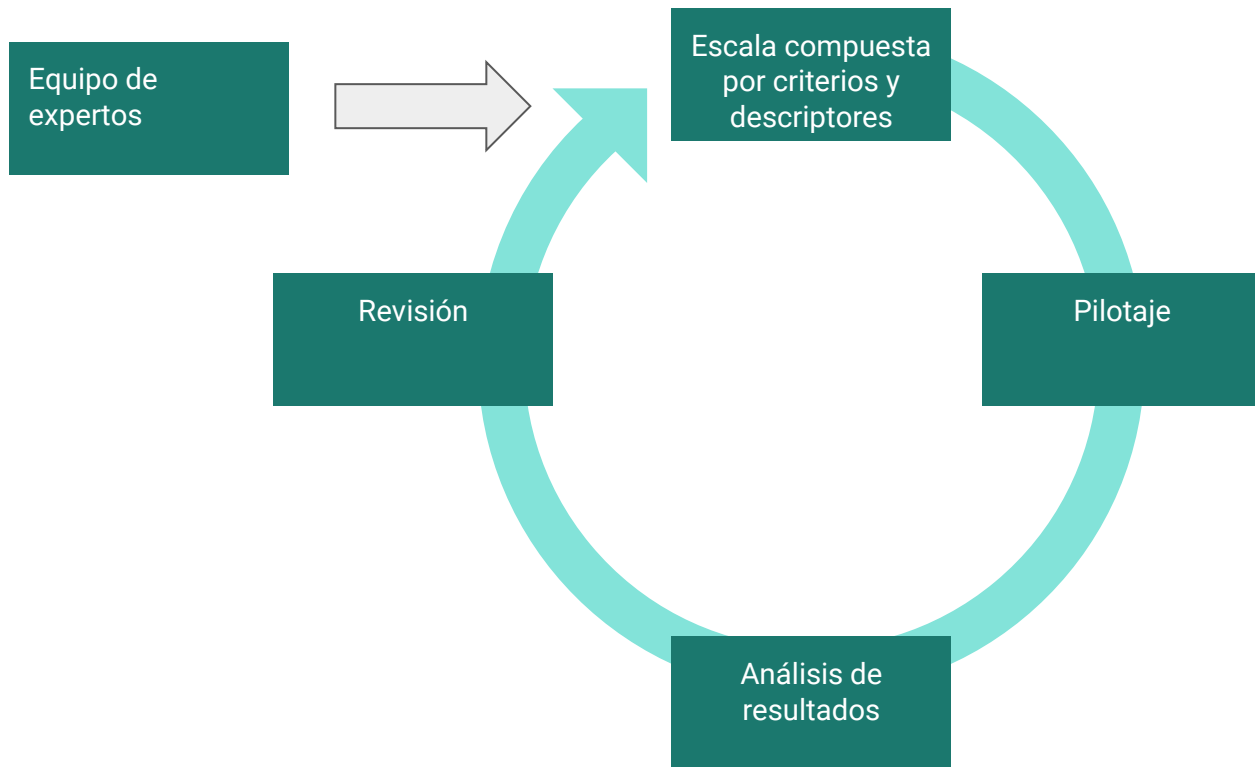
		<b>Criterios / Categorías</b>				
		<b>CORRECCIÓN GRAMATICAL</b>	<b>TURNOS DE PALABRA</b>			
Banda superior <b>5</b>	Se comunica con razonable corrección en situaciones cotidianas; generalmente tiene un buen control gramatical, aunque con una influencia evidente de la lengua materna. Comete errores, pero queda claro lo que intenta expresar.	Es capaz de intervenir en debates sobre temas cotidianos utilizando una frase apropiada para tomar la palabra.	Descriptor <b>B1+</b>	Descriptor <b>B1+</b>	Descriptor <b>B1+</b>	
<b>4</b>						
Banda de apto <b>3</b>	Utiliza con razonable corrección un repertorio de «fórmulas» y estructuras habituales relacionadas con las situaciones más predecibles.	Es capaz de iniciar, mantener y acabar conversaciones sencillas cara a cara sobre temas cotidianos o de interés personal. Es capaz de pedir a alguien que aclare o elabore lo que acaba de decir.	Descriptor <b>B1</b>	Descriptor <b>B1</b>	Descriptor <b>B1</b>	
<b>2</b>						
Banda de no apto <b>1</b>	Utiliza algunas estructuras sencillas correctamente, pero sigue cometiendo errores básicos sistemáticamente; por ejemplo, suele confundir tiempos verbales y olvida mantener la concordancia.	Es capaz de usar técnicas sencillas para iniciar, mantener o acabar una conversación breve. Inicia, mantiene y termina conversaciones sencillas cara a cara.	Descriptor <b>A2</b>	Descriptor <b>A2</b>	Descriptor <b>A2</b>	

		Criterios / Categorías				
		CORRECCIÓN GRAMATICAL	TURNOS DE PALABRA			
Banda superior <b>5</b>	Se comunica con razonable corrección en situaciones cotidianas; generalmente tiene un buen control gramatical, aunque con una influencia evidente de la lengua materna. Comete errores, pero queda claro lo que intenta expresar.	Es capaz de intervenir en debates sobre temas cotidianos utilizando una frase apropiada para tomar la palabra.	Descriptores <b>B1+</b>	Descriptores <b>B1+</b>	Descriptores <b>B1+</b>	
<b>4</b>						
Banda de apto <b>3</b>	Utiliza con razonable corrección un repertorio de «fórmulas» y estructuras habituales relacionadas con las situaciones más predecibles.	Es capaz de iniciar, mantener y acabar conversaciones sencillas cara a cara sobre temas cotidianos o de interés personal. Es capaz de pedir a alguien que aclare o elabore lo que acaba de decir.	Descriptores <b>B1</b>	Descriptores <b>B1</b>	Descriptores <b>B1</b>	
<b>2</b>						
Banda de no apto <b>1</b>	Utiliza algunas estructuras sencillas correctamente, pero sigue cometiendo errores básicos sistemáticamente; por ejemplo, suele confundir tiempos verbales y olvida mantener la concordancia.	Es capaz de usar técnicas sencillas para iniciar, mantener o acabar una conversación breve. Inicia, mantiene y termina conversaciones sencillas cara a cara.	Descriptores <b>A2</b>	Descriptores <b>A2</b>	Descriptores <b>A2</b>	

# Amenazas a la validez

- Irrelevancia al constructo
- Infra-representación del constructo
- Falta de transparencia de las decisiones

# Validación





# Validación

Equipo de  
expertos

“It is questionable whether any mark scheme can wholly capture the definition of a level in a way that examiners could reliably and consistently apply. The definition of a level is not captured merely on paper, but rather through the process of examiner training and standardisation.”

(Shaw y Weir, 2007, citado en Taylor y Galaczi 2011)

Análisis de  
resultados

# Validación

Formación

Familiarización

Estandarización

Muestras ilustrativas

Análisis cualitativo y cuantitativo

Feedback

# Escalas para evaluar la mediación

La evaluación de la mediación se asemeja a las tareas, ya que los dos:

- ocurren en un contexto, del que son dependientes
- tienen una finalidad semejante al de la vida real

Son procesos complejos con muchos elementos, y cada cual tendrá efecto en el resultado. Es muy difícil que todas tengan el mismo nivel de dificultad.

Aumentan en autenticidad, pero no se sabe si son muy “generalizables”.

# Evaluación “fuerte” o “débil”

McNamara (1996) distingue entre:

Evaluación “fuerte”, que se base en el cumplimiento de la tarea - la lengua es el medio por el que se consigue este fin.

Evaluación “débil”, en el que la evaluación se centra menos en la tarea en sí y más en la lengua producida por la persona candidata, la tarea siendo sólo un medio por el que se obtiene la producción.

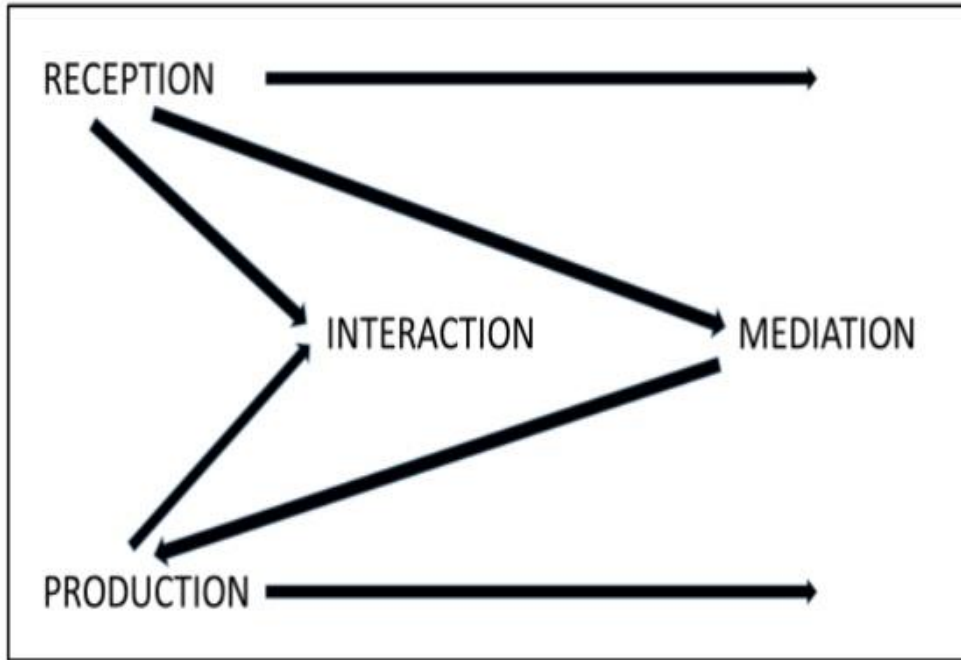
# Redefinición del constructo

Según Wigglesworth, esto lleva a dos formas de definir el constructo.

En la versión “débil”, el constructo se define como habilidad lingüística.

En la versión “fuerte”, incluye todo lo que pueda contribuir al éxito en el cumplimiento de la tarea.

## 4 Modos de comunicación / Actividades de lengua



Comprensión

Expresión (Producción)

Interacción (Co-producción)

Mediación

# Características de las tareas

- a) El significado es lo fundamental
- b) Los aprendices crean sus propios significados, no transmiten los de otros/as
- c) Hay una relación a actividades comparables del mundo real
- d) El cumplimiento de la tarea tiene prioridad
- e) La evaluación de la tarea está basada en el resultado

# Características de las tareas

- a) El significado es lo fundamental
- b) Los aprendices crean sus propios significados, no transmiten los de otros/as
- c) Hay una relación a actividades comparables del mundo real
- d) El cumplimiento de la tarea tiene prioridad
- e) La evaluación de la tarea está basada en el resultado

The assessment of the task is done in terms of the outcome.



# Muchas gracias por vuestra participación

Juan Lago Leis - [juanlagoleis@gmail.com](mailto:juanlagoleis@gmail.com)

Cristina Rodríguez - [crodriguez@eosisantiago.org](mailto:crodriguez@eosisantiago.org)

# Bibliografía

- Alderson, J. C. (1991). Bands and scores. In: J. C. Alderson & B. North (Eds.), *Language testing in the 1990s* (pp. 71–86). London: Macmillan
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: CUP / Council of Europe.
- Fulcher, Glenn. (2010). *Practical Language Testing*. 208-215. London: Hodder Education.
- Harsch, C., & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing*, 17(4), 228–250. doi: 10.1016/j.asw.2012.06.003
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Taylor, L., & Galaczi, E. (2011). Scoring validity. In L. Taylor (Ed.), *Examining speaking: Research and practice in assessing second language speaking* (pp. 171-233). Cambridge: Cambridge University Press.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wigglesworth, G. (2008). Task and performance based assessment. In S. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education (2<sup>nd</sup> Ed.)*, Volume 7: *Language testing and assessment* (pp.111-122). Springer Science+Business Media.