



# Análisis estadístico de las pruebas en moodle: Herramientas y técnicas para mejorar la evaluación



Fecha

**Julio E. Rodríguez Torres , Ph.D.**  
**UPR Río Piedras**  
**21 de octubre de 2024**





# Objetivos

- **Identificar las mejores prácticas al momento de analizar una prueba objetiva.**
- **Presentar algunos elementos sicométricos de las pruebas objetivas.**
- **Aplicar estos conceptos con un ejemplo de las estadísticas de análisis de ítems en Moodle.**
- **Identificar cómo utilizar esta información para mejorar las pruebas futuras.**





# Acceso al informe estadístico

Chrome File Edit View History Bookmarks People Tab Window Help Thu 2:22 PM

online.uprrp.edu/mod/quiz/view.php?id=1306609

Mis cursos Español - Internacional (es)

EDFU3013-001-2020S2

Participantes

Insignias

Competencias

Calificaciones

Información

Conceptos generales de medición, "assessment" y evaluación del aprendizaje.

Aspectos éticos en la evaluación del aprendizaje.

Redacción de objetivos instruccionales.

Elaboración y diseño de pruebas de aprovechamiento.

Tipos de pruebas.

**Examen parcial y final del curso.**

PCMAS.

Créditos

Página Principal

## EDFU3013-001-2020S2 - EVALUACION DEL APRENDIZ

[Página Principal](#) / [Mis cursos](#) / [EDFU3013-001-2020S2](#) / [Examen parcial y final del curso.](#) / [Examen parcial.](#)

### Examen parcial.

El examen parcial tiene un valor de treinta y cuatro (34) puntos. Seleccione la opción correcta en los ejercicios que se presentan.

Intentos permitidos: 1

Este cuestionario se cerró el miércoles, 3 de marzo de 2021, 14:30

Límite de tiempo: 1 hora 40 minutos

Intentos: 19

Volver al curso

↳ [Cómputo y descripción de índices de dificultad y discriminación.](#) (oculto) Ir a... [Materiales PCMAS](#)

https://online.uprrp.edu/mod/quiz/view.php?id=1306609#



# Reporte de estadísticas de examen

Número de primeros intentos completos calificados	19
Número total de intentos completados	19
Promedio de los primeros intentos	73.99%
Promedio de todos los intentos	73.99%
Calificación media de los últimos intentos	73.99%
Calificación media de los mejores intentos	73.99%
Mediana (de intentos con mejores calificaciones)	76.47%
Desviación estándar (para intentos con mejores calificaciones)	13.63%
Asimetría de la distribución de puntuaciones (para intentos con mejores calificaciones)	0.3668
Curtosis de la distribución de puntuaciones (para intentos con mejores calificaciones)	-0.8256
Coefficiente de consistencia interna (para intentos con mejores calificaciones)	50.69%
Ratio de error (para intentos con mejores calificaciones)	70.22%
Error estándar (para intentos con mejores calificaciones)	9.57



# Reporte de estadísticas de examen



El reporte en Moodle da un análisis estadístico (psicométrico) del examen (cuestionario), y de sus preguntas. La parte superior del reporte da un resumen del examen completo.

The screenshot shows a Moodle quiz report page for a partial exam. The page title is "Información sobre el cuestionario". It includes a download button for the full report as a CSV file. The main content is a table with 15 rows of statistical data. The left sidebar shows the course structure, with "Examen parcial y final del curso" highlighted in red. The browser's address bar shows the URL: `online.uprrp.edu/mod/quiz/report.php?id=1306609&mode=statistics`.

Información sobre el cuestionario	
Nombre del cuestionario	Examen parcial.
Nombre del curso	EDFU3013-001-2020S2 - EVALUACION DEL APRENDIZ
Abrir cuestionario	miércoles, 3 de marzo de 2021, 12:50
Cerrar cuestionario	miércoles, 3 de marzo de 2021, 14:30
Abierto por	1 hora 40 minutos
Número de primeros intentos completos calificados	19
Número total de intentos completados	19
Promedio de los primeros intentos	84.52%
Promedio de todos los intentos	84.52%
Calificación media de los últimos intentos	84.52%
Calificación media de los mejores intentos	84.52%
Mediana (de intentos con mejores calificaciones)	88.24%
Desviación estándar (para intentos con mejores calificaciones)	19.03%
Asimetría de la distribución de puntuaciones (para intentos con mejores calificaciones)	-1.9657
Curtosis de la distribución de puntuaciones (para intentos con mejores calificaciones)	4.9924
Coefficiente de consistencia interna (para intentos con mejores calificaciones)	84.11%
Ratio de error (para intentos con mejores calificaciones)	39.86%



# Estadísticas de Preguntas

- Índice de facilidad
- Desviación estándar
- Calificación aleatoria estimada
- Ponderación deseada
- Peso efectivo
- Índice de discriminación
- Eficiencia discriminatoria





# Estadísticas de Preguntas

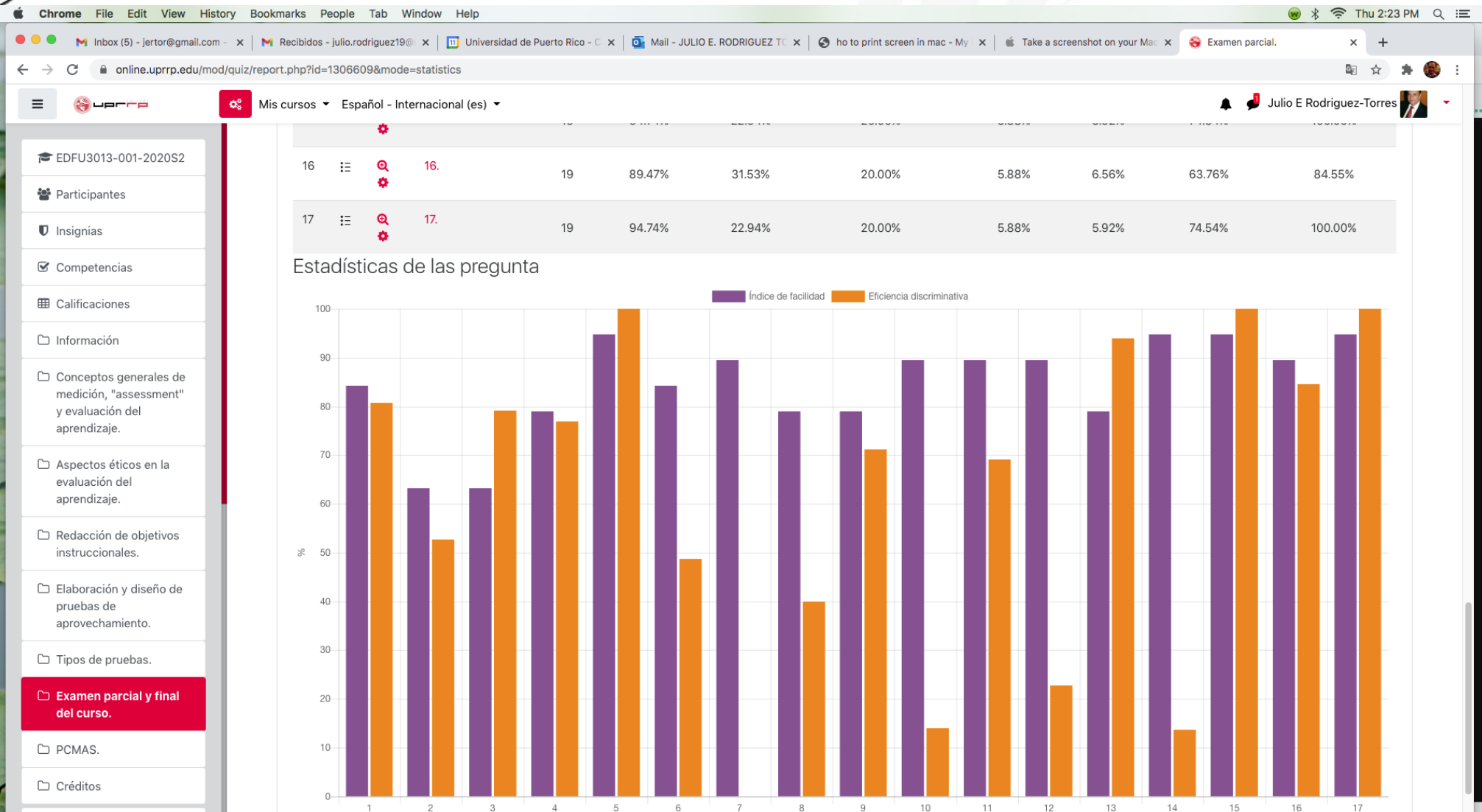
La siguiente sección da un análisis que muestra todas las preguntas en un formato de tabla. Hay ligas en esta sección, para editar preguntas individuales o entrar en un análisis detallado de una pregunta en particular.

Q#	Nombre de la pregunta	Intentos	Índice de facilidad	Desviación estándar	Calificación aleatoria estimada	Ponderación deseada	Peso efectivo	Índice de discriminación	Eficiencia discriminativa
1	1.	19	84.21%	37.46%	20.00%	5.88%	7.14%	62.42%	80.74%
2	2.	19	63.16%	49.56%	20.00%	5.88%	7.25%	41.73%	52.74%
3	3.	19	63.16%	49.56%	20.00%	5.88%	8.13%	58.15%	79.12%
4	4.	19	78.95%	41.89%	20.00%	5.88%	7.46%	59.49%	76.90%
5	5.	19	94.74%	22.94%	20.00%	5.88%	5.92%	74.54%	100.00%
6	6.	19	84.21%	37.46%	20.00%	5.88%	5.84%	36.49%	48.65%
7	7.	19	89.47%	31.53%	20.00%	5.88%	2.42%	-0.29%	-0.41%
8	8.	19	78.95%	41.89%	20.00%	5.88%	5.95%	31.52%	39.94%
9	9.	19	78.95%	41.89%	20.00%	5.88%	7.23%	54.68%	71.12%
10	10.	19	89.47%	31.53%	20.00%	5.88%	3.55%	10.77%	13.96%
11	11.	19	89.47%	31.53%	20.00%	5.88%	6.02%	51.49%	69.11%
12	12.	19	89.47%	31.53%	20.00%	5.88%	4.00%	16.39%	22.76%



# Información gráfica

La última parte de este reporte es una gráfica de barras del porcentaje de respuestas correctas (Índice de Facilidad) y del índice de Eficiencia Discriminatoria.







# Calificación promedio

**Es el promedio porcentual de todos los estudiantes. Se espera que en una prueba balanceada (mediana dificultad) una calificación promedio debe encontrarse entre 50% a 75%. Los valores fuera de estos límites requieren un análisis de las preguntas en términos de su facilidad o dificultad. Los exámenes con múltiples intentos invariablemente conducen a promedios más altos.**





## Mediana de la calificación

- **La mitad de los estudiantes tienen una calificación inferior a este valor. Bajo este valor se encuentra el 50% de los datos y sobre ese valor el otro 50%.**





# Desviación estándar

**Una medida de la dispersión de las calificaciones alrededor de la media. Intente obtener valores entre 12% y 18%. Valores inferiores a 12% sugieren que las calificaciones están demasiado amontonadas.**





# Sesgo

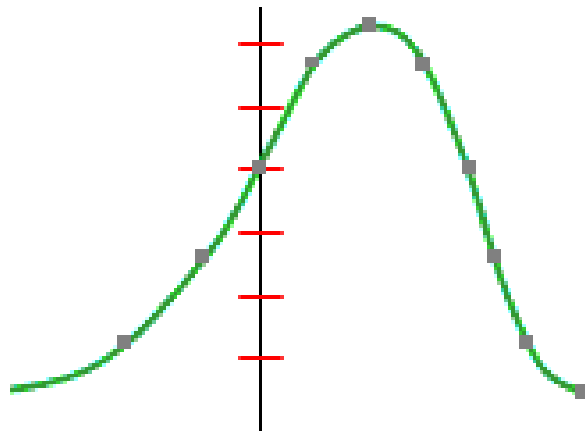
**Una medida de la asimetría de la distribución de calificaciones. Cero implica una distribución perfectamente simétrica, los valores positivos indican una 'cola' a la derecha y los valores negativos una 'cola' a la izquierda.**





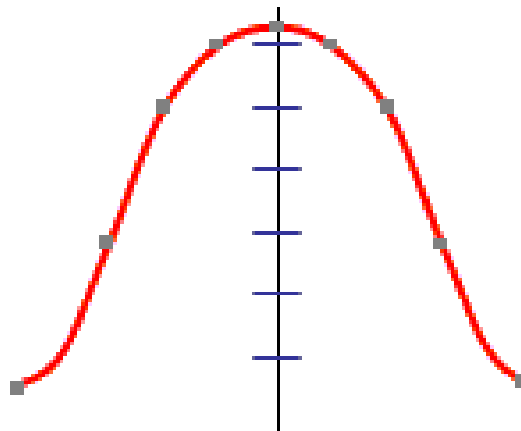
# Sesgo

**Curva de asimetría  
Negativa**



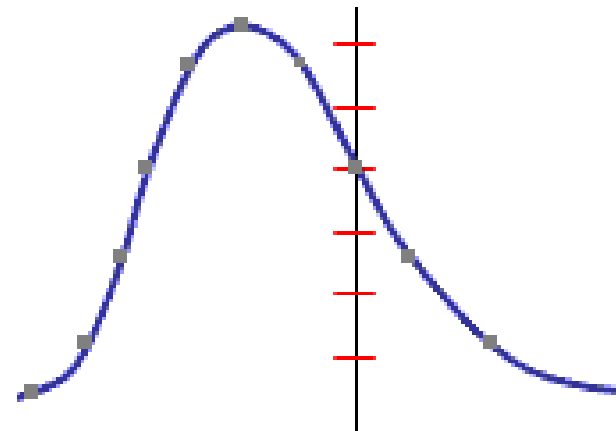
**Eje de simetría**

**Curva simétrica**



**Eje de simetría**

**Curva de asimetría  
Positiva**



**Eje de simetría**





## sesgo

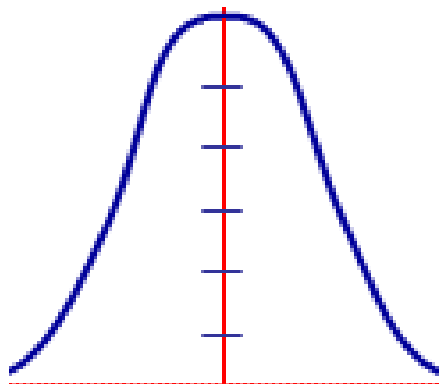
**Intente obtener un valor de sesgo de -1. Si fuera muy negativo, puede indicar falta de discriminación entre los estudiantes a los que les va mejor que al promedio. Similarmente, valores positivos muy grandes (mayores a 1) pueden indicar falta de discriminación cerca de la calificación umbral para pasar/reprobar.**



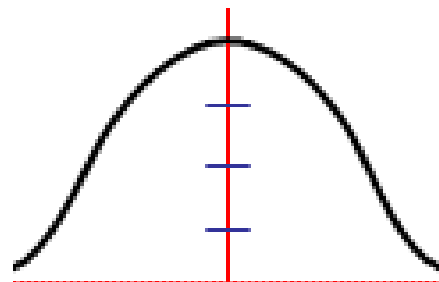


# Curtosis

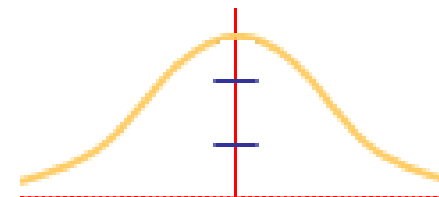
Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución. Por medio del *Coefficiente de Curtosis*, podemos identificar si existe una gran concentración de valores (*Leptocúrtica*), una concentración normal (*Mesocúrtica*) ó una baja concentración (*Platicúrtica*).



**Leptocúrtica**



**Mesocúrtica**



**Platicúrtica**





# Curtosis

**Trate de obtener un valor en el rango entre 0 y 1.  
Un valor mayor que 1 puede indicar que el examen  
no está discriminando muy bien entre los  
estudiantes de alta habilidad (o los de baja  
habilidad) y aquellos que son promedio.**







# Coeficiente de consistencia interna

Es muy difícil obtener consistencia interna mayor al 90%. Cualquier valor arriba de 75% es satisfactorio. Si el valor es inferior a 64%, el examen completo no es muy confiable y debería revisarse. Un valor bajo indica, ya sea que algunas de las preguntas no son muy buenas para discriminar entre estudiantes de diferente habilidad y por esto las diferencias entre las puntuaciones totales están en gran medida asociadas al azar; o puede indicar que algunas de las preguntas están evaluando un rasgo diferente que el resto de las preguntas y que el examen en su conjunto no es muy homogéneo midiendo el rasgo. Cuando los estudiantes tienden a adivinar puede disminuir este coeficiente.





# Tasa de error

**Este estima el porcentaje de la desviación estándar que se debe a efectos aleatorios en lugar de diferencias genuinas de la habilidad entre los estudiantes. Valores de tasa de error superiores al 50% no pueden considerarse satisfactorios; implican que menos de la mitad de la desviación estándar se debe a diferencias en habilidad y que el resto son efectos aleatorios.**





# Tasa de error

Esta está relacionada con el coeficiente de consistencia interna de acuerdo con la tabla siguiente:

<b>Coeficiente de Consistencia Interna</b>	<b>100</b>	<b>99</b>	<b>96</b>	<b>91</b>	<b>84</b>	<b>75</b>	<b>64</b>	<b>51</b>
<b>Tasa de Error</b>	<b>0</b>	<b>10</b>	<b>20</b>	<b>30</b>	<b>40</b>	<b>50</b>	<b>60</b>	<b>70</b>





# Error estándar

**Este es la desviación estándar multiplicada por la tasa de error y dividida entre 100. Estima qué tanto de la desviación estándar se debe a efectos aleatorios, y es una medida de la incertidumbre en la calificación de cualquier estudiante dado. Si el mismo estudiante tomó otro exámen equivalente, su calificación se esperaría que estuviera dentro de más menos un error estándar de la calificación anterior.**





# Puntuación en una prueba

$$X_O = X_R + E$$

$X_O$  = Valor observado

$X_R$  = Valor real

$E$  = Componente de error





# interpretación del Error estándar

- **Cuanto menor sea el valor del error estándar, mejor es el examen, pero es difícil lograr que el error estándar sea inferior al 5% o 6%.**
- **Si el error estándar excediera de 8%, es probable que una proporción sustancial de los estudiantes estén erróneamente calificados, en el sentido que las calificaciones otorgadas no indican con exactitud sus habilidades verdaderas.**





# Estadísticas de Preguntas

- Índice de facilidad
- Desviación estándar
- Calificación aleatoria estimada
- Ponderación deseada
- Peso efectivo
- Índice de discriminación
- Eficiencia discriminatoria





# Índice de facilidad

La puntuación promedio de los estudiantes en el ítem.

Índice de Facilidad	Interpretación
5% o menos	Extremadamente difícil, o algo está mal con la pregunta.
6% - 10%	Muy difícil.
11% - 20%	Difícil.
21% - 34%	Moderadamente difícil.
35% - 65%	Correcta para el estudiante promedio.
66% - 80%	Bastante fácil.
81% - 89%	Fácil.
90% - 94%	Muy fácil.
95% - 100%	Extremadamente fácil.





# Estadísticas de Preguntas

La siguiente sección da un análisis que muestra todas las preguntas en un formato de tabla. Hay ligas en esta sección, para editar preguntas individuales o entrar en un análisis detallado de una pregunta en particular.

Q#	Nombre de la pregunta	Intentos	Índice de facilidad	Desviación estándar	Calificación aleatoria estimada	Ponderación deseada	Peso efectivo	Índice de discriminación	Eficiencia discriminativa
1	1.	19	84.21%	37.46%	20.00%	5.88%	7.14%	62.42%	80.74%
2	2.	19	63.16%	49.56%	20.00%	5.88%	7.25%	41.73%	52.74%
3	3.	19	63.16%	49.56%	20.00%	5.88%	8.13%	58.15%	79.12%
4	4.	19	78.95%	41.89%	20.00%	5.88%	7.46%	59.49%	76.90%
5	5.	19	94.74%	22.94%	20.00%	5.88%	5.92%	74.54%	100.00%
6	6.	19	84.21%	37.46%	20.00%	5.88%	5.84%	36.49%	48.65%
7	7.	19	89.47%	31.53%	20.00%	5.88%	2.42%	-0.29%	-0.41%
8	8.	19	78.95%	41.89%	20.00%	5.88%	5.95%	31.52%	39.94%
9	9.	19	78.95%	41.89%	20.00%	5.88%	7.23%	54.68%	71.12%
10	10.	19	89.47%	31.53%	20.00%	5.88%	3.55%	10.77%	13.96%
11	11.	19	89.47%	31.53%	20.00%	5.88%	6.02%	51.49%	69.11%
12	12.	19	89.47%	31.53%	20.00%	5.88%	4.00%	16.39%	22.76%





# Desviación estándar

Una medida de la dispersión de calificaciones respecto a la media y por tanto, la magnitud de cuánto puede discriminar la pregunta. Si el índice de facilidad fuera muy alto, o fuera muy bajo, es imposible que la dispersión sea grande. Sin embargo, una buena desviación estándar no asegura automáticamente una buena discriminación. Un valor de la desviación estándar inferior a 33% en la tabla anterior *generalmente* es insatisfactorio.





# Calificación aleatoria estimada

**Este es el promedio de calificación que se esperaría que los estudiantes obtuvieran por una adivinanza aleatoria de la pregunta. Los puntajes por adivinanza aleatoria solamente están disponibles para las preguntas que usan algún tipo de opción múltiple.**

**Valores arriba del 40% son insatisfactorios, y demuestran que las preguntas de Falso/Verdadero deben emplearse muy mesuradamente (poco o nada) en exámenes sumativos.**





# Ponderación deseada

**La ponderación o peso de la pregunta, expresado como un porcentaje del puntaje general.**





# Índice de discriminación

Indica qué tan efectiva es la pregunta para clasificar/separar/discernir a los estudiantes más capaces de los menos capaces. Los resultados deberían de interpretarse como sigue:

Índice	Interpretación
50% y superior	Muy buena discriminación.
30% – 50%	Adecuada discriminación.
20% - 29%	Débil discriminación.
0 - 19%	Muy débil discriminación.
valores negativos	La pregunta probablemente sea inválida.



# Eficiencia de discriminación

Esta estadística intenta estimar qué tan bueno es el índice de discriminación en relación con la dificultad de la pregunta.

- Un ítem que sea muy fácil o muy difícil no puede discriminar entre los estudiantes de habilidades diferentes, debido a que la mayoría de ellos obtendrán el mismo puntaje para esta pregunta.
- La discriminación máxima requiere un índice de facilidad que esté en el rango del 30% al 70% (a pesar de que este valor no es garantía de obtener un alto índice de discriminación).
- La eficiencia de discriminación muy rara vez se acercará al 100%, pero deberían poder obtenerse valores superiores al 50%.

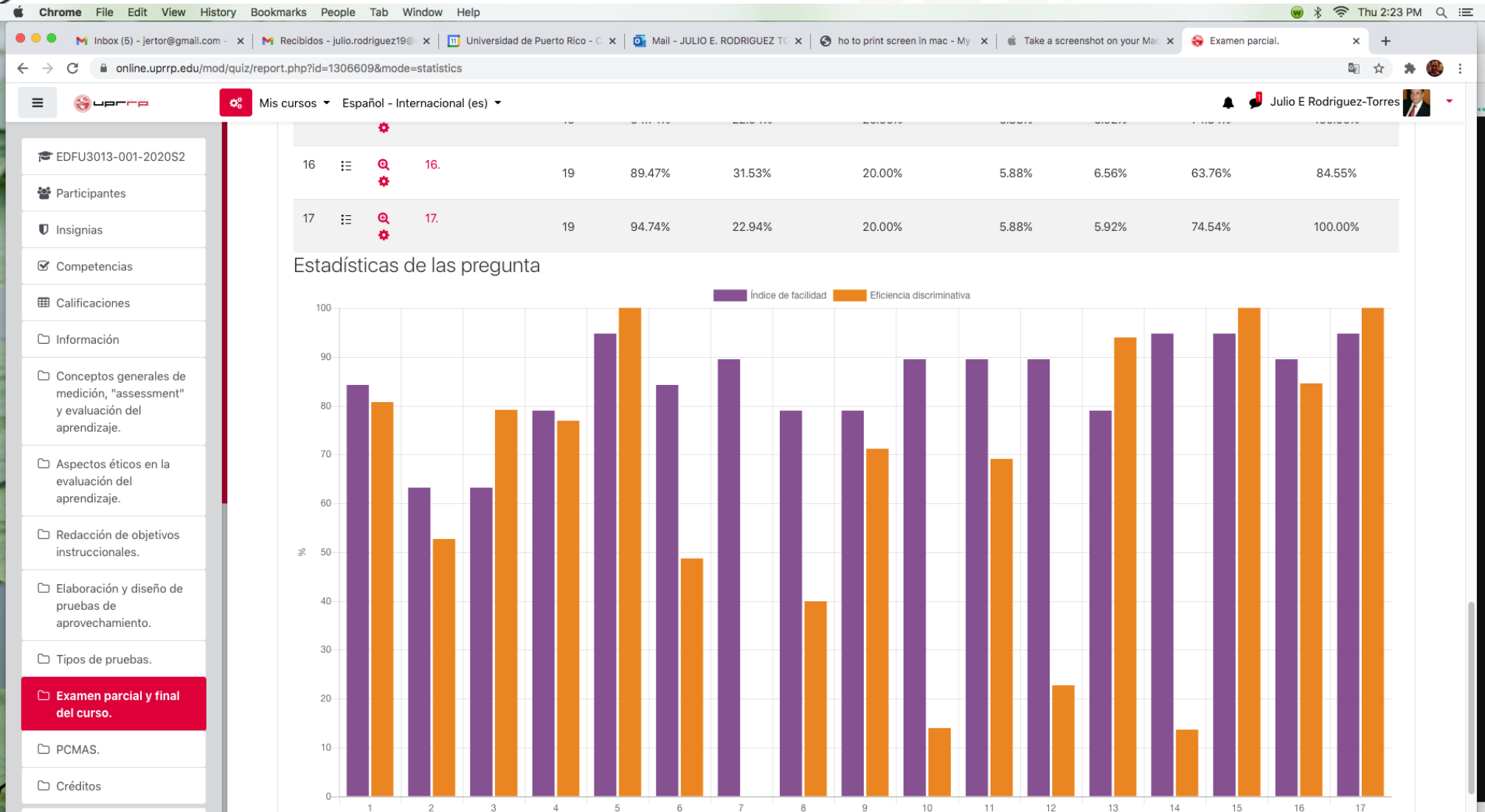
Valores inferiores al 50% y superior	Muy buena discriminación.	que la pregunta discrimina entre estudiantes de diferentes niveles de habilidad de manera muy buena.
30% – 50%	Adecuada discriminación.	
20% - 29%	Débil discriminación.	
0 - 19%	Muy débil discriminación.	
valores negativos	La pregunta probablemente sea inválida.	





# Información gráfica

La última parte de este reporte es una gráfica de barras del porcentaje de respuestas correctas (Índice de Facilidad) y del índice de Eficiencia Discriminatoria.





# ¿Preguntas?

## Referencias:

- R L Thorndike 2004, *Measurement and Evaluation in Psychology and Education (Seventh edition)*, Prentice Hall.
- This looks like it might be worth getting (previous edition cited by J J Barnard):
- L Crocker & J Algina 2006, *Introduction to Classical and Modern Test Theory*, Wadsworth Pub Co.

