



¿Existe sesgo según el sexo en la prueba diagnóstica EDICH?¹

Is there a sex-based bias in the EDICH diagnostic test?

Existe um viés segundo sexo em a prova diagnóstica EDICH?

Eliana Rodríguez.
ORCID: 0000-0001-8018-9242¹

Marcos Davyt.
ORCID: 0000-0001-6230-3654²

Joaquín Hurtado.
ORCID: 0000-0003-1483-7947³

Andrés Abella.
ORCID: 0000-0001-7453-3561⁴

¹ Unidad Académica de Laboratorios Prácticos, Facultad de Ciencias, Universidad de la República, Uruguay.

Contacto: eliana@fcien.edu.uy

² Unidad Académica de Laboratorios Prácticos, Facultad de Ciencias, Universidad de la República, Uruguay.

Contacto: mdavyt@fcien.edu.uy

³ Unidad Académica de Laboratorios Prácticos, Facultad de Ciencias, Universidad de la República, Uruguay.

Contacto: jhurtado@fcien.edu.uy

⁴ Centro de Matemática, Facultad de Ciencias, Universidad de la República, Uruguay.

Contacto: andres@cmat.edu.uy

Recibido: 26-02-24

Aceptado: 08-10-24

Resumen

En este artículo se propuso analizar si existen diferencias significativas entre el rendimiento de estudiantes hombres y mujeres en la prueba de Evaluación Diagnóstica de Conocimientos y Habilidades (EDICH), que es la única evaluación común a todas las licenciaturas de la Facultad de Ciencias. Del análisis de dicha prueba se concluyó que en general sus preguntas exhiben independencia del sexo. Las únicas excepciones se encontraron en algunas instancias del área física, en las cuales las mujeres obtuvieron calificaciones significativamente más bajas de lo que se esperaría si los rendimientos fuesen independientes. Se profundizó en este análisis y se observó un sesgo a favor de los hombres, lo cual podría explicar las diferencias halladas.

Palabras clave: sesgo según sexo, evaluaciones universitarias.

Abstract

In this work we sought to analyze if there were significant differences between men's and women's achievements in the Diagnostic Evaluation of Knowledge and Skills (EDICH), since this is the only test that students of all careers of the Faculty of Science have in common. From the analysis of this test we conclude that in general questions display independence of sex. The only exceptions were found in some cases in Physics, where women obtained significantly lower scores than would be expected if achievements were independent. This analysis was broadened and a bias in favor of men was observed in the Physics questions posed in the test, which could explain the differences we observed.

Keywords: sex-based bias, university evaluations.

1 Artículo aprobado por el equipo editor: Carolina Cabrera y Nancy Peré.

Resumo

Neste trabalho foi proposto analisar se existem diferenças significativas entre os desempenhos de estudantes homens e mulheres na prova de Avaliação Diagnóstica de Conhecimentos e Habilidades (EDICH), já que esta prova é a única avaliação comum a todas as carreiras da Faculdade de Ciências. A análise dessa prova concluiu que em geral as perguntas apresentam independência do sexo. As únicas exceções foram achadas em alguns casos na área física, onde as mulheres obtiveram pontuações significativamente mais baixas do que se esperaria se os desempenhos fossem independentes. Foi aprofundada esta análise e observou-se um viés a favor dos homens em as perguntas apresentadas na área física, o que poderia explicar as diferenças observadas.

Palavras chave: viés segundo sexo, avaliações universitárias.

Introducción y marco teórico

La Evaluación Diagnóstica de Conocimientos y Habilidades (EDICH) es una prueba que se aplica al comienzo de cada año lectivo a la generación de ingreso a la Facultad de Ciencias de la Universidad de la República. Se implementó en 2003 y a partir de entonces se ha seguido aplicando sin interrupciones, aunque con algunas variaciones y ajustes con miras a mejorarla. Esta evaluación consiste en preguntas de múltiple opción agrupadas por áreas. En el caso de las licenciaturas en Ciencias Biológicas y en Bioquímica, las áreas son matemática, física, química y biología. En el caso de las licenciaturas en Matemática y en Física, las áreas se reducen a matemática y física. Es de destacar que la EDICH es de asistencia obligatoria, pero no es eliminativa (Abella et al., 2010); esto implica que no hay presión en obtener buenos resultados, por lo cual para el estudiante funciona como un test de evaluación personal. En la Facultad de Ciencias hay otras licenciaturas, pero tienen una cantidad relativamente baja de ingresos (en comparación con las anteriores), por lo cual no fueron consideradas aquí.

El interés en este estudio es analizar los resultados de la EDICH para saber si presenta algún sesgo (es decir, si sus preguntas favorecen a uno u otro sexo) respecto al sexo de los participantes. Con respecto a esto último, los ingresos generales a la Facultad de Ciencias tienen una proporción 1:1 que es característica del área Tecno-

logías, Ciencias de la Naturaleza y el Hábitat (Dirección General de Planeamiento [DGPlan], Universidad de la República [Udelar], 2022). Esta proporción se mantiene en los ingresos a la Licenciatura en Ciencias Biológicas, pero se transforma en 2:1 a favor de las mujeres en la Licenciatura en Bioquímica, mientras que en las licenciaturas en Matemática y en Física la proporción se invierte, siendo de 2:1 a favor de los hombres, característico de las áreas STEM (Science, Technology, Engineering and Mathematics). Esta variedad en las proporciones hace que la Facultad de Ciencias sea un lugar potencialmente interesante para este tipo de estudios. Cuando se estudia el sesgo por sexo en las evaluaciones en general, se ha observado que el contenido de las preguntas puede influir en los resultados de las evaluaciones. Por ejemplo, Wilson et al. (2016), en un estudio que abarcó exámenes de física de múltiple opción realizados en Australia a lo largo de ocho años, encontraron que había un sesgo a favor de los hombres. Ellos adjudicaron este sesgo principalmente a factores tales como el contenido de las preguntas, el contexto de estas y la presentación de los datos (ecuaciones, diagramas, etc.). Lo mismo se encontró en estudios realizados en Estados Unidos (Docktor y Heller, 2008) y en el Reino Unido (Bates et al., 2013).

También hay que tener en cuenta que el formato de las preguntas, que en la EDICH es de múltiple opción, podría repercutir de manera diferencial según el sexo de quien realiza

la prueba. Existen numerosos estudios relacionados con esta temática en los que algunos autores postulan que el formato de múltiple opción no influenciaría (Wright et al., 2016), mientras que otros mantienen que sí (Graetz y Karimi, 2022; Griselda, 2020; Stanger-Hall, 2012).

Adicionalmente, varios autores han encontrado que el hecho de ser evaluadas presenta una desventaja para las mujeres, ya sea como amenaza del estereotipo o bien como ansiedad (Graetz y Karim, 2022; Griselda, 2020; Régner et al., 2016; Salehi et al., 2019; Spencer et al., 2016; Steele y Aronson, 1995; Wright et al., 2016). La *amenaza del estereotipo* se refiere a la obtención de resultados académicos más bajos de lo esperado en situaciones en que los integrantes de un grupo determinado se sienten amenazados por la posibilidad de que su desempeño confirmará —ya sea ante otros o ante ellos mismos— el estereotipo negativo acerca de las habilidades del grupo al que pertenecen. El hecho de ser evaluado aumenta la preocupación acerca de ser juzgado estereotípicamente, lo que lleva a su vez a un impedimento de procesar eficientemente la información que se presenta, resultando en un mal desempeño. Por ejemplo, las mujeres que se sienten amenazadas por su estereotipo tenderán a presentar un desempeño académico por debajo del de sus pares masculinos, aun poseyendo capacidades equivalentes o superiores (Régner et al., 2016; Spencer et al., 2016; Steele y Aronson, 1995). Respecto a la ansiedad, en un estudio de

Salehi et al. (2019) que involucró a más de 5.000 estudiantes en pruebas universitarias de nivel introductorio, se encontró que las mujeres tuvieron rendimientos más bajos que los hombres. Estos autores encontraron una correlación entre los rendimientos bajos de las mujeres en estas pruebas y la ansiedad que les generaba la prueba. Sin embargo, en otras instancias de evaluación que no implicaban una prueba formal, como por ejemplo actividades de laboratorio, no solo no se encontraron sesgos a favor de los hombres, sino que hubo casos en que las mujeres tuvieron un rendimiento académico por encima del de sus pares masculinos.

Finalizando esta introducción, se observa que la Mesa Interinstitucional Mujeres en Ciencia, Innovación y Tecnología (2020) concluyó que el ámbito educativo no es neutral al género, sino que habría ciertas normas de género que condicionan las trayectorias de hombres y mujeres y resultan en el alejamiento de estas últimas de las áreas STEM a lo largo de los distintos niveles educativos (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2017).

Objetivos

El objetivo de este estudio es analizar la EDICH para investigar si en ella existe un sesgo hacia uno u otro sexo y, en caso afirmativo, tratar de determinar sus causas.

Contexto

El análisis que se detalla a continuación se refiere a los resultados de la aplicación de la EDICH en cinco años consecutivos, de 2015 a 2019 inclusive, restringido a cuatro cohortes de estudiantes correspondientes a

las licenciaturas en Ciencias Biológicas (abreviado Biología), Bioquímica, Física y Matemática. En ese período hubo un total de 1.494 estudiantes, de los cuales 748 eran hombres y 746 mujeres. Más precisamente, hubo 332 hombres en Biología, 105 en Bioquímica, 166 en Física y 143 en Matemática, y 416 mujeres en Biología, 197 en Bioquímica, 65 en Física y 70 en Matemática.

Las cuatro cohortes se dividieron en dos bloques, uno formado por Biología y Bioquímica y el otro por Física y Matemática. Quienes integraban el primer bloque debieron responder preguntas en cuatro áreas: matemática, física, química y biología, mientras que los del segundo solo tuvieron preguntas de matemática y física. Cabe destacar que de las 16 preguntas del área física presentes en la prueba del segundo bloque, las primeras 12 coincidían con las del primer bloque. Algo similar sucedió con las preguntas del área matemática, en que de las 20 preguntas en la prueba del segundo bloque, las primeras 12 coincidían con las del primer bloque.

Metodología

Lo primero que se hizo fue calcular, para cada uno de los dos bloques, en cada año, las *medias* (es decir, el puntaje promedio) obtenidas en las pruebas de las distintas áreas. Así, para el bloque Biología-Bioquímica se calcularon las medias obtenidas por los estudiantes hombres en cada una de las pruebas de matemática, física, química y biología. Lo mismo se hizo con las de las estudiantes mujeres y con el total de estudiantes. Eso se repitió en el bloque Física-Matemática con las pruebas de matemática y física (en la sección siguiente están los resultados). Al analizar las medias se observó que los rendimientos generales en cualquiera de las distintas áreas fueron ba-

jos para ambos sexos y que las medias de las mujeres fueron casi siempre inferiores a las de los hombres.

Se pasó luego a analizar si esta diferencia en el rendimiento entre mujeres y hombres era significativa. Dado los bajos valores de las medias, no se pudo utilizar el simple criterio de aprobado/no aprobado para analizar los rendimientos en la prueba EDICH, ya que rara vez se alcanzó el nivel de suficiencia (más de 50 puntos en 100). Fue por ello que se decidió establecer tres categorías de rendimiento: alto, medio y bajo, para lo cual se convirtieron las notas obtenidas a porcentajes y luego se halló la media aritmética para cada área temática y para cada cohorte; también se halló el desvío estándar correspondiente a cada caso. Se definieron así las siguientes categorías:

Rendimiento bajo = toda nota menor que la media aritmética menos el desvío estándar

Rendimiento medio = toda nota entre la media aritmética menos el desvío estándar y la media aritmética más el desvío estándar

Rendimiento alto = toda nota mayor que la media aritmética más el desvío estándar

Esta categorización de los rendimientos usando el desvío estándar está inspirada en el método utilizado por Régner et al. (2016).

Una vez establecidas las tres categorías, se procedió a realizar el test de chi-cuadrado para analizar si había diferencias significativas entre los distintos rendimientos, es decir, para saber si las diferencias obtenidas podían deberse o no al azar (Agresti, 2007; Cortés et al., 2014). La diferencia se considera *significativa* cuando el valor del chi-cuadrado es mayor de 5,99 (para dos grados de libertad y un valor p menor a 0,05). Este análisis se hizo para los cinco años en estudio y para ambos bloques. El problema del test del chi-cuadrado es que su valor aumenta con el tamaño de la muestra,

tendiendo a hacer pensar que existe correlación cuando no la hay. Eso nos afecta porque en nuestro estudio vamos a tener que comparar el comportamiento de las cuatro cohortes en cinco períodos distintos, con variaciones importantes en la cantidad de alumnos. Para remediarlo se suele aplicar, y fue lo que hicimos, otro estadístico llamado el coeficiente *V* de Cramer, que toma valores entre 0 y 1. Solo cuando el valor del chi-cuadrado es superior a 5,99 y el valor del *V*

de Cramer es 0,30 o mayor, se considera que hay una diferencia significativa. Esto se debe a que a partir del valor 0,30 el estadístico de Cramer nos dice que existe una asociación de magnitud moderada entre las variables en estudio (Baguley, 2009; Cortés et al., 2014).

Como resultado de aplicar los estadísticos anteriores se detectaron diferencias significativas en los resultados de algunas preguntas del área física, por lo cual se analizó la *facilidad relativa* de

cada pregunta de esta área, para intentar determinar si estaba relacionada con el sexo de quien respondía la pregunta, según lo descrito por Wilson et al. (2016). Con este fin, para cada pregunta (y en cada año) se calculó el *porcentaje de aprobación de estudiantes hombres*, obtenido realizando el cociente entre la cantidad de estudiantes hombres que respondieron bien la pregunta sobre el total de estudiantes hombres que respondieron la pregunta y luego multiplicándolo por 100:

$$\text{porcent. aprob. estud. hom.} = \left(\frac{\text{estud. hombres con respuesta correcta}}{\text{total estud. hombres}} \right) 100$$

En forma análoga se obtuvo el *porcentaje de aprobación de estudiantes mujeres*. Finalmente se obtuvo la *facilidad*

relativa, definida como la resta del porcentaje de aprobación de estudian-

tes hombres menos el porcentaje de aprobación de estudiantes mujeres:

$$\text{facilidad relativa} = \text{porcent. aprob. estud. hom.} - \text{porcent. aprob. estud. muj.}$$

Los valores positivos de la *facilidad relativa* señalan una ventaja para los hombres; los negativos, para las mujeres. Nótese que si las preguntas tuviesen la misma *facilidad* para hombres y mujeres, este valor sería 0. En su aplicación, cuando el valor de la *facilidad relativa* es superior a +1,5 se consideró que está a favor de los hombres, cuando el valor es inferior a -1,5, que está a favor de las mujeres, y cuando la diferencia en *facilidad relativa* está entre -1,5 y +1,5 que no favorece a ningún sexo. Además, a cada pregunta le asociamos una *tendencia* de acuerdo al siguiente criterio: cuando en tres o más años a los hombres les va mejor se le adjudica la letra M (masculino), cuando ocurre lo mismo con las mujeres se le adjudica la F (femenino) y cuando no hay predominancia de uno sobre el otro se le adjudica la N (neutro). Recuérdese que estamos evaluando un período de cinco años. En el análisis de los resultados del párrafo anterior resultó que había preguntas de las tres tendencias, por lo

cual para tratar de entender mejor lo que ocurría con las preguntas del área física se decidió analizar cada pregunta de manera individual, nuevamente tomando como referencia el trabajo realizado por Wilson et al. (2016). En dicho trabajo, se establecían cinco dimensiones distintas para clasificar las preguntas de las pruebas de física:

a. *Contenido*: los conocimientos que un estudiante debe tener para responder la pregunta.

b. *Proceso*: las habilidades necesarias para responder la pregunta (interpretar una gráfica, aplicar una fórmula, etc.).

c. *Presentación*: cómo se presenta la información, *i.e.*, si hay ecuaciones, gráficas o si es mayoritariamente texto.

d. *Contexto*: grado de abstracción a partir de experiencias personales, *i.e.*, si el estudiante pudo haber vivido o experimentado una situación similar.

e. *Dificultad*: qué tan compleja es la pregunta, por ejemplo, el número de pasos o de conceptos necesarios para resolverla.

Cada una de estas dimensiones tiene, a su vez, distintas categorías. Por ejemplo, en la dimensión *contenido* las categorías fueron: ecuación, gráfica, texto y diagrama (tabla 9, al final). Lo mismo sucede con las otras. Luego, para cada pregunta con tendencia M o F, se miraron las dimensiones anteriores para ver en qué categorías caía. Es decir, en las preguntas que parecían tener cierto sesgo, se fijó en el análisis de sus dimensiones con el fin de intentar determinar las causas.

Resultados y discusión

En las tablas 1 y 2 se presentan las medias por área para cada uno de los dos bloques, Biología-Bioquímica y Física-Matemática. En ambas tablas G es la media general, ♀ es la media obtenida por las mujeres del bloque y ♂ es la media de los hombres. El puntaje máximo para cada prueba es 100.

Tabla 1: Medias de rendimiento por área, en el bloque Ciencias Biológicas-Bioquímica

Año	Matemática			Física			Química			Biología		
	G	♀	♂	G	♀	♂	G	♀	♂	G	♀	♂
2015	47,3	45,8	49,3	31,4	27,9	35,8	45,8	44,7	47,1	54,3	54,2	54,4
2016	48,2	47,8	48,8	30,5	28,7	32,6	43,1	41,3	45,1	54,5	50,3	59,7
2017	45,2	45,0	45,5	27,9	26,4	30,1	43,0	41,6	45,3	54,4	53,3	56,3
2018	48,7	47,8	50,2	29,2	27,7	31,5	44,1	42,6	46,3	55,0	54,3	56,1
2019	41,6	40,7	43,1	27,5	24,7	31,6	39,1	37,0	42,2	52,0	49,7	55,2

Tabla 2: Medias de rendimiento por área, en el bloque Física-Matemática

Año	Matemática			Física		
	G	♀	♂	G	♀	♂
2015	53,7	47,7	57,1	34,0	26,8	37,9
2016	62,3	61,6	62,6	42,8	36,4	45,6
2017	60,4	61,7	59,7	44,8	44,0	45,3
2018	63,9	54,2	67,1	39,5	32,3	41,9
2019	67,0	62,1	69,2	44,5	37,3	47,6

Analizando las tablas anteriores se observa que las medias de las mujeres son siempre más bajas que las de los hombres, con la única excepción del año 2017 para la prueba del área matemática correspondiente al bloque Física-Matemática.

A continuación se busca determinar en qué casos esas diferencias en las medias son significativas. Las tablas 3 y

4 resumen los resultados de aplicar los test de chi-cuadrado y V de Cramer. En ambas tablas, junto al año figura entre paréntesis el número total de estudiantes que integra cada cohorte; M representa las preguntas del área matemática; F las de física; Q las de química, y B las de biología. Estas tablas presentan las diferencias en el rendimiento por sexo dentro de cada una de las dos licen-

ciaturas que componen el bloque. La palabra *No* indica que las diferencias halladas no fueron significativas; en cambio, cuando las diferencias entre lo esperado y lo observado resultaron significativas se consigna *Sí* junto con el valor del V de Cramer. Además, en los casos en que la V de Cramer es superior a 0,30, la zona correspondiente aparece sombreada.

Tabla 3: Resultados del test de chi-cuadrado para el bloque Biología-Bioquímica

	2015 (n=179)				2016 (n=212)				2017 (n=193)				2018 (n=224)				2019 (n=242)			
	M	F	Q	B	M	F	Q	B	M	F	Q	B	M	F	Q	B	M	F	Q	B
Intralic. en Cs. Biológicas	No	No	No	No	No	No	No	Sí 0,22	No	No	No	No	No	No	No	No	No	No	No	Sí 0,20
Intralic. en Bioquímica	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Sí 0,41	Sí 0,29	No

Tabla 4: Resultados del test de Chi-cuadrado para el bloque Física/Matemática

	2015 (n=67)		2016 (n=97)		2017 (n=81)		2018 (n=97)		2019 (n=102)	
	M	F	M	F	M	F	M	F	M	F
Intralic. en Física	No	No	No	No	No	No	No	No	No	Sí 0,35
Intralic. en Matemática	No	Sí 0,50	No	No	No	No	No	No	Sí 0,36	No

En términos generales se puede observar que no existen diferencias significativas en los rendimientos de las distintas áreas según el sexo de quien responde las preguntas. Solo hay cuatro excepciones, tres de las cuales corresponden a preguntas del área física y una al área matemática (véanse las casillas sombreadas en las tablas 3 y 4). Estas diferencias implican que en dichos casos el desempeño de las mujeres está por debajo de lo que se esperaría si los resultados de alto, medio y bajo se distribuyeran de manera independiente del sexo.

Adicionalmente, al analizar las diferencias intralicenciatura del año 2019 para las licenciaturas en Física y en Matemática (tabla 4, fila 1/columna 10 y fila 2/columna 9), se observa que las mujeres inscritas en Física no tienen diferencias significativas con sus pares masculinos a la hora de responder las preguntas del área matemática; sin embargo, a estas mismas mujeres les va significativamente peor que a los hombres cuando tienen que responder las preguntas del área física.

Algo análogo ocurre en la Licenciatura en Matemática, pero intercambiando los roles de las preguntas de física con las de matemática. Esto podría estar indicando la ocurrencia del fenómeno de amenaza del estereotipo. Cabe aclarar que no es la única explicación posible para estos resultados del año 2019, ya que no se realizaron otros estudios, como por ejemplo para medir ansiedad.

En los datos obtenidos, llama la atención que tres de los cuatro casos en los que se encontraron diferencias significativas en el rendimiento según sexo correspondían al área física (tablas 3 y 4). Además, las medias generales de los resultados en las preguntas del área física (en ambos bloques) son las más bajas de todas las áreas, con un valor promedio de 29,2 en 100 para el bloque Biología-Bioquímica y un promedio de 41,5 en 100 para el bloque Física-Matemática (tablas 1 y 2). Por lo tanto se decidió profundizar en las preguntas del área física, en ambos bloques.

A continuación se presentan cuatro tablas numeradas 5, 6, 7 y 8, en las que aparece el valor de la *facilidad relativa* de cada pregunta del área física, por año y por licenciatura, según lo descrito en la sección “Metodología”. En las tablas 5 y 6 se ve la facilidad relativa de cada una de las 12 preguntas del área física para el bloque Biología-Bioquímica. En las tablas 7 y 8 se ve la facilidad relativa de las 16 preguntas del área física para el bloque Física-Matemática. Recuérdese que las 12 primeras preguntas de física del bloque Física-Matemática son las mismas que las del bloque Biología-Bioquímica. Cuando el valor de la facilidad relativa es superior a +1,5 se colorea la casilla en verde (está a favor de los hombres), cuando el valor es inferior a -1,5 se colorea en violeta (está a favor de las mujeres) y cuando está entre -1,5 y +1,5 aparece en blanco. Finalmente se agregó la fila *Tendencia*, determinada de acuerdo al criterio de predominancia de facilidad relativa por tres años, descrito en la sección “Metodología”.

Tabla 5: Facilidad relativa de cada pregunta del área física para la Licenciatura en Biología

Licenciatura en Biología	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
2015	-6,1	0,1	-3,7	3,5	6,7	17,8	-7,5	18,9	9,0	15,1	-2,0	10,3
2016	17,7	3,4	-7,9	-19,6	6,8	12,9	-2,8	14,0	14,0	15,5	0,2	11,9
2017	0,1	8,3	3,4	15,3	9,5	8,5	-8,3	4,1	-3,7	18,9	0,2	6,0
2018	5,8	7,3	-8,2	-4,7	5,1	14,5	-0,7	20,7	5,7	1,9	4,9	5,1
2019	-4,6	12,3	2,2	1,4	17,4	-1,4	2,1	14,2	17,9	3,1	0,6	7,5
Tendencia	N	M	F	N	M	M	F	M	M	M	N	M

Tabla 6: Facilidad relativa de cada pregunta del área física para la Licenciatura en Bioquímica

Licenciatura en Bioquímica	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
2015	12,6	9,0	11,9	2,9	20,5	1,4	0,5	40,5	24,0	16,4	4,3	5,7
2016	-18,6	0,0	0,0	-2,6	20,8	10,0	-2,6	20,3	5,2	-25,1	-5,6	5,2
2017	5,6	14,0	-3,5	1,1	1,1	-19,4	-6,5	3,0	-16,9	-0,8	-4,6	-3,5
2018	-8,8	9,3	-6,6	14,1	10,6	10,0	9,1	-19,4	6,2	-6,8	8,4	22,1
2019	21,6	5,6	2,3	12,9	18,8	35,2	-2,0	11,3	30,1	10,2	2,7	15,8
Tendencia	M	M	N	M	M	M	F	M	M	N	M	M

Si se analizan las 12 preguntas para este primer bloque en su conjunto, se ve que la mitad de ellas (preguntas 2, 5, 6, 8, 9 y 12) presentan un sesgo a favor de los hombres, mientras

que solo en una (pregunta 7) hay un sesgo a favor de las mujeres. Además, en el caso de Biología a las mujeres también les va mejor en la pregunta 3.

Siguiendo con Biología, hay tres casos de neutralidad (preguntas 1, 4 y 11), mientras que en Bioquímica hay dos casos (preguntas 3 y 10).

Tabla 7: Facilidad relativa de cada pregunta del área física para la Licenciatura en Física

Licenciatura en Física	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
2015	-5,5	-3,2	24,5	3,6	18,6	27,7	-18,2	13,6	7,7	-9,5	30,0	24,5	-6,8	-7,3	6,8	1,8
2016	-1,2	-5,5	-17,1	12,2	7,4	12,5	3,8	10,0	47,4	-3,0	21,4	22,9	-6,2	5,8	23,8	6,1
2017	0,0	6,7	-16,7	-10,0	-13,3	3,3	-16,7	-6,7	-20,0	-3,3	-16,7	6,7	3,3	0,0	-20,0	-10,0
2018	4,2	11,8	10,9	10,1	26,1	34,0	18,1	32,8	21,8	-0,8	12,2	30,7	12,6	18,5	20,6	21,0
2019	-3,3	8,9	4,4	14,4	-19,4	16,1	-6,7	26,1	27,8	16,7	21,7	42,8	7,2	20,0	5,0	18,3
Tendencia	F	M	M	M	M	M	F	M	M	F	M	M	M	M	M	M

Tabla 8: Facilidad relativa de cada pregunta del área física para la Licenciatura en Matemática

Licenciatura en Matemática	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
2015	23,7	43,5	14,7	20,4	44,5	-17,7	8,7	6,4	34,4	28,1	-2,3	4,0	12,4	-26,4	14,0	27,1
2016	18,8	3,2	8,4	-3,9	9,1	24,0	-14,9	-13,6	7,1	-13,0	20,1	-8,4	-2,6	6,5	33,8	13,6
2017	1,2	8,8	32,1	7,1	17,2	-6,1	-0,2	5,6	-8,8	20,6	3,2	-33,8	2,7	-0,2	16,7	4,9
2018	1,3	18,5	4,1	13,8	23,8	-16,9	0,5	23,6	18,7	5,9	-12,1	-29,0	-7,4	3,1	-2,3	-4,4
2019	3,8	13,4	5,0	-4,3	13,9	-6,1	16,6	12,1	10,5	8,8	-2,1	28,8	-1,3	21,8	-7,3	15,5
Tendencia	M	M	M	M	M	F	N	M	M	M	F	F	N	M	M	M

Analizando las 16 preguntas para este segundo bloque en su conjunto, se ve que nueve de ellas (preguntas 2, 3, 4, 5, 8, 9, 14, 15 y 16) presentan un sesgo a favor de los hombres. En el caso de la Licenciatura en Física, las mujeres responden mejor en tres preguntas: 1, 7 y 10, y no existen preguntas neutras. En el caso de la Licenciatura en Matemática, las mujeres responden

mejor a tres preguntas: 6, 11 y 12, y hay dos preguntas neutras, la 7 y la 13. La presencia de diferencias significativas entre hombres y mujeres en el rendimiento en las preguntas del área física de la EDICH podría vincularse a una facilidad relativa a favor de los hombres en estas. Para estudiar esto último, se asignaron categorías en cada una de las cinco dimensiones

descritas en la sección “Metodología” y se clasificó cada pregunta de acuerdo a estas categorías. En base a esto se construyó la tabla 9, en la que aparecen las categorías en las cuales se clasifica cada pregunta, junto con su tendencia (tomada de las tablas 5 a 8); esto para todas las preguntas del área física en las cuatro licenciaturas.

Tabla 9: Categorización y tendencia relativa de cada pregunta del área física

Dimensiones	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Presentación	ecuación	gráfica	gráfica	texto	texto	texto	texto	texto	texto	diagrama	diagrama	texto	texto	diagrama	texto	diagrama
Contexto	abstracto	libro de texto	libro de texto	libro de texto	libro de texto	libro de texto	libro de texto	libro de texto	real/no personal	libro de texto	libro de texto	real/no personal	libro de texto	libro de texto	real/no personal	libro de texto
Contenido	análisis dimensional	cinemática/proyectiles	cinemática/unidim.	cinemática/unidim.	cinemática/mov. circular	dinámica, trabajo y energía	dinámica choques	dinámica mov. oscilatorio	ondas	electrostática	corriente y resistencia	óptica	electrostática	electrostática	corriente circuitos	óptica
Proceso	identif. fuerza	interpretar diagrama	interpretar diagrama	identif. fuerza	identif. fuerza	identif. fuerza	identif. fuerza	identif. fuerza	hacer cálculos	identif. fuerza	identif. fuerza	identif. fuerza	identif. fuerza	interpretar ecuación	interpretar ecuación	interpretar diagrama
Dificultad	concepto único	más de un concepto	más de un concepto	más de un concepto	concepto único	más de un concepto	más de un concepto	más de un concepto	más de un concepto	más de un concepto	dim. espacial única	dim. espacial única	más de un paso	dim. espacial única	más de un concepto	más de un concepto
Tendencia Lic. Fís.	F	M	M	M	M	M	F	M	M	F	M	M	M	M	M	M
Tendencia Lic. Mat.	M	M	M	M	M	F	N	M	M	M	F	F	N	M	M	M
Tendencia Lic. Cs. Biol.	N	M	F	N	M	M	F	M	M	M	N	M	N/A	N/A	N/A	N/A
Tendencia Lic. Bioq.	M	M	N	M	M	M	F	M	M	N	M	M	N/A	N/A	N/A	N/A

Nota: Las cuatro filas de tendencias resumen las correspondientes de las tablas 5 a 8. Las 12 primeras preguntas fueron comunes a todas las licenciaturas, mientras que las preguntas 13 a 16 fueron solo para física y matemática, de ahí el N/A (no aplica) en las dos últimas filas. Se emplearon las siguientes abreviaciones: unidim. = unidimensional, mov. = movimiento, identif. = identificar, dim. = dimensión.

Analizando las preguntas en su conjunto y teniendo en cuenta la tendencia asignada a cada una, se ve que solamente en una (pregunta 7) les va mejor a las mujeres. Hay siete (preguntas 2, 5, 8, 9, 14, 15 y 16) en que la tendencia siempre favorece a los hombres. Finalmente, hay cinco (preguntas 1, 3, 10, 11 y 13) que resultan un poco más balanceadas ya que a veces les va mejor a los hombres, otras veces a las mujeres y en otras la tendencia es neutra.

Al analizar la pregunta 7 se observa que se diferencia en la dimensión contenido, ya que es la única pregunta de dinámica de choques. Un análisis similar de las siete preguntas con predominancia masculina absoluta deja en evidencia que también es esta dimensión, la del contenido, que hace que estas preguntas se diferencien de las demás. En este grupo aparece, por ejemplo, la única pregunta sobre proyectiles. Del mismo modo, la única pregunta sobre movimiento circular, la única sobre ondas, sobre dinámica del movimiento oscilatorio y sobre corriente y circuitos pertenecen a este grupo “masculino”. Las otras dimensiones, *i.e.*, presentación, contexto, proceso y dificultad, no estarían incidiendo en la facilidad relativa de cada pregunta.

Wilson et al. (2016) hallaron que las tres dimensiones que influían más en la facilidad relativa de cada pregunta eran contexto, presentación y contenido. En cuanto a la dimensión contexto, encontraron que la categoría “real y posiblemente experimentado” era la que les otorgaba una clara ventaja a los hombres. En nuestro caso, no tenemos preguntas con este contexto. En lo que se refiere a la dimensión presentación, los autores encontraron ventajas a favor de los hombres en preguntas que presentaban la información a través de un diagrama. En nuestro caso, hay cuatro preguntas que caen en esta categoría, dos de ellas están en el grupo de predomi-

nancia masculina absoluta mientras que las otras dos están en el grupo de las preguntas más “balanceadas”. Esto indicaría que, en nuestro caso, esta dimensión por sí sola no sería la responsable de conferir ventajas a un sexo u otro. Finalmente, ellos encuentran que la dimensión contenido también estaría influyendo en la facilidad relativa de las preguntas, en especial cuando se trata de proyectiles. Nuestro análisis coincide totalmente con el de estos autores en cuanto a que la dimensión contenido estaría influenciando la facilidad relativa de las preguntas planteadas.

Conclusiones finales

En cuanto al análisis de la existencia de diferencias en los rendimientos entre alumnos hombres y mujeres en la EDICH, se puede concluir que, si bien las mujeres tuvieron casi siempre rendimientos inferiores, estas diferencias en la amplia mayoría de los casos no llegan a ser significativas: solo en 4 casos en 60 se encontraron diferencias significativas (tablas 3 y 4), lo cual nos dice que en el 93% de los casos no las hubo. Por lo tanto, se puede concluir que en términos generales en las preguntas de la EDICH existe independencia del sexo. Este resultado está en consonancia con lo obtenido por Dietz et al. (2012) en Estados Unidos (en el sentido de que hay diferencias, pero estas no son significativas como para poder catalogarse como un sesgo) y en contraposición con lo de Bates et al. (2013) en el Reino Unido y de Wilson et al. (2016) en Australia. Sin embargo corresponde notar que los contextos de los estudios citados tienen diferencias con los nuestros: en los casos de Estados Unidos y Reino Unido, se refieren a los resultados del FCI (el Force Concept Inventory, que es un test de conocimientos de física), realizado en cursos introductorios de

física a nivel universitario, mientras que los de Australia corresponden al examen para integrar el equipo de estudiantes que asistirá a olimpiadas de física. En los dos primeros casos los estudiantes requieren ciertas calificaciones para asistir a esos cursos, mientras que en el tercero son elegidos por los profesores como los mejores alumnos para representar a sus instituciones. Estas situaciones son distintas de la nuestra, en la cual no se exigen requisitos particulares para realizar la EDICH y tampoco les genera ninguna repercusión académica el resultado obtenido. Es decir, se tomaron los artículos citados para obtener modelos estadísticos que se pudiesen adaptar a nuestro caso, pero no se repitieron sus experiencias.

Volviendo al resultado obtenido en el análisis, hay que tener en cuenta de que si bien no se obtuvo evidencia de que la diferencia de rendimiento entre estudiantes hombres y mujeres se deba a la existencia de un sesgo en las preguntas, este podría ser consecuencia del formato de la prueba (múltiple opción) como mantienen algunos autores (Graetz y Karim, 2022; Griselda, 2020; Stanger-Hall, 2012). Eso sería difícil de comprobar en nuestro caso, dado que la prueba EDICH tiene un formato único, además de las dificultades que se presentan para realizar evaluaciones masivas con otros tipos de formatos.

Más allá de lo anterior, se encontraron cuatro casos (uno en 2015 y tres en 2019, tablas 3 y 4) en los cuales las mujeres obtuvieron rendimientos significativamente más bajos de lo que se esperaría si estos fuesen independientes del sexo. Esto sucedió en los dos bloques en estudio, Biología-Bioquímica y Física-Matemática. Tres de estos cuatro casos se encontraron en preguntas del área física. Por lo tanto, se profundizó en el análisis de las preguntas de esta área y en la mayoría de los casos se observó un sesgo a favor de los hombres. Luego, aplicando

la metodología desarrollada por Wilson et al. (2016), se encontró que la dimensión *contenido* de las preguntas podría ser determinante a la hora de conferir ventajas a uno u otro sexo. Esto estaría en consonancia con los

trabajos de Wilson et al. (2016) y de Salehi et al. (2019), en el sentido de que en las preguntas del área física habría un sesgo que favorece a los hombres.

En base a lo anterior, se entiende que se deberían revisar las preguntas del área física de la EDICH, priorizando aquellas con contenidos más balanceados.

Referencias bibliográficas

- Abella, A., Cabrera, C., De León, L., Varela, G., y Villalba, V. (2010). *Análisis de la Evaluación Diagnóstica de Conocimientos y Habilidades en la Facultad de Ciencias* (Informe para el Consejo de la Facultad de Ciencias, Universidad de la República).
- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. John Wiley & Sons Hoboken, NJ.
- Baguley, T. (2009). Standardized or simple effect size: What should be reported? *Br J Psychol*, 100(Pt 3), 603-617.
- Bates, S., Donnelly, R., MacPhee, C., Sands, D., Birch, M., y Walet, N. R. (2013). Gender differences in conceptual understanding of Newtonian mechanics: A UK cross-institution comparison. *Eur J Phys*, 34(2), 421.
- Cortés, F., Rubalcava, R. M., y Fernández, T. (2014). *Estadística Social Básica* (2.ª ed., 3.ª reimpr.). Departamento de Sociología, Facultad de Ciencias Sociales. Universidad de la República- Programa Universitario de Estudios sobre el Desarrollo, Universidad Nacional Autónoma de México.
- Dietz, R. D., Pearson, R. H., Semak, M. R., y Willis, C. W. (2012). Gender bias in the Force Concept Inventory? *AIP Conf Ser*, 1413, 171.
- Dirección General de Planeamiento, Universidad de la República. (2022). *Estadísticas Básicas 2020 de la Universidad de la República*.
- Docktor, J., y Heller, K. (2008). Gender differences in both Force Concept Inventory and introductory physics performance. *AIP Conf Ser*, 1064, 15.
- Graetz, G., y Karimi, A. (2022). Gender gap variation across assessment types: Explanations and implications. *Economics of Education Review*, 91, id:102313. <https://doi.org/10.1016/j.econedurev.2022.102313>
- Griselda, S. (2020). Different Questions, Different Gender Gap: Can the Format of Questions Explain the Gender Gap in Mathematics? 2020 *Papers pgr710*, Job Market Papers.
- Mesa Interinstitucional Mujeres en Ciencia, Innovación y Tecnología. (2020). *Mujeres en ciencia, tecnología e innovación en Uruguay: Un factor clave para avanzar en igualdad de género y desarrollo sostenible*.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2017). *Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM)*. París, Francia.
- Régner, I., Selimbegovic, L., Pansu, P., Monteil, J. M., y Huguet, P. (2016). Different Sources of Threat on Math Performance for Girls and Boys: The Role of Stereotypic and Idiosyncratic Knowledge. *Front Psychol*, 7, 637. <https://doi.org/10.3389/fpsyg.2016.00637>
- Salehi, S., Cotner, S., Azarin, S. M., Carlson, E. E., Driessen, M., Ferry, V. E., Harcombe, W., McGaugh, S., Wassenberg, D., Yonas, A., y Ballen, C. J. (2019). Gender Performance Gaps Across Different Assessment Methods and the Underlying Mechanisms: The Case of Incoming Preparation and Test Anxiety. *Front Educ*, 4, 107. <https://doi.org/10.3389/educ.2019.00107>
- Spencer S. J., Logel C. y Davies P. G. (2016). Stereotype Threat. *Annual Review of Psychology*, 67, 415-437.
- Stanger-Hall, K. F. (2012). Multiple-choice exams: an obstacle for higher-level thinking in introductory science classes. *CBE Life Sci Educ*, 11, 294-306.
- Steele, C. M., y J Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol*, 69(5), 797-811.
- Wilson, K., Low, D., Verdon, M., y Verdon, A. (2016). Differences in gender performance on competitive physics selection tests. *Phys Rev Phys Educ Res*, 12, 020111-16.
- Wright, C. D., Eddy, S. L., Wenderoth, M. P., Abshire, E., Blankenbiller, M., y Brownell, S. E. (2016). Cognitive Difficulty and Format of Exams Predicts Gender and Socioeconomic Gaps in Exam Performance of Students in Introductory Biology Courses. *CBE Life Sci Educ*, 15, 1-16. <https://doi.org/10.1187/cbe.15-12-0246>

Contribución de autoría

Participación: 1. Concepción y diseño del estudio 2. Adquisición de datos 3. Análisis de datos 4. Discusión de los resultados 5. Redacción del manuscrito 6. Aprobación de la versión final del manuscrito

E. Rodríguez: 1, 2, 3, 4, 5, 6

M. Davyt: 2, 3, 4, 6

J. Hurtado: 2, 3, 4, 6

A. Abella: 1, 3, 4, 5, 6

Disponibilidad de datos

El conjunto de datos utilizados en este estudio tiene carácter reservado y pertenece a la Unidad de Enseñanza de la Facultad de Ciencias.

Agradecimientos

A la Unidad de Enseñanza de la Facultad de Ciencias, por suministrar los datos que permitieron hacer este análisis.

A Arturo Martí (Instituto de Física de la Facultad de Ciencias) y Álvaro Suárez (Administración Nacional de Educación Pública), por el análisis detallado de las preguntas de física presentes en la EDICH.

A Cecilia Muniz, por sus aportes para mejorar el manuscrito, basados en su formación en Ciencias Sociales y en Estadística.

