

CÓMO EVALUAR DE MANERA OBJETIVA AL ALUMNADO

José Muñiz

Universidad de Oviedo

Importancia de una evaluación rigurosa del alumnado

- **Nivel Científico-Profesional**
 - Disponer de indicadores fiables
 - Base del diagnóstico e intervención eficaz
- **Nivel Deontológico**
 - Derecho del alumnado a ser evaluado de forma objetiva y equitativa
 - Necesidad de formación del profesorado

Paradoja

- Todo el mundo parece saber cómo hay que evaluar a los alumnos
- Nadie (casi nadie) ha recibido formación sobre cómo hacerlo
 - Evaluar bien no es innato...

Objetivos

- Analizar la problemática implicada en la evaluación rigurosa y objetiva del alumnado
- Transversal: Estimular vuestra pasión evaluadora

Esquema

- Marco General de la evaluación universitaria
- Sistemas de evaluación del alumnado
- Perspectivas de futuro

Caso Práctico

- Un estudiante lleva a los tribunales a su profesora de *Didáctica de la Lengua* porque le suspendió habiendo obtenido un 4,5. Alega que la prueba carece de precisión, objetividad y de validez
- El Juez exige a la profesora que presente evidencias empíricas de la precisión, objetividad y validez del examen realizado.
- **¿Cómo se defiende la Profesora?**



Marco General de la Evaluación Universitaria

Hechos

- De repente en la Universidad Española todo el mundo se ha puesto a evaluar a todo el mundo a todas horas
- Pandemia Evaluadora

**¿Por qué esta obsesión
evaluadora?**

**Para Mejorar el sistema
Universitario**

Cadena de Mejora

- Evaluación rigurosa
- Diagnóstico preciso
- Intervención eficaz

¿Qué aspectos hay que mejorar?

- Investigación
- Docencia
- Gestión

Complejidad del Proceso Evaluador

- ¿Por dónde empezar a evaluar una universidad?
- ¿Qué evaluar?
- ¿Quién evalúa?
- ¿Qué se hace con los resultados?

.....



**SE NECESITA
UN MODELO INTEGRAL
DE EVALUACIÓN**

Modelo General de Evaluación Institucional

- A. Qué se evalúa
- B. Partes legítimamente implicadas
- C. Quién evalúa
- D. Cómo se evalúa: Metodología
- E. Feedback a las partes
- F. Planes de Mejora
- G. Opinión de las partes

¿Qué se evalúa en la Universidad?

- Individuos
- Productos
- Sistemas

Individuos

- Alumnos
- Profesores
- PAS
- Gestores
 - Rectores
 - Decanos
 - Directores...

Productos

- Proyectos de Investigación
- Papers
- Tesis
- Libros
- Curricula
- Planes de Estudios...

Sistemas

- Departamentos
- Facultades
- Institutos
- Grupos de investigación
- Másters
- Bibliotecas
- Universidades



**¿Qué es una buena
Universidad?**

Dos Parámetros básicos

- Calidad del Profesorado
- Calidad del Alumnado

Calidad del Profesorado

- **Contratar los mejores**
 - Limitaciones de la Universidad
 - ICREA
 - IKERBASQUE
 - POMPEU FABRA
- **Potenciar la variabilidad**
 - Sexenios *Ad libitum*

Calidad del alumnado

- Buena Selección
- Docencia de calidad
- Evaluación objetiva

Variables colaterales

- Ratio alumnos profesor
- Bibliotecas
- Facilidades informáticas
- Tamaño
- Gestión
- PAS
- Aspectos físicos: aulas, Laboratorios

....

The background of the slide features a serene landscape with a clear blue sky and a calm blue ocean meeting at a distant horizon. The text is overlaid on this background in a bright yellow color.

Evaluación de la Calidad

Rankings de Universidades

Shanghai Jiao Tong University Ranking

Es una de las clasificaciones más conocidas mundialmente, se trata de un listado recopilado por un grupo de especialistas en bibliometría de la Universidad Jiao Tong de Shanghai en China.

Criterios

- Número de galardonados con el Premio Nobel (10%)
- Ganadores de la Medalla Fields (20%)
- Número de investigadores altamente citados (20%)
- Número de artículos publicados en Science y Nature (20%)
- Impacto de los trabajos registrados en Science Citation Index (20%)
- Tamaño de la institución (10%)

Resultados Globales

1. Universidad Harvard
2. Universidad Stanford
3. Universidad de California, Berkeley
4. Universidad de Cambridge
5. Instituto Tecnológico de Massachusetts

Universidades de la *Ivy League*

- Brown (Rhode Island, 1764)
- Columbia (Nueva York, 1754)
- Cornell (Ithaca, Nueva York, 1865)
- Dartmouth (New Hampshire, 1769)
- Harvard (Massachusetts, 1636)
- Pensilvania (Filadelfia, 1740)
- Princeton (New Jersey, 1746)
- Yale (Connecticut, 1701)

Lecciones de los Rankings para la Evaluación Universitaria

- Están ahí, y van a seguir estando
- La gente los consulta
- Movilidad de los estudiantes
- Criterios simples, operativos y Públicos
- Dominio claro de EEUU
- En Europa: Inglaterra

¿Hacia qué Europa quiere converger España?

¿Y la Docencia qué?

- No puntúa en los Rankings
- Nos contratan por ello
- Nos pagan por ello
- Nos evalúan por ello
- Como al soldado el valor, al profesor se le supone su capacidad para enseñar
 - No se evalúa al contratar
 - Nadie nos enseña el oficio

Profesores Eficientes en Docencia

- Condición *sine qua non*:

**Saben: Dominan su campo
de especialización**

Profesores Eficientes en Docencia

- Actitud positiva ante los problemas
- Actitud Reflexiva
- Serenidad
- Optimismo
- Metodología eficaz
- Buen Clima en el aula
- Toma de decisiones flexible
- Establecen unos objetivos claros

Profesores Eficientes

- Contenido en pequeños pasos
- Mucha práctica
- Guían a los alumnos
- Corrigen y proporcionan *feedback*
- Fomentan el trabajo independiente
- Buscan aplicaciones de lo aprendido
- Usan síntesis periódicas y final
- Evalúan con rigor y precisión

Cuestionario Evaluación Profesores

- 1. Lleva las clases bien preparadas
- 2. Motiva a estudiar
- 3. Explica de forma clara y organizada
- 4. Me ayuda y orienta en el estudio
- 5. Responde a las dudas que le planteamos
- 6. Me anima a resolver las actividades con autonomía
- 7. Su sistema de evaluación es objetivo
- 8. Parece ilusionado con su trabajo como profesor
- 9. Entra y sale de clase a la hora establecida
- 10. En líneas generales, estoy contento con su labor como profesor

Evaluación de los alumnos: La Gran Ausente

- Necesidad de formación de los Profesores
 - Tecnología
 - Deontología
- Niveles de Evaluación
 - Académicos
 - Transversales

Niveles del conocimiento: Del Saber al hacer

- **Saber**
 - Conocimiento teórico
- **Saber cómo se hace**
 - Conocimiento práctico
- **Saber hacerlo**
 - Competencia
- **Hacerlo**
 - Desempeño

Jerarquía DIKW



Competencias

- Saber hacer algo de forma eficiente en un contexto determinado
 - Específicas
 - Transversales
- Definición académica
 - Gran variabilidad
 - Cano (2008): 18 definiciones



Competencias Transversales

Éxito Académico-Éxito Profesional

Regla General

- **Éxito Académico:**
Condición Necesaria, pero NO
Suficiente para el Éxito Profesional
- **Razón: Distintas Reglas de juego**

Fórmula del Rendimiento académico

Rendimiento = Capacidad x Esfuerzo

Otras variables implicadas

1. Características Personales

- Actitudes
- Personalidad
- Inteligencia Emocional

2. Pares

3. Profesor

4. Tipo de tarea

5. Centro

6. Ambiente familiar

7. Entorno Socio-cultural

Inteligencia Emocional: Los Techos de Cristal

**Nos ayuda a optimizar nuestro
rendimiento a partir de un esfuerzo y
de unas capacidades intelectuales
determinadas**

Facetas de la Inteligencia Emocional

- **Inteligencia INTRA-personal**
 - Autoconocimiento
 - Control Emocional
 - Motivación
- **Inteligencia INTER-personal**
 - Empatía
 - Habilidades Sociales

Cualidades Personales más valoradas por las empresas

1. Responsabilidad
2. Iniciativa
3. Capacidad Intelectual
4. Tenacidad
5. Creatividad
6. Ética
7. Lealtad
8. Seguridad en sí mismo
9. Discreción
10. Sinceridad

Habilidades Sociales más valoradas por las empresas

1. Trabajo en equipo
2. Adaptación
3. Ganas de aprender
4. Superación
5. Disponibilidad
6. Organización
7. Capacidad de comunicación
8. Integración
9. Capacidad para las relaciones sociales
10. Dotes de mando

Las cinco Aptitudes demandadas por Google

- Razonamiento analítico
- Capacidad de Comunicación
- Deseo de experimentar
- Saber trabajar en equipo
- Pasión y Liderazgo

Actitud emprendedora: Dimensiones

- Motivación de logro
- Toma de riesgos
- Innovación
- Autonomía
- Autoeficacia
- Tolerancia al estrés
- Locus de control
- Optimismo



**Evaluación
del Rendimiento de los
Estudiantes**

Evaluación rigurosa

- **Difícil de realizar**
 - Muchos factores implicados
- **Compleja**
 - Cada caso es distinto a los demás
 - Complejidad humana
- **Necesidad de personal especializado**
 - No se empieza de cero
 - Los errores queman el terreno
 - Basada en evidencias

Complejidad: Grandes números de lo Humano

- Vida: 3.000 millones de años
- 10 millones de especies
- Homo Sapiens: 150.000 años
- 20.000 genes
- 7.000 millones de habitantes
- 6.000 Lenguas
- 100.000 millones de neuronas
- 100 billones de conexiones

Leyendas Urbanas sobre la *condición Humana*

- Somos seres racionales
 - Kahneman, D. (2012). *Pensar rápido, pensar despacio*. Barcelona: Editorial Debate.
- Sólo utilizamos el 10% de nuestro cerebro
- La música de Mozart mejora la inteligencia de los bebés
- Se pueden aprender idiomas mientras se duerme

Leyendas Urbanas sobre la *condición Humana*

- Las personas opuestas se atraen
- Los hombres son de Marte y las mujeres de Venus
- Si a los niños se les educa igual funcionarán de forma parecida
- Los enfermos mentales son violentos

Programa a Evaluar

- **Explícitamente definido**
 - Programa
 - Sistema de Evaluación
- **Claro**
- **Objetivo**
- **Detallado**
- **Realista**
 - Ajustado al desarrollo docente
- **Referencias bibliográficas**
 - Papel de los Manuales: Grado y Postgrado

Norma ISO 10667: Evaluación de Personas en contextos Laborales

- Marco de Referencia reciente
- Evaluaciones basadas en evidencias
- Instrumentos de medida
 - Equitativos
 - Objetivos
 - Fiables
 - Válidos

Tipos de Pruebas

- Preguntas cortas
- Problemas
- Preguntas amplias
- Trabajos
- Casos Prácticos
- Evaluación *auténtica*
- Pruebas objetivas

¿Cuál es la más adecuada?

- Todas tienen pros y contras
- No todas valen para todo
- Combinación de varios tipos
 - Prueba objetiva
 - Ejercicio práctico
 - Trabajo tutorizado

Preguntas cortas

- **Definición**
 - Exigen una respuesta breve
- **Puntos fuertes**
 - Objetivas
 - Fáciles de corregir
 - Permiten muestrear los contenidos
- **Puntos débiles**
 - Memorísticas si no están bien formuladas

Problemas

- **Definición**
 - Resolver una situación planteada (cuantitativa o cualitativa)
- **Puntos fuertes**
 - Plantear situaciones novedosas
 - Corrección objetiva
- **Puntos débiles**
 - Formulación alambicada
 - Posible Capitalización del error
 - Dificultad para muestrear el contenido

Preguntas amplias

- Definición
 - Desarrollar una temática
- Puntos fuertes
 - Creatividad
 - Argumentación comprensiva
 - Vocabulario
 - Estilo
- Puntos débiles
 - Evaluación subjetiva
 - Muestreo del programa

Preguntas amplias

- **Variación**
 - Limitar la longitud
 - Orales
- **Sesgos de corrección**
 - Tipo de letra
 - Efecto de halo
 - Anclajes
 - Priming
- **Corrección**
 - Generar Instrucciones detalladas (*Rúbricas*) (PISA)
 - Corrección automatizada

Trabajos

- **Definición**
 - Desarrollo detallado de un objetivo
- **Puntos fuertes**
 - Madurez
 - Trabajo autónomo
 - Manejo de fuentes
 - Dominio de técnicas
- **Puntos débiles**
 - Corrección subjetiva
 - Tiempo de corrección y tutorización
 - Fraudes

Casos Prácticos

- Definición
 - Aplicación concreta del conocimiento
- Puntos fuertes
 - Evaluación del Conocimiento aplicado
 - Integración de conocimientos
- Puntos débiles
 - Dificultad de hacerlos *en vivo*
 - Validez de contenido

Evaluación *Auténtica*

- **Definición**
 - Trabajos realizados
- **Puntos fuertes**
 - Trabajo realmente desarrollado
- **Puntos débiles**
 - Objetividad de la evaluación
 - Tiempo de evaluación y tutorización
 - Discriminación socio-económica

Pruebas Objetivas

- **Definición**
 - Elección de la respuesta adecuada
- **Puntos fuertes**
 - Objetividad
 - Muestreo del programa
 - Corrección fácil y rápida
- **Puntos débiles**
 - Reconocimiento vs. Generación
 - Memorísticas si están mal formuladas

Tipos de Pruebas de Elección Múltiple

- Elección múltiple propiamente dicha
- Elección alternativa
- Verdadero-Falso
- Verdadero-Falso Múltiple
- Emparejamiento
- Elección múltiple compleja

Elección múltiple propriadamente dicha

- Estructura
 - Tronco o cuerpo
 - Respuesta correcta
 - Distractores
- Ejemplo: ¿Qué Universidad fue fundada por el inquisidor Valdés Salas?
 - Salamanca
 - Oviedo
 - Sevilla

Verdadero-Falso

- **Ventajas**
 - Fáciles de formular
 - Se adaptan a todo tipo de áreas
 - Fáciles de puntuar
 - Minimizan el tiempo de lectura
- **Inconvenientes**
 - Potencian el aprendizaje memorístico
 - Aciertos al azar 50%
 - Algo menos fiables que el formato clásico

Feedback

- Rápido
- Preciso
- Imprescindible



Desarrollo de Pruebas Objetivas

Contenido

1. Cada ítem debe reflejar un contenido específico y preciso
2. Mantener el contenido de cada ítem independiente del contenido de otros ítems del test.
3. Usar un vocabulario acorde con el grupo de estudiantes que están siendo evaluados

Estilo

4. Construir el ítem de forma vertical, no horizontal
5. Minimizar la cantidad de lectura en cada ítem

Redacción del enunciado

6. Asegurarse de que el sentido del enunciado resulta muy claro e inequívoco.
7. Incluir la idea central en el enunciado y no en las opciones.

Redacción del enunciado

8. Expresar el enunciado en términos positivos, evitando formas negativas tales como **NO** o **EXCEPTO**. Si usa términos negativos, hágalo con sumo cuidado y asegúrese que aparecen en mayúsculas o negritas

Redacción de las opciones

9. Escriba tantas opciones como pueda, aunque la investigación psicométrica sugiere que tres es lo adecuado.
10. Asegúrese que sólo una de esas opciones es la respuesta correcta.
11. Varíe la colocación de la respuesta correcta

Redacción de las opciones

12. Coloque las opciones en un orden lógico o numérico
13. Construya las opciones independientes entre sí, no deben solaparse
14. Mantenga a las opciones homogéneas en contenido y estructura gramatical

Redacción de las opciones

15. Escriba las opciones con una longitud aproximadamente igual.
16. La opción *Ninguna de las anteriores* debe ser usada con mucho cuidado, mejor no usarla
17. Evite la opción *Todas las anteriores*.
18. Escriba las opciones en términos positivos, evite negativas tales como NO.

Redacción de las opciones

19. Haga plausibles a todos los distractores

20. Use el humor con mucho cuidado, y sólo si es compatible con el profesor y con el ambiente de aprendizaje



Escalas tipo Likert

Medición de Actitudes

- Modelos:
 - Thurstone (1928)
 - Likert (1932)
- Ejemplos:
 - Aplicación a todos los ámbitos (Personas, Ideas, Instituciones...)

Ideas directrices para su construcción

- Seguir los pasos generales de construcción
- Afirmaciones
 - Cortas
 - Unívocas
 - Simples
 - Discriminativas
 - Efecto suelo
 - Efecto techo
 - Unidimensionalidad

Evaluación Integral

- Prueba Objetiva
- Ejercicio práctico
 - Problemas
 - Casos prácticos
- Trabajo tutorizado

Evaluación centrada en las competencias

- Centros de Evaluación
- Prácticum
- Simulaciones
- Portafolios
- Pruebas informatizadas
- Trabajos

Algunos errores habituales

- No valorar los conocimientos
 - Sin saber no hay competencia
 - La Universidad y el saber...
- Evaluaciones no discriminativas
 - Efecto suelo y Efecto techo
- Evaluaciones no objetivas
 - Inter-pares
 - Actitudes
 - Participación
 - Esfuerzo
 - Autoevaluación



Propiedades Métricas de los Instrumentos

Principales propiedades métricas

- **Fiabilidad (Precisión)**
 - ausencia de errores de medida
- **Validez**
 - mide lo que pretende medir
- **Calidad de los ítems**

Fiabilidad

- Concepto
- Estimación
 - Interjueces
 - Coeficiente de fiabilidad
 - Alfa
 - Otros coeficientes
- Fiabilidad y Longitud de la prueba
 - Fórmula de Spearman-Brown

Validez

- Concepto
- Tipos
 - Contenido
 - Predictiva
 - Constructo: Dimensionalidad

Puntos de Corte

- Concepto
- Tipos
 - Centrados en la materia
 - Centrados en las personas



Análisis de los Ítems

Tipos de Análisis

- Índice de Dificultad
- Índice de Discriminación
- Análisis de las alternativas incorrectas
- Número óptimo de alternativas
- Corrección del azar
- Sesgo
- Conocimiento parcial

Índice de Dificultad

- **Concepto**
 - Pruebas cognoscitivas
 - Escalas tipo Likert
- **Cálculo**
 - $ID = A/n$

Índice de Discriminación

- Concepto
- Cálculo
 - Correlación Ítem-Test (r_{ix})
 - Tipos de correlaciones

Análisis de Alternativas incorrectas

Ejemplo: N = 200

| | Alternativas | | | | |
|------------|--------------|----|----------|----|---|
| | A | B | <u>C</u> | D | E |
| Masters | 5 | 15 | 70 | 10 | 0 |
| No Masters | 15 | 65 | 20 | 0 | 0 |

Número Óptimo de Alternativas

- Número de alternativas y aciertos al azar
- Tres alternativas frente a cuatro y cinco

Corrección del azar

- Concepto
- Cálculo: $P = A - E/(n-1)$
 - P: puntuación corregida
 - A: número de aciertos
 - E: número de errores
 - n: número de opciones del ítem
- Alternativas a la fórmula canónica
 - Pesos especiales

Sesgo

- **Concepto**
- **Cálculo**
 - Métodos clásicos: Mantel-Haenszel
 - Teoría de Respuesta a los Ítems

Conocimiento Parcial

- **Concepto**
- **Formas de evaluarlo**
 - Juicios de probabilidad
 - Responder Hasta Acertar
 - Ponderación de las alternativas

Perspectivas de Futuro

- Evaluación Informatizada
 - Tele-evaluación
 - Items Multimedia
 - Ítems Interactivos
 - Gameficación
- Nuevo Marco Métrico
 - Teoría de Respuesta a los Ítems
 - Bancos de Ítems
 - Pruebas Adaptativas Informatizadas



Muchas gracias por vuestra atención

jmuniz@uniovi.es

Referencias Bibliográficas

- Carretero-Dios, H., y Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of Clinical and Health Psychology*, 5, 521-551.
- Downing, S. M. (2006). Twelve steps for effective test development. En S. M. Downing y T. M. Haladyna (Eds.), *Handbook of test development* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum Associates.
- Downing, S. M., y Haladyna, T. M. (2006). *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fernández, J. (2008). *Valoración de la calidad docente*. Madrid: Editorial Complutense.

Referencias Bibliográficas

- Morales, P., Urosa, B., y Blanco, A. B. (2003). *Construcción de escalas de actitudes tipo Likert*. Madrid: La Muralla.
- Moreno, R., Martínez, R., y Muñiz, J. (2006). New guidelines for developing multiple-choice items. *Methodology, 2*, 65-72.
- Muñiz, J. (1997a) Introducción a la teoría de respuesta a los ítems. Madrid: Pirámide.
- Muñiz, J. (1997b). Aspectos éticos y deontológicos de la evaluación psicológica. En A. Cordero (ed.), *La evaluación psicológica en el año 2000*. Madrid: Tea Ediciones.

Referencias Bibliográficas

- Muñiz, J. (2000). *Teoría Clásica de los Tests*. Madrid: Pirámide.
- Muñiz, J. (2004). La validación de los tests. *Metodología de las Ciencias del Comportamiento*, 5, 121-141.
- Muñiz, J., y Bartram, D. (2007). Improving international tests and testing. *European Psychologist*, 12, 206-219.
- Muñiz, J., Fidalgo, A. M., García-Cueto, E., Martínez, R., y Moreno, R. (2005). *Análisis de los ítems*. Madrid: La Muralla.
- Muñiz, J. y Fonseca, E. (2008). Construcción de instrumentos de medida para la evaluación universitaria. *Revista de Investigación en Educación*, 5, 13-25.