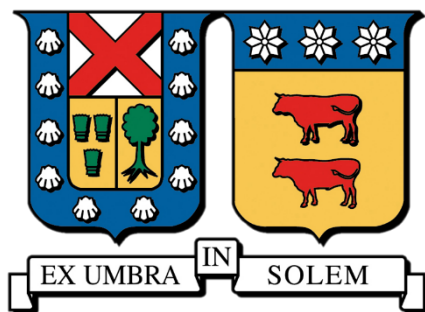


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE FÍSICA

VALPARAÍSO — CHILE



**“TEORÍA DE RESPUESTA AL ÍTEM Y SUS APLICACIONES CON *JMETRIK*
PARA EL ANÁLISIS DE DATOS DE EVALUACIONES DE DIAGNÓSTICO EN
LA FÍSICA NEWTONIANA”**

Ricardo Lizana Cortés

**PROYECTO DE INVESTIGACIÓN PARA OPTAR AL GRADO DE LICENCIADO
EN FÍSICA**

Profesor Guía

Dr. Hugo Alarcón

Comisión Evaluativa

Dr. Gonzalo Fuster

Prof. Claudio Figueroa

FEBRERO, 2025

Índice

1.- Introducción.....	2
2.- Marco Teórico	4
2.1.- <i>Preconceptos Erróneos</i>	4
2.1.1.- <i>Conductismo y Constructivismo</i>	5
2.1.2.- <i>Soluciones</i>	6
2.2.- <i>FCI y FMCE</i>	7
2.3.- <i>Análisis de Datos en Evaluaciones de Selección Múltiple</i>	9
2.3.1.- <i>Item Response Theory</i>	10
2.4.- <i>jMetrik</i>	11
2.4.1.- <i>Curvas No-Paramétricas</i>	12
3.- Metodología.....	14
4.- Recopilación de Datos y Análisis.....	19
4.1.- <i>FCI 2023: Análisis iniciales</i>	19
4.2.- <i>FCI 2012 v/s FMCE 2013</i>	25
4.3.- <i>FCI 2012 v/s 2024: A doce años de diferencia</i>	34
4.4.- <i>FMCE 2014: Una Retroalimentación Más Precisa</i>	42
5.- Conclusiones.....	46
6.- Bibliografía.....	47

1.- Introducción

A la hora de enseñar, el mayor desafío para un instructor, y su objetivo principal, es promover en sus estudiantes un aprendizaje significativo que les permita alcanzar una comprensión profunda y clara de los temas tratados en la asignatura. La gran interrogante que enfrenta cualquier persona dedicada a la labor de enseñar es: ¿cómo fomentar un aprendizaje efectivo y adaptado a las necesidades individuales de cada estudiante? O, en palabras más simples y centrándose en el sujeto del aprendizaje, ¿cómo aprenden las personas?

A primera vista, la respuesta a esta pregunta podría parecer obvia, por ejemplo, “una persona aprende cuando alguien le enseña”, pero el problema del aprendizaje no es trivial, y ha ocupado el esfuerzo de personas desde la antigüedad dentro de la rama de la filosofía de la epistemología, cuyo objetivo es estudiar el conocimiento, sobre todo el científico, y su naturaleza, origen, y manera de comprenderlo.

El cómo un ser humano adquiere y formula sus propios conocimientos es un asunto complejo, y no se tiene un consenso fijo que responda a esta pregunta, aunque sí existen varias líneas de pensamiento que pueden ayudar a obtener herramientas para alcanzar a una mejor forma, no solo de enseñanza, sino también de cómo retroalimentar la manera en la que esta enseñanza es adquirida por los estudiantes, y con cuánta eficiencia los profesores plantean en el aula los distintos conocimientos a ser transmitidos.

Independiente de la postura que elijamos para entender cómo una persona aprende, la gran mayoría de interpretaciones comparten un punto en común que será el centro de gran parte de esta investigación: los preconceptos erróneos, ideas alternativas, o, por su término en inglés “*misconceptions*” (Mestre & Docktor, 2021), que son obstáculos para que los estudiantes logren un aprendizaje significativo. Estos preconceptos erróneos serán el principal objeto de este estudio. En particular, se buscará reconocer y estudiar la evolución de los preconceptos erróneos de un grupo de estudiantes haciendo el seguimiento de una asignatura a lo largo de un semestre. Para este propósito, se hará uso de diferentes instrumentos de análisis de datos, aplicados a los resultados de evaluaciones de selección múltiple.

El primero de estos instrumentos es el software JMetrik (Meyer, 2014), una herramienta diseñada exclusivamente para el estudio y análisis de evaluaciones con múltiples formas de

análisis de datos en evaluaciones de selección múltiple. Este software contiene opciones tales como: análisis de modelo, teoría de respuesta al ítem, (*IRT*, por sus siglas en inglés) en sus diferentes modelos (uno, dos y tres parámetros), y análisis de gráficos no-paramétricos, entre otras. En este estudio utilizaremos principalmente el análisis de gráficos no paramétricos.

Por otro lado, se usará como instrumento de medición el Inventario de Conceptos de Fuerza (Hestenes, Wells, & Swackhamer, 1992), o *FCI*, por sus siglas en inglés. Esta herramienta de evaluación de conceptos de mecánica newtoniana fue diseñada a finales de los noventa, usando una metodología cuidadosa y exhaustiva, construyendo cada pregunta y cada posible respuesta con el fin de identificar y resaltar los preconceptos erróneos de los estudiantes en el ámbito de la Física newtoniana, buscando detectar a aquellos estudiantes que presentan un mejor “pensamiento newtoniano”, y a los estudiantes que presentan un menor conocimiento previo en el área con el fin de poder apoyarlos a lo largo del semestre en el que se dicta la asignatura correspondiente.

Además, se usará un segundo instrumento de medición: la Evaluación Conceptual de Fuerza y Movimiento, (Thornton & Sokoloff, 1998) o *FMCE*, por su sigla en inglés. Con esta segunda evaluación, se podrán resaltar las características únicas de ambos instrumentos de medición, así como también la manera en la que el software elegido para el análisis de los datos se comporta con dos evaluaciones de naturaleza similar pero con claras diferencias en su diseño y objetivos específicos.

Un primer objetivo del presente trabajo consiste en identificar errores conceptuales de los estudiantes y monitorear si hay cambios en la comprensión conceptual de los a lo largo de un semestre, comparando los resultados de un mismo grupo entre el inicio (pre test) y el final (post test) del semestre.

Un segundo objetivo de este estudio consiste en comparar resultados intergeneracionales, es decir, comparar los resultados de estudiantes de cohortes separadas por diez años de diferencia. Dentro de este segundo análisis, además, se aprovechará de comparar directamente las diferencias en los resultados obtenidos en ambas evaluaciones de diagnóstico (*FMCE* y *FCI*) para comprender mejor cómo sus diferencias de diseño y naturaleza se reflejan en los resultados obtenidos y, sobre todo, en análisis realizados con JMetrik.

2.- Marco Teórico

2.1.- Preconceptos Erróneos

Este trabajo está centrado en la identificación de preconceptos erróneos (“*misconceptions*”) en el aprendizaje de la Física newtoniana, y en su evolución a lo largo de una asignatura de un semestre. Para esto se aplicará un pre test y un post test del Inventario de Conceptos de Fuerza *FCI*, al igual que la Evaluación Conceptual de Fuerza y Movimiento *FMCE*, cuyos resultados serán estudiados con análisis de gráficos no paramétricos producidos con el software JMetrik. Al analizar y comparar los resultados de ambas evaluaciones, se podrá no solo observar la manera en que el software puede ser utilizado para la detección de preconceptos erróneos, sino también cómo las diferencias de diseño entre los instrumentos de evaluación cambiará la manera en los que estos se presentan en los estudiantes y sus resultados en cada evaluación.

Los preconceptos erróneos son ideas, creencias o entendimientos incorrectos que presenta una persona acerca de un concepto, fenómeno o experiencia, que suelen estar basados en sus vivencias previas del mismo con una información incompleta que les permita explicar estos fenómenos. Al estar fundados en experiencias personales, estos preconceptos suelen estar profundamente arraigados e interferir con el aprendizaje de conceptos científicos o académicos nuevos que sean correctos y entren en conflicto con los preconceptos erróneos.

En la educación, los preconceptos erróneos son particularmente comunes en las ciencias y matemáticas, donde la intuición o la “lógica” de los estudiantes, basada en sus observaciones del mundo y sus fenómenos, puede diferir significativamente de las teorías científicas formales que expliquen el comportamiento de la realidad.

Un ejemplo clásico de preconcepto erróneo en Física es la confusión de los términos de aceleración y velocidad, que se da cuando un estudiante utiliza estos términos de manera intercambiable. Así, un estudiante con este preconcepto erróneo describiría la aceleración de un objeto en el punto más alto de un lanzamiento vertical como nula, en lugar de reconocer que su valor es constante a lo largo de todo su movimiento, mientras que es la velocidad la que sería cero en este punto. Otro preconcepto bastante frecuente es que los estudiantes no comprendan de manera correcta el principio de acción y reacción de la tercera ley de Newton,

por ejemplo, confundiendo la fuerza normal que ejerce una superficie horizontal sobre un objeto con el peso del mismo. Un cierto porcentaje de estudiantes tiende a pensar que estas dos fuerzas forman un par de acción-reacción, cuando en realidad se trata de fuerzas de diferente naturaleza ejercidas sobre un mismo cuerpo. Como último ejemplo, se puede mencionar el caso de un objeto en movimiento con rapidez constante, con el preconcepto erróneo de que este cuerpo debe estar bajo el efecto de una fuerza constante en el sentido de su desplazamiento, lo cual se contradice directamente con las leyes de Newton.

Más allá de ser una complicación para comprender e interpretar de manera correcta una situación física, los preconceptos erróneos pueden además ser obstáculos al momento de seguir construyendo el conocimiento y desarrollar las habilidades científicas y analíticas de un estudiante: es común que un estudiante puesto en una posición donde se ve obligado a trabajar sus preconceptos erróneos, elija encontrar maneras de justificar o adaptar sus propias ideas previas a la información nueva en lugar de reformular su propio entendimiento incompleto o incorrecto del tema. Para explicar los obstáculos que presentan los preconceptos erróneos para un aprendizaje eficaz, debemos estudiar cómo aprenden las personas, o al menos, ciertas corrientes de pensamiento que permitan comprender la manera en la que los preconceptos erróneos se relacionan con la construcción de nuevo conocimiento y los problemas que puede traer el generar conocimiento con preconceptos erróneos sin corregir.

2.1.1.- Conductismo y Constructivismo

En los últimos años se ha llegado al entendimiento de que el aprendizaje se consigue a través de un método denominado constructivista, en contraste con el anterior paradigma del aprendizaje, el conductista (Mestre & Docktor, 2021).

Para entender bien ambas posturas, primero se debe aceptar que las dos conciben el aprendizaje como el proceso a través el cual una persona forma conceptos y los trabaja y evoluciona a lo largo de su formación. La manera conductista de ver el aprendizaje busca abordar este a través de la separación de una tarea o conocimiento en diferentes procesos que compongan el concepto en su totalidad, y entonces enseñar cada etapa de manera independiente hasta que el estudiante sea capaz de replicar cada una de las fases y, finalmente, pueda llegar a hacerlo todo en conjunto. Los dos mayores problemas de esta manera de abordar la enseñanza es, primero, que pierde parte de los matices y detalles

intermedios de etapa a etapa, enfocando en su lugar el aprendizaje como si de una lista de tareas a cumplir se tratara; y segundo, que puede tender a mecanizar a los estudiantes, llevándolos a repetir procesos y pasos sin aplicar el pensamiento crítico ni el análisis que indicarían un auténtico aprendizaje.

En contraste, la postura constructivista entiende la adquisición de conocimiento como un proceso que se da en consecuencia a la formación de nuevas ideas a través de las experiencias y observaciones del individuo, sumándose entre sí y, sobre todo, comparándose y relacionándose con todas las ideas y conocimientos previos que el estudiante tenga de experiencias y observaciones anteriores. Así, el conocimiento vendría a venir de la suma de diferentes “bloques” conceptuales que una persona forma a lo largo de su vida y se interconectan poco a poco, a lo largo que las nuevas experiencias van causando contrastes, contradicciones y justificaciones entre ellos. De esta manera, un concepto erróneo bajo el lente constructivista sería un bloque de información que se encuentra mal asociado, conectado, o que fue obtenido a través de una interpretación errónea de una experiencia en la realidad. Volviendo al ejemplo anterior, un estudiante podría tener asociado el concepto erróneo de la aceleración de gravedad porque, a lo largo de toda su vida, siempre ha conectado que un objeto con velocidad cero *no se mueva*, y de similar manera, lo más probable es que se haya tomado con muchos ejemplos experimentales de objetos que, si poseen una aceleración igual a cero, tampoco se moverán, creando una correlación errónea entre ambos conceptos.

2.1.2.- Soluciones

Con la postura constructivista frente al conocimiento planteada, se puede cuestionar cómo abordar por fin los preconceptos erróneos para solucionarlos, y aunque esto sigue sin ser una tarea sencilla, existen dos posibles vías.

La primera busca producir un conflicto directo entre los conceptos erróneos del estudiante y los conceptos físicamente acertados que se quieren enseñar, empujando de esta manera a los estudiantes a un punto de quiebre, donde, o bien logran desprenderse del concepto erróneo y comienzan a asimilar el pensamiento científico correspondiente, o (y esta es una de las desventajas de este acercamiento) prevalece el concepto erróneo y el estudiante, en lugar de

asimilar el nuevo conocimiento, memoriza la información necesaria para abordar los problemas de la asignatura, pero continúa con su pensamiento erróneo en su día a día.

La segunda se aprovecha de la visión constructivista del conocimiento de manera más directa, aferrándose a la idea del conocimiento como un conjunto de bloques de información. De manera más específica, este acercamiento entiende no solo que los conceptos erróneos vienen de la formación constructivista del conocimiento, sino que se tratan, puntualmente, de bloques de información que pueden o no ser correctos, pero utilizados de forma incorrecta para resolver un problema. Usando nuevamente el ejemplo del lanzamiento vertical: el bloque de información mal utilizado sería el que le dice al estudiante que, si un objeto no se encuentra en movimiento, su aceleración será cero. Para solucionar este problema entonces, se busca, en lugar de reemplazar el bloque de información con el concepto correcto, enseñar al alumno el concepto científicamente acertado como un bloque de información nuevo, que el estudiante pueda utilizar de manera correcta para resolver problemas de este estilo, dejando de utilizar el preconcepto usado hasta ese momento. Aun así, esta metodología tiene su propio problema, y es que podría llevar a la coexistencia de conceptos erróneos e ideas científicamente correctas. Por ejemplo, un estudiante podría entender que todos los objetos caen a la misma velocidad *en ausencia de un roce*, pero seguir considerando que, en presencia de uno, los objetos más pesados caerán más rápido, en vez de asumir el primer caso como el general y el segundo como una excepción a la norma.

En la práctica, para poder sobrepasar los conceptos erróneos de manera efectiva se busca conseguir por parte de los estudiantes un aprendizaje significativo (Ausubel D. P., 1963), que es un concepto y modo de aprendizaje que sigue un paradigma en línea con el constructivismo, y se suele alcanzar a través de técnicas que sigan los principios de la enseñanza activa, constructiva y que fomente que cada estudiante se impulse por sus propios objetivos (Shuell, 1992).

2.2.- FCI y FMCE

Habiendo establecido el significado y el efecto de los conceptos erróneos en el aprendizaje, indicaremos el método específico que se usará en grupos masivos de estudiantes para reconocer un preconcepto en particular. Se usarán dos herramientas de evaluación, ambas centradas en la Física newtoniana, diseñadas con el objetivo de identificar los errores

conceptuales de los estudiantes al rendir las mismas: El *Force Concept Inventory*, y el *Force and Motion Conceptual Evaluation*. Ambas son evaluaciones de selección múltiple puramente conceptuales centradas en diferentes aspectos de la física Newtoniana, y tanto sus preguntas y sus posibles respuestas están elegidas con el objetivo de detectar los preconceptos erróneos que acarrearán los estudiantes. Aunque ambos test contienen preguntas que abordan los mismos temas, esto no quita que desde su concepción, las dos evaluaciones tengan enfoques diferentes (Thornton, Kuhl, Cummings, & Marx, 2009) y, por lo tanto, esta comparación será más que nada limitada a estos aspectos comunes entre los dos instrumentos evaluativos, llegando incluso a estudiar preguntas con el mismo enunciado y donde su única diferencia son las alternativas dadas y el formato cada evaluación para poder responderlas. Específicamente, el *FCI* presenta las preguntas y sus posibles respuestas en un formato de evaluación de selección múltiple tradicional, la pregunta seguida de cinco posibles respuestas correctas, mientras que explicar el *FMCE* requiere de más detalle.

En esta segunda evaluación, en lugar de presentarse al estudiante una serie de alternativas para cada pregunta, los temas se encuentran divididos por agrupaciones de preguntas correspondientes a una misma situación física. Por ejemplo, las primeras preguntas del *FMCE* tienen todas relación con un trineo moviéndose horizontalmente sobre una superficie de hielo con roce despreciable, y con una persona aplicando una fuerza externa al mismo. Notar que en esta primera descripción no se establece ni la dirección ni la cualidad del movimiento (acelerado, constante u otros detalles), sino que solo se plantean los elementos relevantes para el mismo. Luego, se describen las posibles alternativas a ser seleccionadas para *todas las siguientes preguntas*, que por lo general serán descripciones de fuerzas y su manera de relacionarse con los objetos involucrados. Por último, las preguntas en cuestión corresponden a la misma situación previa, pero en vez de ser general y sin detalle, asignándole magnitud, características y dirección a la velocidad y aceleración de los cuerpos, siendo entonces la manera de responder asignando cuál de las anteriormente descritas interacciones de fuerzas darían esta situación como resultado.

Si bien se estudiarán resultados obtenidos por diferentes cohortes, el análisis se centrará sobre todo en datos extraídos de grupos de estudiantes que rindieron el *FCI*, y por lo cual, es en esta evaluación en la que más se ahondará.

El *FCI* (Hestenes, Weiss y Swackhamer, 1992), consiste de treinta preguntas de alternativas múltiples, a través de las cuales se evalúa el concepto de fuerza en 6 dimensiones: cinemática, primera, segunda y tercera Leyes de Newton, tipos de fuerzas, y principio de superposición. A continuación, se describen estas seis dimensiones.:

- Cinemática. Discriminar velocidad de posición; aceleración de velocidad; relación entre aceleración constante con órbitas parabólicas y velocidades en constante crecimiento, suma vectorial de velocidades.
- Primera Ley de Newton, en ausencia de fuerzas y en presencia de fuerzas.
- Segunda Ley de Newton. Fuerzas de impulso, fuerzas constantes y aceleración asociada a estas.
- Tercera Ley de Newton, para fuerzas de impulso y constantes.
- Principio de Superposición. Suma de vectores, suma y cancelación de fuerzas.
- Tipos de Fuerza. De contacto sólido (pasiva, impulsiva o fricción en oposición al movimiento), de contacto fluido (resistencia del aire, de flotación) y de gravedad (aceleración independiente del peso; trayectoria parabólica).

Todos estos temas son abordados y cuestionados a lo largo de las treinta preguntas, cada una con cinco alternativas diferentes, donde cada alternativa a su vez está planteada con el objetivo de servir como distractor para ciertos estudiantes que se dejen llevar por ideas alternativas en específico, permitiendo así diferenciar qué preconcepto erróneo sobre la materia presentan los estudiantes dependiendo de qué alternativa errónea elijan para cada pregunta.

2.3.- Análisis de Datos en Evaluaciones de Selección Múltiple

La evaluación por sí sola, sin embargo, no es suficiente para poder analizar en profundidad los conceptos erróneos de un grupo de estudiantes. Si bien los puntajes obtenidos por cada estudiante y sus alternativas seleccionadas harán salir a flote individualmente sus preconceptos erróneos en el área de la física Newtoniana, esto no resulta de utilidad en el contexto de un profesor buscando retroalimentación en el conocimiento previo de un grupo de decenas de estudiantes, ni mucho menos a un nivel administrativo cuando se querrá evaluar el estado de toda una generación estudiantil que rinda una asignatura relacionada con

estos contenidos, cuando la dimensión de los datos se encontrará en los centenares. Es aquí cuando entran en acción las herramientas de análisis de datos.

La gran mayoría de acercamientos al análisis de evaluaciones de selección múltiple trabajan de manera similar (Ding & Beichner, 2009): se proponen un objetivo en concreto a conseguir, ya sea determinar la dificultad de la evaluación o la capacidad de la misma para discriminar de estudiantes que manejen o no el tema evaluado (como con *Classic Test Theory*), o poder agrupar a los estudiantes según sus diferentes rasgos y respuestas (*Cluster Analysis*), y luego plantean un algoritmo que permite obtener valores en base a los datos obtenidos de las respuestas de los estudiantes a la evaluación en cuestión; estos algoritmos, por supuesto, dependerán del objetivo específico que se quiera conseguir, y como tal, variarán según qué modelo de análisis se esté realizando. Así, se partirá de la base de tener que trabajar con grandes muestras de estudiantes, para aumentar así la precisión de los datos obtenidos, y el próximo paso es considerar qué modelo de análisis es el que más se ajusta al objetivo final, que en este caso es determinar en qué errores conceptuales e ideas alternativas caen con más frecuencia los estudiantes, si es que estos errores se corrigen entre el inicio y el final de una asignatura, y qué tanto cada estudiante mejora su nivel de habilidad relativo al finalizar el semestre.

2.3.1.- Item Response Theory

Dentro de los diferentes modelos de análisis de datos para este tipo de evaluaciones, se hará uso de la teoría de respuesta de ítems, o *IRT*, por sus siglas en inglés. La *IRT* tiene como objetivo estimar las curvas características de cada pregunta y el nivel de habilidad de los estudiantes (Livingston, 2020). La curva característica de una pregunta es una medida de la probabilidad de que, al elegir a un estudiante con cierto nivel de habilidad, su respuesta a esta pregunta sea correcta. De esta manera, la curva característica de una pregunta cuya dificultad se encuentre aproximadamente en el promedio, debería tener un cincuenta por ciento de probabilidad de que, al elegir un alumno con un nivel de habilidad promedio, este haya respondido correctamente.

Con este estilo de análisis no solo se obtendrá una curva que permita determinar la dificultad de las preguntas y la habilidad de los estudiantes pertenecientes a la muestra, sino que también se llegará a una lista de parámetros a tener en consideración según qué tipo de

modelo de *IRT* se esté utilizando. Existen tres listas de parámetros principales a tener en consideración, cada una de las cuales toma en cuenta más parámetros que la anterior a la hora de modelar las curvas de cada pregunta.

El modelo *IRT* de un parámetro, también conocido como Modelo Rasch, considerará solamente el parámetro que definiremos como “dificultad de la pregunta” a la hora de definir si un estudiante puede o no responder correctamente una pregunta. Luego, de manera progresiva, los modelos de dos y tres parámetros, añadirán la *discriminación* y la *facilidad para adivinar la respuesta correcta*, respectivamente. Aquí, la *discriminación* no solo considera lo “difícil” de escoger la alternativa correcta, sino que también la presencia de alternativas discriminantes, que permitan diferenciar estudiantes de diferentes niveles de habilidad, entre las posibles respuestas. El último parámetro se centra exclusivamente en lo fácil que podría ser, para un estudiante de poco nivel de habilidad en el tema, adivinar la respuesta, ya sea por chance o por descarte.

Si bien estos dos parámetros adicionales son importantes a la hora de hacer análisis de modelos *IRT* más profundos, el propósito de este estudio se centra menos en lo numérico y más en lo cualitativo, obteniendo curvas de respuestas para cada una de las preguntas, con cada una de las alternativas marcadas por cada estudiante en la evaluación al inicio y al final del semestre. De esta manera, se detectará qué grupos de estudiantes, entre qué valores de puntaje final (o nivel de habilidad) responden qué alternativa en específico, y cómo estas respuestas evolucionan a lo largo de la asignatura. Esto dejará en evidencia bajo qué valores los estudiantes caen en qué preconceptos erróneos con más frecuencia, sobre qué umbral de conocimiento estos errores desaparecen, y qué nuevos errores pueden o no existir solamente para estudiantes con un nivel de manejo de la materia ya superior al promedio. Esto también podría revelar preguntas de la evaluación cuyo fin último no haya sido alcanzado; alternativas incorrectas que no cumplen con su propósito de atraer a estudiantes con menor nivel de habilidad, o alternativas correctas que no son lo suficientemente claras como para que todos los estudiantes en los mayores tramos de habilidad respondan correctamente.

2.4.- jMetrik

jMetrik es el software escogido para este análisis. Este programa, creado como un software de psicometría, permite aplicar diferentes modelos y métodos de análisis de datos en

evaluaciones y test de variados tipos, entre ellos *IRT* de uno, dos, tres parámetros, o no-paramétricos. Debido al objetivo de esta investigación, se usará el software principalmente para este último tipo de análisis. Tras importar los datos de todas las respuestas marcadas por todos los estudiantes en una evaluación, y cargar las respuestas correctas de cada pregunta, el software jMetrik creará gráficas para cada una de las preguntas de la evaluación, cada una con las curvas de cada una de las respuestas marcadas por los estudiantes. Así se podrá visualizar con facilidad los conceptos erróneos más comunes de cada alumno, obteniendo gráficas que representarán el comportamiento de toda la muestra de estudiantes en cada una de las preguntas, y en cada tema en específico que busca abordar tanto el *FCI* como el *FMCE*.

Además, se puede aplicar a los datos un Modelo de Rasch o *IRT* de más parámetros para poder obtener, como mínimo, los valores modelados de dificultad de cada pregunta. Esto permite ahorrar tiempo de análisis, identificando específicamente las gráficas de las preguntas cuyo valor de dificultad, discriminación o facilidad para adivinar parezcan fuera de lugar en comparación con el resto de la evaluación.

2.4.1.- Curvas No-Paramétricas.

El tipo de análisis específico a utilizar en este estudio será la creación de gráficas no-paramétricas con jMetrik. Este tipo de curvas realizan una regresión lineal directamente de la tabla de datos inicial en lugar de aplicar un modelo previo, obteniendo así gráficas que, si bien son menos suaves y definidas, se acercarán más al comportamiento real de los estudiantes en la evaluación. De manera puntual, la forma de regresión que utiliza el software para crear las curvas plantea que la posibilidad de que un alumno de cierto nivel de habilidad elija una alternativa específica corresponderá a:

$$P_{jk}(\theta) = \frac{\sum_{i=1}^N K\left(\frac{x_i - \theta}{h}\right) \gamma_{ijk}}{\sum_{i=1}^N K\left(\frac{x_i - \theta}{h}\right)}$$

Donde K es la función Kernel que asigna un diferente peso a los estudiantes según su nivel de habilidad, aumentándolo para los estudiantes que se acerquen más al valor real de su puntaje obtenido para un estimado dado, y reduciéndolo para posibles *outliers* (*¿hay algún*

equivalente en español, o explicar el significado). Además, h es el “ancho de banda” de la función, que define la longitud de cada tramo de habilidad de los estudiantes que será aproximado a un único punto de este valor para la regresión lineal. Así, se obtendrán curvas más suaves para valores de h mayores, y líneas más precisas pero con menor suavidad con un h muy pequeño. Por último, el valor de θ corresponde a la estimación del nivel de habilidad de los estudiantes, extraído directamente del puntaje obtenido en la evaluación.

3.- Metodología

Con la información necesaria cubierta, se puede dar comienzo al respectivo análisis de las dos evaluaciones a tomar en consideración, aplicadas a distintos grupos de estudiantes en diferentes años, tanto para sus instancias pre y post dentro de la asignatura. De manera específica, se realizarán análisis de Modelo de Rasch y gráficas no-paramétricas para los resultados de muestras de entre cien y cuatrocientos estudiantes que rindieron la evaluación *FCI* en los años 2012, 2013, 2023 y 2024; y la *FMCE* en el año 2013. Se estudiarán los resultados obtenidos por los mismos grupos de estudiantes a inicios del semestre, aplicando la evaluación antes del inicio de las clases (categorizados como resultados PRE), y a finales del mismo semestre, sirviendo estos resultados como retroalimentación de la asignatura (categorizados como POST). Además, puntualmente para los resultados del *FCI* el año 2024 se realizará un análisis más fino, comparando, en lugar de los resultados de los estudiantes en función de su nivel de habilidad calculado en base a su puntaje final, las probabilidades de marcar una alternativa por los estudiantes al rendir la evaluación POST en función de sus niveles de habilidad previos al rendir la evaluación PRE. De esta manera, se obtendrán curvas representativas de las respuestas dadas por los estudiantes al finalizar el semestre, pero según los mismos niveles de habilidad iniciales de estos en la asignatura, pudiendo así hacer una relación uno a uno de los estudiantes y su aprendizaje de la materia en lugar de un estudio promedio de la muestra en su totalidad.

Realizar el análisis de cualquiera de estas muestras de datos y sus gráficas correspondientes se consigue con los mismos pasos, así que la descripción del proceso será generalizada, guardando los análisis puntuales para la próxima sección de este documento. Así, lo primero a tener en consideración es el conjunto de los datos requeridos para el análisis, y el formato en que estos se almacenan. jMetrik lee exclusivamente archivos de texto, donde los datos pueden estar separados por comas, tabulaciones, dos puntos o punto y coma, por lo cual hay que tener esto en consideración a la hora de exportar datos. Para este caso, todas las bases de datos se encuentran en formato Excel, por lo que solo hace falta exportar los datos necesarios a cualquiera de los compatibles con jMetrik.

Por otro lado, los datos necesarios para poder realizar los análisis deseados son: identificadores para los estudiantes (en caso de que se desee realizar alguna agrupación o

revisar datos individuales), una numeración arbitraria para las preguntas, y las alternativas de respuesta de cada uno de los estudiantes, para cada pregunta de la evaluación. De esta manera, deberíamos tener una tabla con la alternativa marcada por cada estudiante en cada pregunta, lo cual es lo justo y necesario para iniciar cualquier análisis dentro del software.

En primera instancia se creará una base de datos dentro de JMetrik, en la cual se podrán importar tantas tablas de datos como se desee. Para esta investigación, cada base de datos creada corresponderá con una evaluación específica y el año de su rendimiento, mientras que las tablas de datos serán las de las evaluaciones PRE o POST en el semestre correspondiente. A continuación, hay que añadir la forma de la evaluación. Esto se consigue en la sección *Transform* del programa, bajo la opción de *Advanced Item Scoring*, donde se podrá asignar para cada pregunta la respuesta correcta, el valor en puntaje que entrega la misma (asociando a la alternativa correcta 1 o más, según el puntaje que otorgue), las alternativas erróneas (asociadas a un valor de 0) y, opcionalmente, los valores para respuestas omitidas (que se tabulan como OM) o preguntas a partir de la cual el resto de la evaluación de un estudiante no fue completada (marcadas como NR). Una vez completada la puntuación para cada ítem, se puede actualizar la tabla de datos, que ahora tendrá en consideración los puntajes de cada alternativa, y como tal, permitirá aplicar a los mismos un Modelo de Rasch, o análisis *IRT* de más parámetros.

Aplicando a estos datos un análisis con Modelo de Rasch, se obtendrá una serie de datos y tablas. La primera será una tabla de valores para la dificultad de cada pregunta, el error estándar de este con respecto a los valores reales, y valores para la media estándar con o sin un peso variado de cada uno de los datos, así como el error de ambos valores. Dentro de esta tabla, la dificultad tendrá un valor que puede tomar valores positivos o negativos, donde una pregunta con una dificultad igual a cero presentará una probabilidad aproximada del cincuenta por ciento para que un estudiante con un nivel de habilidad media responda correctamente a dicha pregunta, aumentando esta probabilidad para valores más bajos, y disminuyendo para valores altos (así, una dificultad negativa indica más facilidad para responderla de manera correcta; y una positiva, una mayor dificultad). La siguiente tabla dará la transformación realizada por el modelo entre los puntajes obtenidos en la evaluación y el nivel de habilidad asignado al alumno, representado como un valor θ . Así, θ variará aproximadamente entre -3 y 3, y dependiendo de los valores y puntajes de la muestra de cada

evaluación, cada puntaje total le asignará a un estudiante un valor de θ puntual; siendo -3 el otorgado a quienes consigan el mínimo puntaje posible, y 3 para los máximos.

Por último, la tercera tabla de datos generada por el Modelo de Rasch entregará datos estadísticos respecto a los ajustes realizados por el modelo mismo; varianza, desviación estándar, error cuadrado medio, entre otros. Estos datos, si bien relevantes para motivos de análisis estadístico y para definir el ajuste a la realidad del modelo, no serán de mayor importancia para el resto del estudio.

Dentro de toda esta información obtenida gracias al Modelo de Rasch, el mayor énfasis estará en los datos de dificultad y habilidad de los estudiantes, ya que darán un indicativo previo de qué preguntas de la evaluación tuvieron un mayor número de estudiantes con problemas para responderla correctamente, lo cual reducirá el tiempo de análisis de los gráficos uno a uno hasta encontrar aquellos que sean destacables. Por otro lado, la tabla de niveles de habilidad de los estudiantes y sus valores de error estándar dirán hasta qué punto la distribución de puntajes y la aproximación de los niveles de habilidad es acertado.

Pasando ahora al análisis gráfico de los datos, se utilizará la información obtenida tanto de nuestras bases de datos como de los Modelos de Rasch para crear gráficos no-paramétricos de los mismos. Estas gráficos en cuestión tendrán como base el uso de una de las diferentes funciones Kernel disponibles para realizar la regresión lineal de los datos. Podemos usar diferentes funciones Kernel si queremos analizar distintas aproximaciones a los datos reales, pero usualmente es suficiente con una función Gaussiana (la opción determinada por defecto dentro del software), que sigue la forma:

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$$

Además, a la hora de configurar los gráficos no-paramétricos que se desean obtener, se usará un valor de ancho de banda modificado al predeterminado de 1, en su lugar asignando $h = 0.75$, lo que generará gráficos con curvas menos suaves, pero también más cercanas a los porcentajes reales de estudiantes que respondieron cada alternativa. Se podría variar este valor si lo que se desea conseguir es un análisis más representativo de la realidad, o por el contrario, si se busca un estudio que permita relacionar el comportamiento de los datos con

un modelo más predictivo. Antes de ejecutar el programa se debe tomar la precaución de marcar la opción correspondiente a graficar todas las posibles respuestas de cada pregunta, ya que con la configuración por defecto, jMetrik entregará solamente las curvas correspondientes a las alternativas correctas, y lo que se desea estudiar es el comportamiento tanto de las respuestas correctas como de todos los posibles preconceptos erróneos de cada pregunta.

Habiendo obtenido todas las gráficas deseadas, el último detalle a pulir es meramente visual, pero igualmente importante a la hora de interpretar las gráficas en comparaciones entre sí. Como podría esperarse de cualquier evaluación realizada de manera previa y posterior a cursar una asignatura, es lógico esperar que los resultados PRE sean drásticamente diferentes a los resultados POST. De esta manera, hay que cuidar que las gráficas obtenidas se encuentren en la escala correcta: jMetrik, por defecto, centrará la gráfica según los datos obtenidos, no según los puntajes mínimos y máximos de la evaluación. Esto quiere decir que, si el *FCI* tiene un máximo de treinta puntos y un mínimo de cero, pero ningún estudiante obtuvo el puntaje máximo, la gráfica se centrará usando como corte para el eje de habilidad el puntaje más alto obtenido entre los estudiantes. Esto es usual en evaluaciones previas y de diagnóstico, mientras que en los datos de POST suele ocurrir lo contrario: que, al no presentarse estudiantes con resultados mínimos, se obtengan gráficas donde el eje x inicia desde un valor mayor al cero. Para esto, tendremos que reconfigurar manualmente los ejes de las gráficas, lo cual dará espacios en blanco dentro de cada una, pues para ciertos intervalos de puntaje no existirán datos correspondientes. Esto no es un problema, pues el objetivo de mayor interés es comparar el desarrollo de los estudiantes a lo largo de la asignatura: que desaparezcan todos los estudiantes por debajo de cierto nivel de habilidad no resulta inesperado.

Aun así, hay información que se pierde en este proceso, y eso se debe a que no se sabrá qué grupos de estudiantes de cierto nivel de habilidad en las evaluaciones PRE corresponden a los mismos grupos en la evaluación POST. Para esto, lamentablemente jMetrik no tiene una opción correspondiente, y se deberá ajustar la base de datos utilizando un software alternativo para poder obtener la tabla deseada. Usando Excel se puede realizar una relación uno a uno de todos los estudiantes que dieron la misma evaluación a inicio y a final de semestre, haciendo uso del método de numeración de los estudiantes, ignorando los que

iniciaron el semestre dando la evaluación, pero no lo terminaron, o los que se sumaron a mitad del proceso educativo, y por lo tanto no tienen respuestas PRE. Una vez reconocidos todos los estudiantes que se repiten en la evaluación PRE y POST, se pueden tomar los datos de sus respuestas en la evaluación POST y asociarlos con sus puntajes totales obtenidos en la evaluación PRE. De esta manera, se conservarán los datos que jMetrik entrega como nivel de habilidad del estudiante al inicio de semestre, pero ahora obtendremos las respuestas que marcan al responder la misma evaluación una vez que se presentaron todos los contenidos de la asignatura.

Este análisis tendrá el mayor peso a la hora de hacer un seguimiento de manera más cercana o individual de los estudiantes, pues las variables de nuestro eje x no cambiarán. Si bien se sigue haciendo una regresión lineal con los datos, virtualmente, veremos en esta gráfica a los mismos estudiantes que en la gráfica de PRE responder la evaluación, después que sus conocimientos han sido adaptados y actualizados al pensamiento Newtoniano deseado, lo que permite ver directamente cómo los conceptos erróneos que cada uno de ellos presentaba ha evolucionado tras un semestre de estudio. En contraste, el gráfico de POST seguirá siendo de utilidad, pero más que como una herramienta directa de retroalimentación de los estudiantes, como una parte del análisis para la asignatura en su totalidad: indica, sin mayores matices, cuál es el nivel de habilidad promedio de todos los estudiantes al finalizar el semestre; y la tabla de puntajes PRE con las respuestas POST será útil para saber qué tanto evolucionó cada estudiante en su totalidad.

4.- Recopilación de Datos y Análisis

4.1.- FCI 2023: Análisis iniciales

Para comenzar a mostrar la utilidad de jMetrik, se observarán las gráficas no paramétricas para un grupo de estudiantes que rindieron la evaluación *FCI* el primer semestre del año 2023. Al realizar en una primera instancia los análisis de *IRT* de un parámetro, jMetrik generará las tablas de datos para la evaluación realizada al inicio del semestre (PRE) y al final del mismo (POST).

La tabla 1 muestra los valores ya mencionados que se obtiene tras realizar un análisis con modelo de Rasch a los datos de una evaluación, en este caso y de izquierda a derecha: numeración de la pregunta, dificultad, error estándar, error cuadrado medio considerando el peso de los datos y su variación estándar, y el error cuadrado medio sin considerar el peso de los datos y su variación estándar. De todos estos datos, resultan de mayor interés para el estudio del comportamiento de los estudiantes la dificultad, pues a menor el valor, más respuestas correctas se detectarán de parte de los estudiantes, y de la misma manera, si el valor es demasiado alto, se esperará encontrar un gran porcentaje de las respuestas cayendo en preconceptos erróneos típicos.

Item	Difficulty	Std. Error	WMS	Std. WMS	UMS	Std. UMS
r1	-0,26	0,15	0,98	-0,30	0,95	-0,64
r2	0,68	0,18	1,10	0,93	1,11	0,80
r3	-0,65	0,14	1,06	1,19	1,03	0,51
r4	0,12	0,15	1,08	1,01	1,13	1,21
r5	1,70	0,24	1,08	0,50	1,25	0,94
r6	-2,02	0,15	1,04	0,64	1,12	1,01
r7	-1,26	0,14	1,01	0,25	1,03	0,45
r8	-1,16	0,14	0,96	-0,86	1,05	0,73
r9	-0,11	0,15	1,13	1,93	1,23	2,45
r10	-0,29	0,15	0,90	-1,83	0,85	-1,92
r11	1,21	0,20	0,87	-0,95	0,78	-1,14
r12	-1,77	0,15	1,01	0,15	1,01	0,10
r13	1,05	0,19	0,89	-0,85	1,05	0,32
r14	-0,05	0,15	0,95	-0,77	0,96	-0,42
r15	0,45	0,17	1,09	1,03	1,06	0,49
r16	-0,18	0,15	1,00	0,01	0,99	-0,13
r17	1,30	0,21	0,99	-0,05	0,96	-0,13
r18	0,81	0,18	0,97	-0,20	0,96	-0,21
r19	-1,08	0,14	0,96	-0,88	0,92	-1,10
r20	-0,63	0,14	0,93	-1,59	0,88	-1,73
r21	0,65	0,17	0,98	-0,11	0,95	-0,28
r22	-0,14	0,15	1,04	0,63	1,07	0,80
r23	-0,05	0,15	1,01	0,21	0,98	-0,17
r24	-0,99	0,14	0,97	-0,80	0,94	-0,91
r25	1,02	0,19	1,05	0,45	1,05	0,33
r26	2,76	0,35	1,05	0,26	0,93	0,01
r27	-0,89	0,14	1,00	0,09	0,98	-0,31
r28	0,42	0,16	0,90	-1,10	0,83	-1,44
r29	-1,70	0,15	1,05	0,83	1,24	2,26
r30	1,05	0,19	0,95	-0,33	1,03	0,21

Tabla 1: FCI2023-1 PRE, Análisis con Modelo Rasch

Se nota de inmediato ciertas preguntas con un valor de dificultad mayor que las demás en la Tabla 1, como las pregunta 5, 13, 17, 25, 26 y 30, siendo las de mayor interés las preguntas

5 y 26, ya que al ser las de mayor valor de dificultad deberían mostrar un comportamiento más marcado hacia una o más alternativas que indiquen un error conceptual en concreto.

La tabla 2 indica el mismo Modelo de Rasch aplicado a la misma evaluación, pero en su instancia POST, por lo que ahora se podrán comparar los valores de dificultad obtenidos en la tabla 1 con la tabla 2 y ver la evolución de estos valores para detectar así cambios de interés en el comportamiento de las respuestas de los estudiantes.

Item	Difficulty	Std. Error	WMS	Std. WMS	UMS	Std. UMS
r1	0,18	0,18	0,99	-0,16	0,94	-0,46
r2	1,36	0,22	1,05	0,41	1,10	0,54
r3	-0,63	0,18	1,21	3,07	1,20	1,48
r4	0,18	0,18	1,00	0,03	0,98	-0,17
r5	1,08	0,21	1,03	0,34	1,11	0,66
r6	-1,49	0,20	1,10	1,07	1,23	1,11
r7	-1,23	0,19	1,14	1,68	1,25	1,38
r8	-1,23	0,19	1,13	1,67	1,25	1,38
r9	0,18	0,18	1,04	0,61	1,01	0,11
r10	-0,53	0,18	0,87	-2,20	0,83	-1,46
r11	-0,28	0,18	0,94	-0,90	0,89	-1,00
r12	-2,24	0,23	0,93	-0,43	0,79	-0,63
r13	0,63	0,19	0,86	-1,59	0,81	-1,39
r14	0,18	0,18	1,07	1,00	1,18	1,50
r15	0,67	0,19	1,32	3,17	1,53	3,28
r16	-0,37	0,18	0,87	-2,15	0,82	-1,64
r17	0,99	0,21	1,17	1,48	1,32	1,78
r18	0,04	0,18	1,07	1,02	1,18	1,51
r19	-0,60	0,18	0,88	-1,91	0,89	-0,89
r20	-0,92	0,18	0,87	-2,03	0,78	-1,53
r21	0,83	0,20	0,92	-0,74	0,85	-0,95
r22	0,38	0,19	0,94	-0,74	1,00	0,08
r23	0,63	0,19	0,90	-1,11	0,80	-1,51
r24	-0,73	0,18	0,87	-2,07	0,79	-1,63
r25	1,12	0,21	0,93	-0,58	0,84	-0,88
r26	2,15	0,28	0,88	-0,56	0,81	-0,50
r27	-0,41	0,18	1,00	0,00	0,99	-0,00
r28	0,63	0,19	0,93	-0,82	0,89	-0,81
r29	-1,45	0,19	1,04	0,50	1,11	0,60
r30	0,87	0,20	0,95	-0,46	0,86	-0,89

Tabla 2: FCI2023-1 POST, Análisis con Modelo Rasch

Por ejemplo, al realizar esta comparación se evidencia un comportamiento interesante en la pregunta 2, ya que la dificultad de la misma aumenta entre el PRE y el POST. Esto no es imposible ni necesariamente extraño, pero la magnitud del valor dificultad, que alcanza casi el doble, es llamativo, por lo que también deberá ser considerado al realizar las gráficas no paramétricas.

Es relevante también notar que, para la muestra completa de estudiantes en la evaluación PRE de este año del FCI, no se obtuvo ningún valor máximo entre los estudiantes; y en el POST, si bien se registró un único puntaje de treinta respuestas correctas, el número tan bajo hace imposible que la regresión lineal de los datos considere puntos para todo el espectro de valores de habilidad de los estudiantes. A pesar de esto, se puede graficar el comportamiento no-paramétrico de las respuestas dadas por los estudiantes en algunas de las preguntas de interés según las próximas imágenes; sin embargo, antes de adentrar en aquellas que llaman la atención por su comportamiento extraño, se utilizará de ejemplo la pregunta 11, que

presenta un nivel bajo de dificultad, para poder poner en práctica las múltiples herramientas hasta ahora mencionadas a prueba.

Al momento de analizar la pregunta y todas sus alternativas, es fácil detectar los preconceptos erróneos que busca sacar a flote: la pregunta 11 plantea la situación de un objeto en movimiento a velocidad constante sobre una superficie sin roce, y pide identificar las fuerzas actuando sobre el objeto. Aquí, la alternativa C menciona una fuerza en el sentido del movimiento, un concepto erróneo que fue planteado como un ejemplo en este mismo informe, a la vez que las dos fuerzas que realmente actúan en tal movimiento, la gravedad y la normal. Con esto en cuenta, se puede realizar el análisis correspondiente de las curvas no-paramétricas de esta pregunta.

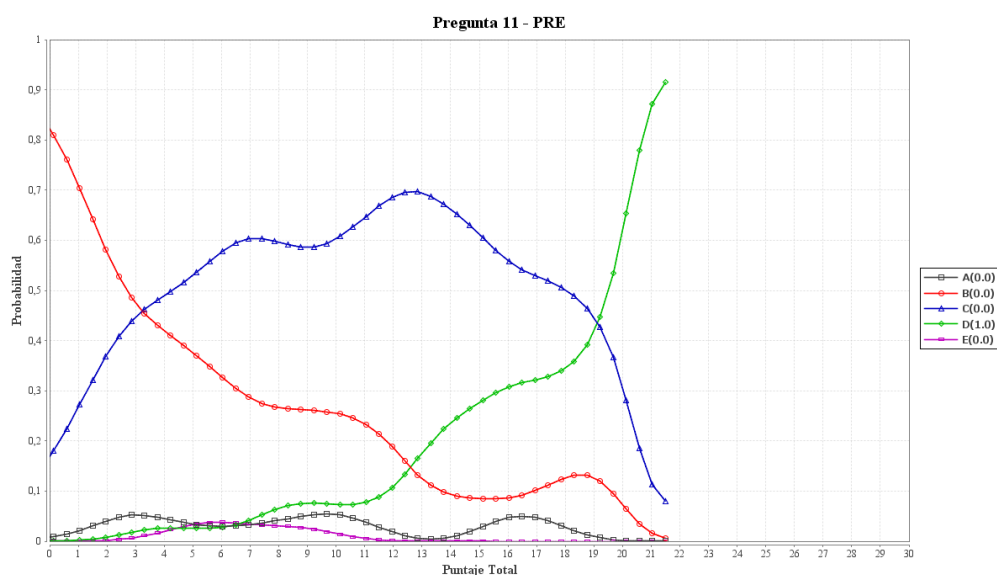


Figura 1: Gráfico Pregunta 11 PRE FCI2023-1

Aquí se puede evidenciar de inmediato el efecto de la diferencia de nota promedio, así como también la ausencia de una nota final máxima en la evaluación PRE, al ver que el gráfico obtenido en la *Figura 1* no presenta valores pasado el puntaje total de los estudiantes de aproximadamente 22, mientras que en la *Figura 2* para estos valores más altos los datos no solo se extienden hasta casi 30, sino que pasados los 24 puntos, se estabilizan, marcando la mucho más alta probabilidad para los estudiantes en este nivel de habilidad de responder la respuesta correcta una vez atravesado el semestre.

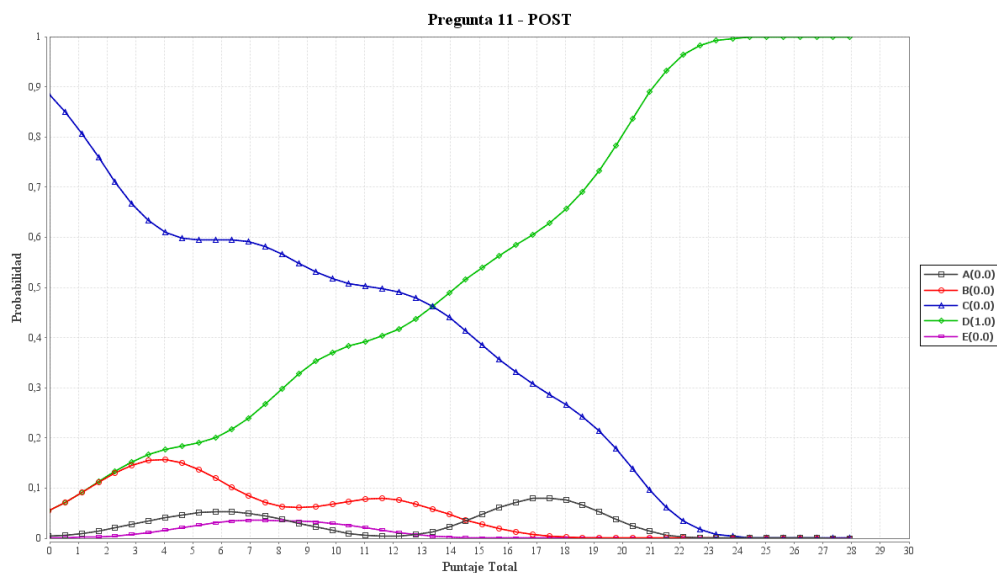


Figura 2: Gráfico Pregunta 11 POST FCI2023-1

En la *Figura 1*, además, es evidente que para el grupo de habilidad intermedio-bajo la alternativa C es una idea alternativa que les resulta atractiva. En el gráfico POST se evidencia que esta atracción desaparece gradualmente a medida que aumenta el nivel de habilidad de los estudiantes.

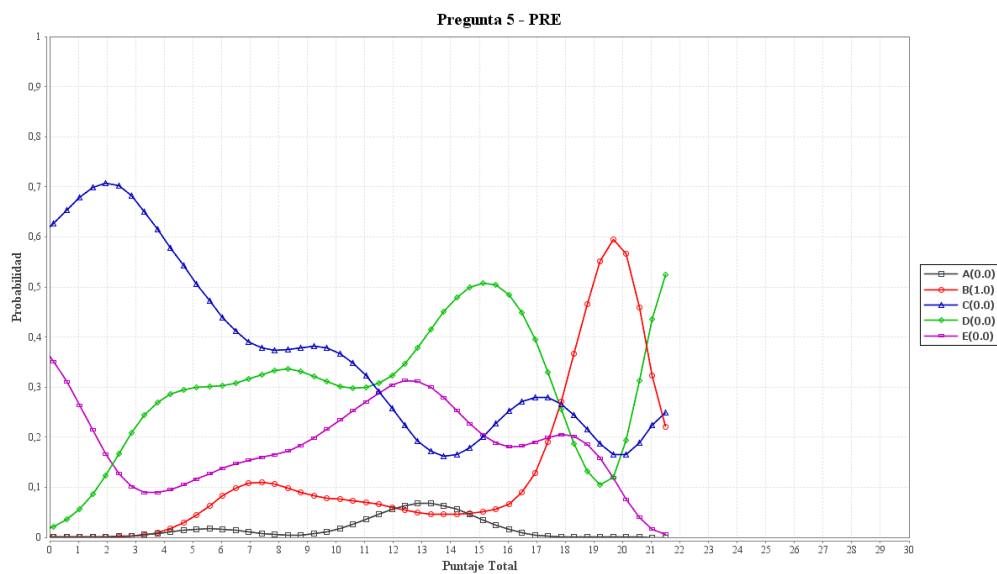


Figura 3: Gráfico Pregunta 5 PRE FCI2023-1

En cuanto a las dos preguntas con un aumento de dificultad, en la *Figura 3* notamos de inmediato cómo en la pregunta 5 existen dos preconceptos erróneos con alto grado de atracción para los estudiantes; en los valores más bajos de habilidad la alternativa C; y en los

intermedios, la alternativa D. Esta pregunta plantea el movimiento de una bola sobre una superficie, siguiendo el camino de una canaleta circular, y por lo tanto, trazando una circunferencia con centro en un punto “O” en su trayectoria. Luego, la evaluación pide identificar las fuerzas haciendo efecto sobre la bola, donde la alternativa D reconoce la fuerza de gravedad y la fuerza que apunta hacia el centro de la circunferencia, pero añade una tercera fuerza que va a favor del movimiento; y la alternativa C, además de considerar también esta tercera fuerza inexistente, ignora la fuerza hacia el centro “O”.

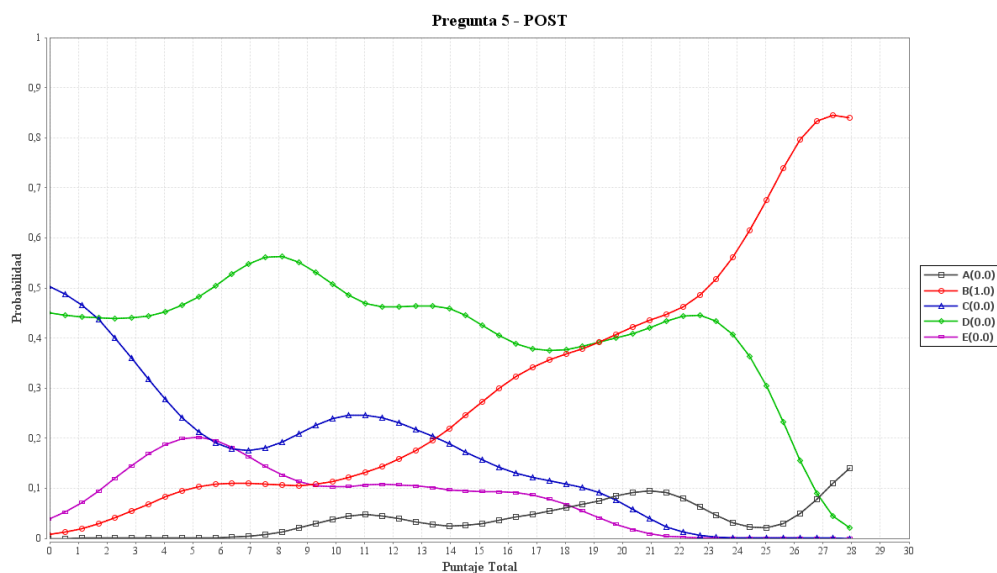


Figura 4: Gráfico Pregunta 5 POST FCI2023-1

La alternativa correcta, B, de esta pregunta identifica la fuerza de gravedad y la fuerza centrípeta del movimiento (como entre las alternativas iniciales no se exige detectar la fuerza normal, esta es irrelevante), y se esperaría que una vez pasada la asignatura estas gráficas mejoren, y si bien la tendencia a responder la alternativa correcta es claramente mayor, en la Figura 4 es evidente que una de las dos ideas alternativas previas gana presencia, y esta corresponde a la alternativa D.

No es difícil realizar el análisis de que esto se debe a que, una vez cubierto el contenido de la asignatura, los estudiantes comprenden con éxito la presencia de fuerzas centrales en movimientos curvilíneos, pero siguen sin poder deshacerse del todo de la idea errónea de que existe en todo momento una fuerza con la misma dirección que el movimiento de un cuerpo con velocidad constante.

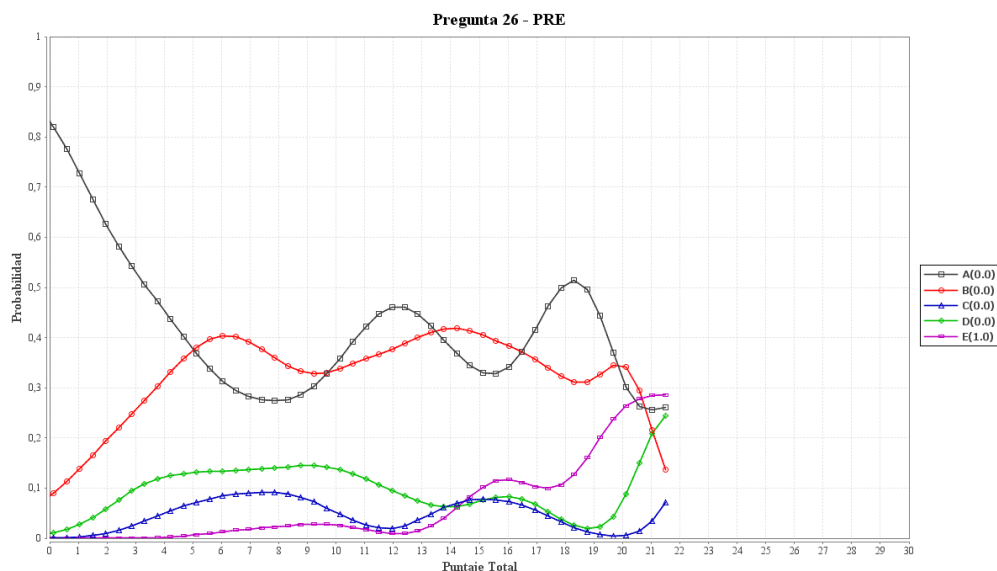


Figura 5: Gráfico Pregunta 26 PRE FCI2023-1

Pasando a la siguiente pregunta de interés, en la *Figura 5* se observa un comportamiento similar al caso anterior, donde dos conceptos erróneos atraen a gran parte de los estudiantes con un nivel de habilidad por debajo del punto medio y, a medida que los valores de puntaje se acercan a los máximos conseguidos, estos errores conceptuales se reducen y en su lugar la respuesta correcta gana presencia. A diferencia de la pregunta 5, sin embargo, aquí las dos ideas alternativas más prominentes, las alternativas A y B, comparten presencia aproximadamente en los mismos rangos de habilidad de los estudiantes, y se mantienen en una magnitud similar incluso una vez terminada la asignatura, como se evidencia en la *Figura 6*, en vez de que solo uno de los dos se mantenga presente en estudiantes que hayan atravesado la asignatura. Además, a diferencia de las dos gráficas anteriores, aquí en la evaluación POST los estudiantes que respondieron la pregunta 26 presentaron un nuevo error conceptual importante para el grupo de menor habilidad, ganando la alternativa C bastante presencia en el rango de puntaje total por debajo de los 5 puntos y luego nuevamente en los valores intermedios.

Por último, es de importancia notar que los valores de puntaje total para los que la alternativa correcta se vuelve predominante son prácticamente los mismos. Esto indica que, si bien hay más estudiantes con mejor manejo de la asignatura, el umbral de conocimiento requerido para detectar la respuesta correcta de la pregunta no ha variado.

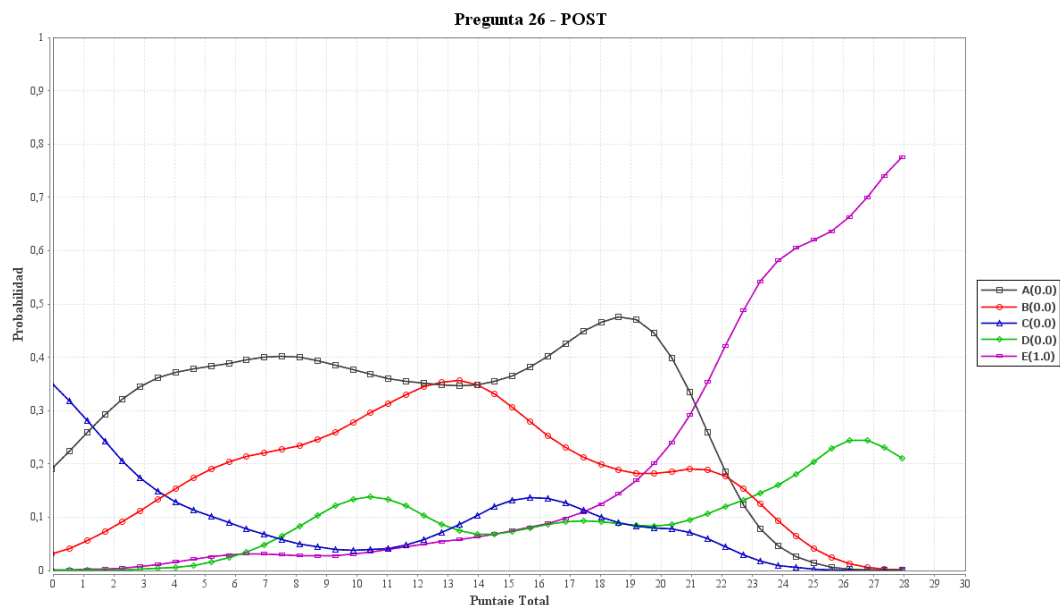


Figura 6: Gráfico Pregunta 26 POST FCI2023-1

La pregunta en cuestión plantea una situación con una fuerza externa siendo aplicada sobre un objeto en movimiento. La pregunta justo anterior plantea una velocidad de movimiento constante como consecuencia de la fuerza externa, para, en este caso, *duplicar* dicha fuerza, y luego se pregunta cuál será la velocidad resultante. Ahora resulta evidente, al ser la alternativa A la que plantea una velocidad constante igual al doble que en el caso anterior, y la B la que plantea otra velocidad constante mayor pero no necesariamente el doble, el por qué estas resultan atractivas para los estudiantes de un nivel de habilidad menor; al encontrarse con una fuerza externa constante de empuje dan por asumido que la velocidad del movimiento empujado no podrá ser variable, ignorando que toda fuerza que acabe con una sumatoria total mayor a cero sobre un objeto producirá una *aceleración*, en lugar de rapidez.

4.2.- FCI 2012 v/s FMCE 2013

Pasando a la comparación entre las dos evaluaciones que se tienen para el análisis de conceptos erróneos, se tienen datos de dos años cercanos del *FCI* y el *FMCE*. Tras analizar los modelos de Rasch de ambas evaluaciones, y repasando con cuidado todas las preguntas de cada evaluación, se encuentran dos grupos de pregunta que son casi la misma situación, y además reflejan cierto comportamiento interesante por parte de los estudiantes. La primera

de estas dos es la pregunta número 4 en el *FCI*, y la serie de preguntas de la 30 a la 34 en el *FMCE*. Aquí es útil recordar la diferencia que existe entre cómo son planteadas las preguntas del *FMCE* en comparación con las del *FCI*, pero principalmente el hecho de que la primera de estas evaluaciones presentará más alternativas y, por lo mismo, una capacidad de detectar preconceptos erróneos más puntuales.

Entonces, en la pregunta 4 del *FCI* se describe la colisión entre un camión y un automóvil, donde el camión tiene más masa que el automóvil, y se pide a los estudiantes reconocer correctamente la relación entre las magnitudes de las fuerzas producidas por cada cuerpo sobre el otro, durante la colisión: la respuesta correcta es que, de acuerdo con la tercera ley de Newton, las magnitudes de ambas fuerzas son iguales, independientemente de las masas de ambos cuerpos.

El *FMCE* va más allá, y plantea cinco preguntas sobre el mismo problema, con pequeñas variaciones. Para la primera instancia, la situación es exactamente la misma, solo que especifica que ambos vehículos se mueven a la misma velocidad (que no es relevante para esta primera pregunta, pero que se utilizará en las siguientes preguntas).

Así, al comparar esta primera especificación de la misma situación para ambas evaluaciones en su instancia PRE, en *Figura 7* y *Figura 8* observamos un comportamiento muy similar para los mismos conceptos erróneos, ya que la alternativa A de ambas evaluaciones es la que plantea que el camión, al poseer una mayor masa, ejercerá la mayor cantidad de fuerza sobre el automóvil. Aun así, también es importante notar la diferencia del punto donde la cantidad de estudiantes cayendo en el error conceptual de que a mayor masa se ejerce una mayor fuerza en la colisión, es drásticamente diferente para ambas evaluaciones, encontrándose en el *FCI* por debajo del punto intermedio de habilidad, mientras que para el *FMCE* se encuentra con aproximadamente el mismo desplazamiento de la mediana de puntaje total, pero hacia los niveles de habilidad más altos.

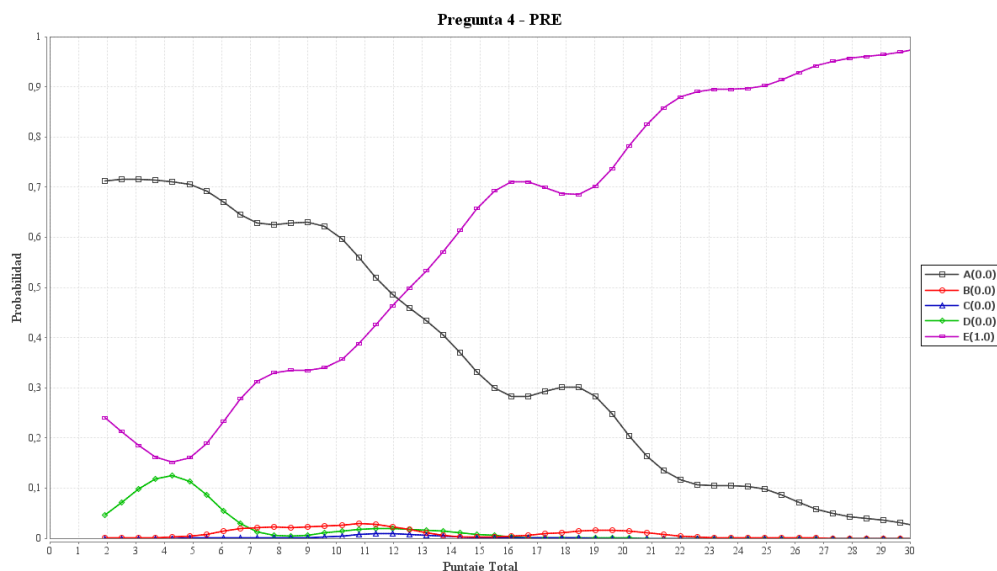


Figura 7: Gráfico Pregunta 4 PRE FCI 2012-2

Es importante también mencionar que, mientras en el *FCI* se evidencia la aparición esporádica de otra idea alternativa en la opción D (que es muy baja para considerarla con más peso), en dos puntos de la gráfica del *FMCE* es notoria la alternativa F; esto se explica bastante fácil, ya que es la respuesta que se asigna para cuando los estudiantes consideran que el enunciado no otorga suficiente información para dar una respuesta concisa. Es importante también mencionar que, mientras en el *FCI* se evidencia la aparición esporádica de otra idea alternativa en la opción D (que es muy baja para considerarla con más peso), en dos puntos de la gráfica del *FMCE* es notoria la alternativa F; esto se explica bastante fácil, ya que es la respuesta que se asigna para cuando los estudiantes consideran que el enunciado no otorga suficiente información para dar una respuesta concisa.

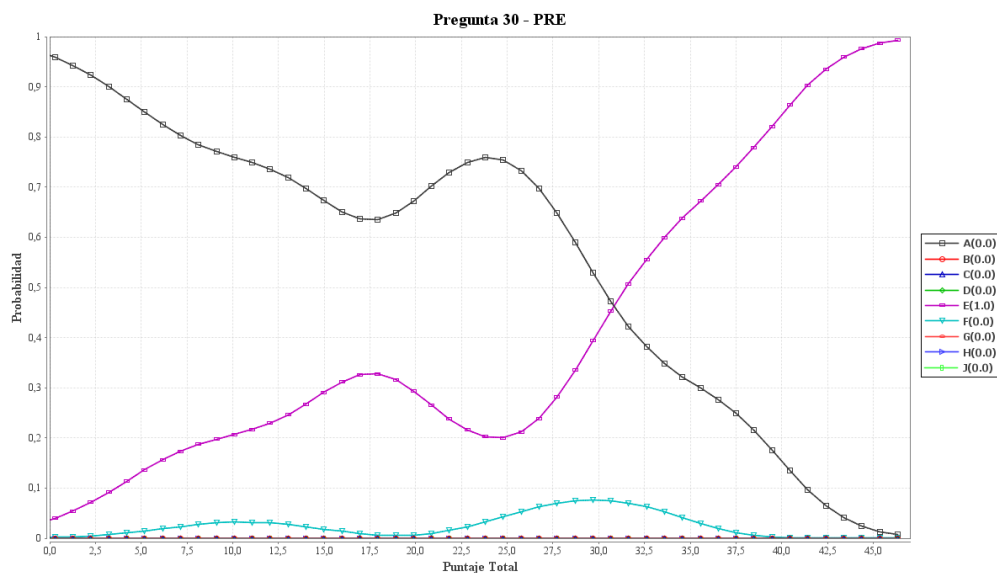


Figura 8: Gráfico Preguntar 30 PRE FMCE 2013-2

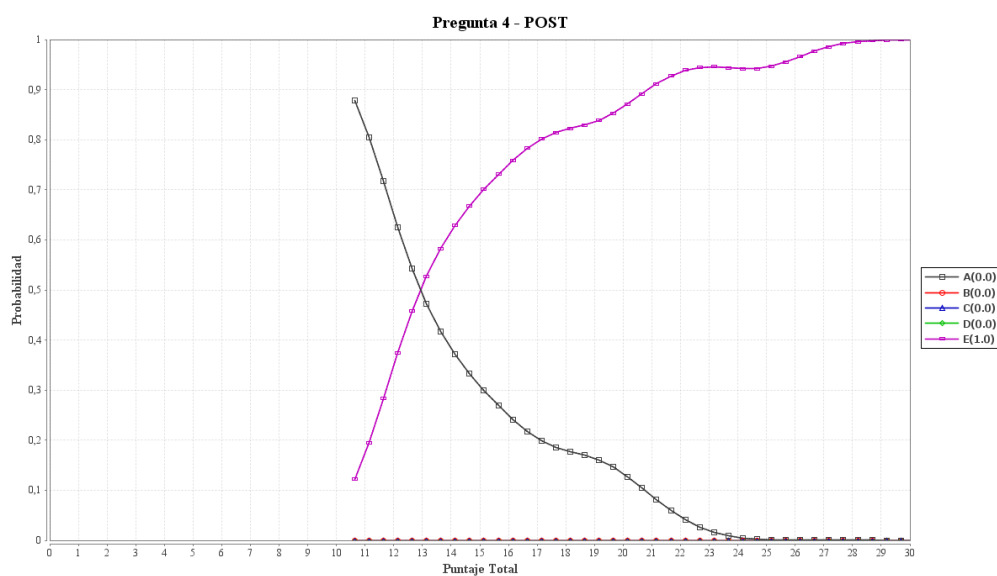


Figura 9: Gráfico Preguntar 4 POST FCI 2012-2

Estudiando los gráficos POST (9 y 10) tampoco se encuentran muchas sorpresas: los resultados más altos de puntaje total mínimo producen un desplazamiento general de los datos, pero dentro de estos el comportamiento es el esperado: la idea alternativa resulta menos atractiva para estudiantes con menor nivel de habilidad, y el umbral mínimo para el cual la respuesta mayoritaria es la correcta se reduce; incluso se mantienen las diferencias de estos umbrales en las gráficas de la evaluación POST; estando el punto de “intercambio” del

concepto erróneo con la alternativa correcta más bajo para la pregunta en el *FCI*, y más alto para el *FMCE*.

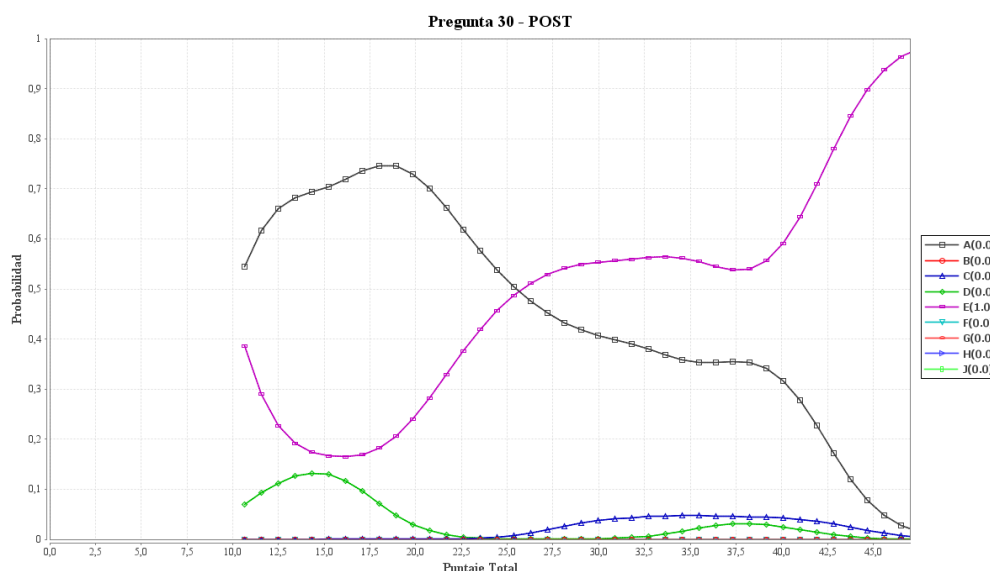


Figura 10: Gráfico Pregunta 30 POST FMCE 2013-2

En contraste con este caso, la pregunta 32 de la evaluación *FMCE* establece que el camión se encuentra detenido y el automóvil es el causante del choque, para luego dar las mismas opciones de posibles relaciones entre las fuerzas ejercidas por ambos cuerpos. Aquí, a pesar de que la variación en el planteamiento no cambia de ninguna manera ni la respuesta correcta ni el motivo por el cual la lógica de los estudiantes deberían llegar a esta respuesta, las gráficas obtenidas por los estudiantes enfrentándose a esta pregunta resultan interesantes.

En la *Figura 11*, con la evaluación PRE, se nota de inmediato que, a diferencia de la manera anterior de plantear la pregunta, esta situación produce mucha más confusión y los estudiantes caen en varias alternativas en lugar del patrón anterior centrado en la respuesta correcta y una sola idea alternativa predominante. Para empezar, el error conceptual con más fuerza en los estudiantes de menor nivel de habilidad, en lugar de ser aquel que confunde la masa de un objeto con la fuerza que ejerce (A), aquí es el que confunde velocidad con fuerza, pues la alternativa B dice que el automóvil ejerce una mayor fuerza sobre el camión. Los resultados POST, también, a diferencia del caso anterior, no evidencian una claridad en los términos poco comprendidos por los estudiantes de la misma manera, pues sigue existiendo incluso una vez pasada la asignatura una cantidad igualmente considerable de respuestas en el resto de alternativas.

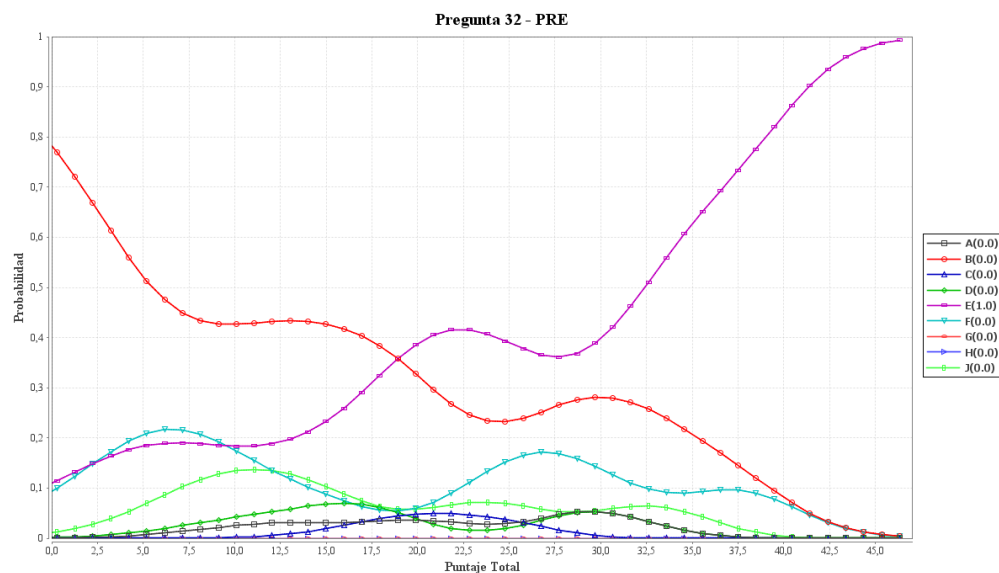


Figura 11: Gráfico Pregunt 32 PRE FMCE 2013-2

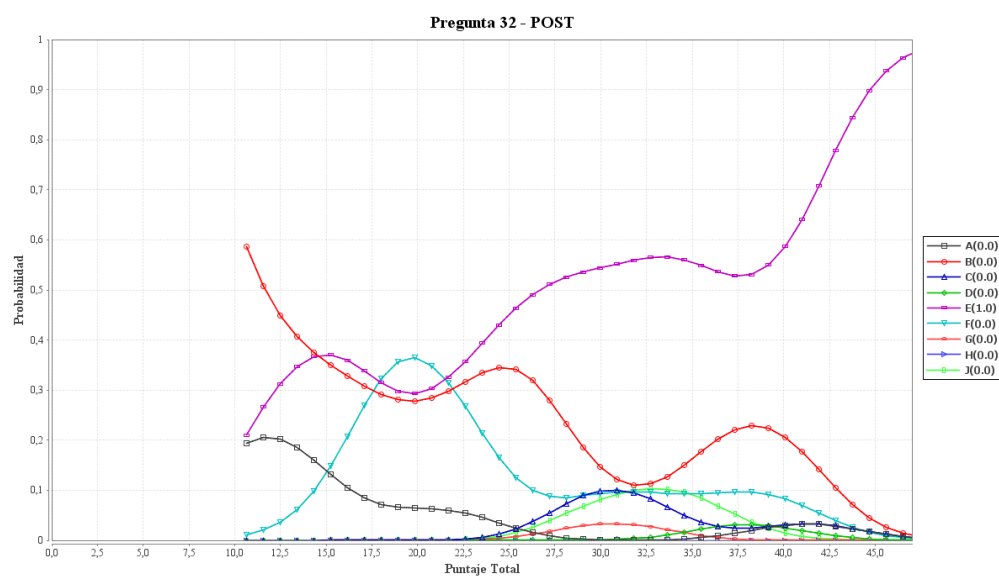


Figura 12: Gráfico Pregunt 32 POST FMCE 2013-2

Por último, en lo que a esta situación se refiere, la pregunta 33 del *FMCE* le da otro giro a la situación, planteando el choque como uno de dos cuerpos de la misma masa y moviéndose a la misma velocidad (un automóvil y una camioneta, esta vez del mismo tamaño que el otro vehículo). Como podría esperarse, en la *Figura 13* no se obtienen resultados muy reveladores, ya que al eliminar los dos factores externos que en casos anteriores causaban confusión en los estudiantes (diferencia de masa y de velocidad), no es extraño que los

resultados obtenidos incluso antes de que el grupo de estudio pase los contenidos de la asignatura sepan encontrar la alternativa correcta por mera intuición.

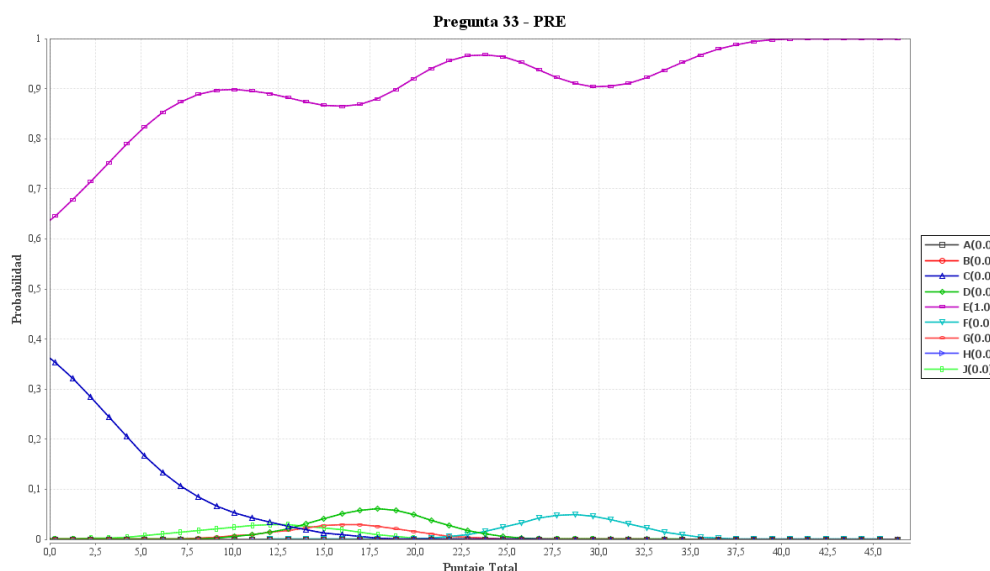


Figura 13: Gráfico Pregunta 33 PRE FMCE 2013-2

La segunda pregunta que es compartida entre ambas evaluaciones corresponde a la número 28 en el *FCI* y la número 39 en el *FMCE*. A diferencia del caso anterior, esta pregunta es la misma tanto a nivel de redacción como posibles respuestas, con la única diferencia para el *FMCE* de que esta evaluación, por su concepción misma, presenta más alternativas, que permiten tanto dar la respuesta ya mencionada antes de que la pregunta no da la información suficiente para encontrar una respuesta, como también una de “ninguna de las alternativas anteriores”. Esta pregunta describe una situación con dos estudiantes, uno con mayor masa que el otro, sentados uno frente al otro sobre sillas de escritorio idénticas. Entonces, el estudiante con más masa se apoya en el otro para empujarse. El enunciado entonces pregunta por las fuerzas ejercidas entre los estudiantes *mientras* siguen en contacto a lo largo del empuje.

El mayor logro que tiene el *FCI* respecto a esta pregunta es, con diferencia, la reducción de las pequeñas confusiones y conceptos erróneos de menor prominencia en los niveles de habilidad más bajos de los estudiantes; en la *Figura 14* se nota de inmediato que casi todas las respuestas posibles tienen cierto grado de aparición por debajo de los 12 puntos, a pesar de la mayoría que prefiere hasta la mitad del rango de valores la alternativa D.

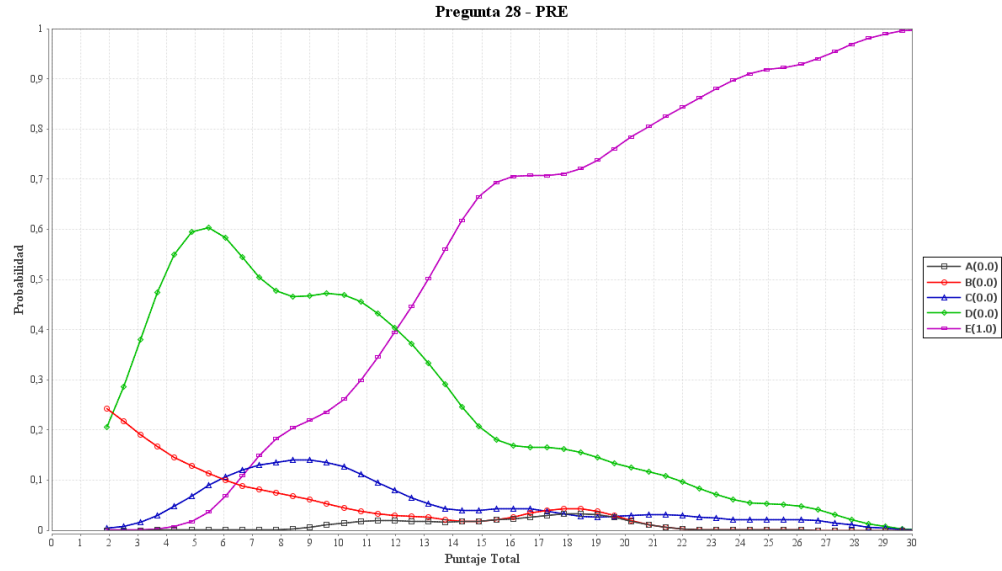


Figura 14: Gráfico Pregunta 28 PRE FCI 2012-2

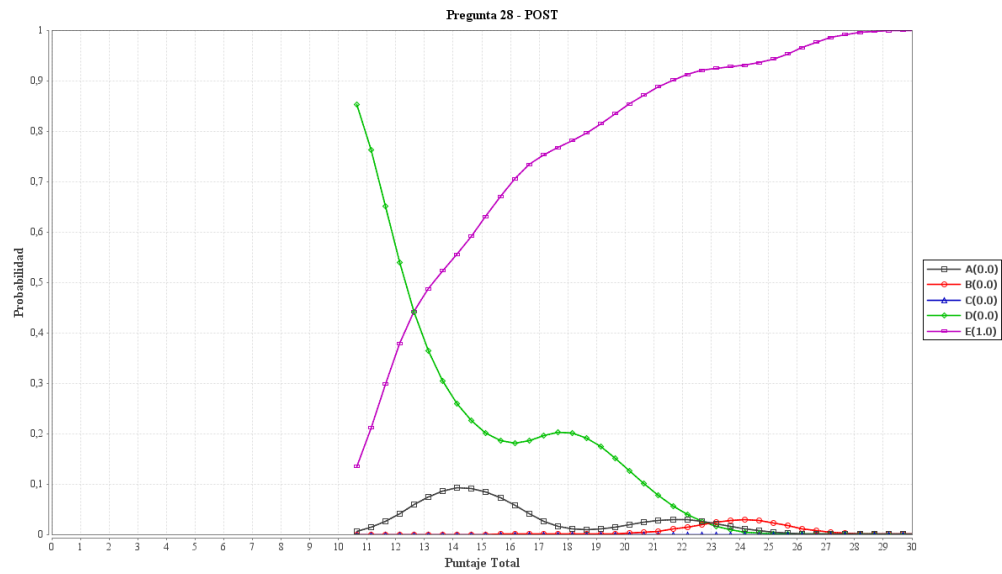


Figura 15: Gráfico Pregunta 28 POST FCI 2012-2

En la *Figura 15*, por otro lado, casi toda la preferencia por las otras alternativas se desvanece, y el decaimiento de las alternativas D marcadas es mucho más abrupto. En contraste con esto, en la *Figura 16* es notoria la alternativa H como la prominente en los valores más bajos de habilidad, y el decaimiento de la alternativa D es mucho más suave que siquiera en los resultados PRE del *FCI*.

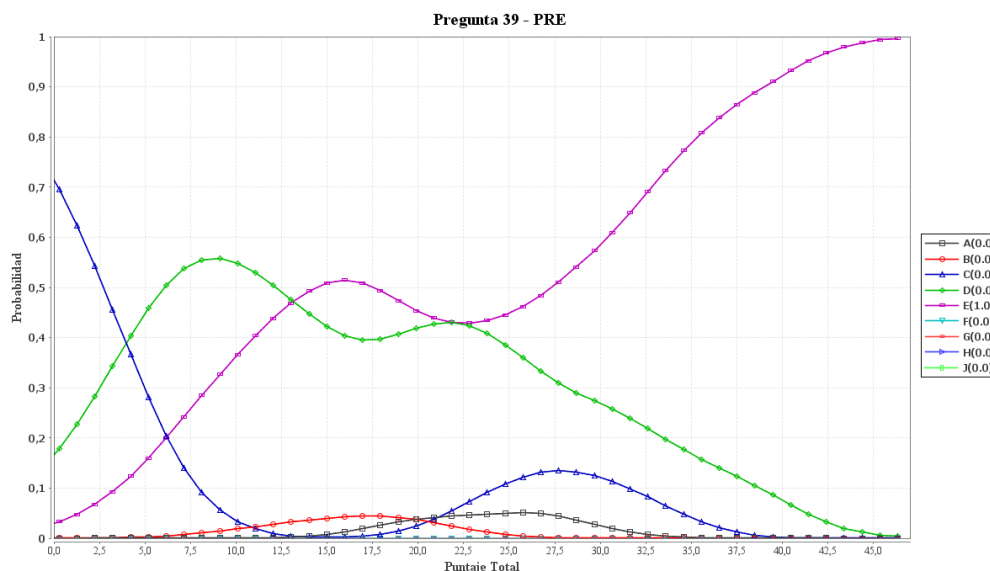


Figura 16: Gráfico Pregunta 39 PRE FMCE 2013-2

Otro detalle que notar es que en la *Figura 17* gana presencia una nueva idea alternativa en la forma de la alternativa A, una que en la evaluación PRE ni siquiera había aparecido con la excepción de una probabilidad muy pequeña en los valores de habilidad medios de los estudiantes. Para entender esto, se deberá realizar el mismo análisis a las alternativas en cuestión.

Siendo la alternativa correcta la misma para ambas evaluaciones (E), y la alternativa errónea con más porcentaje de respuestas la D, esta cae en el mismo error que el ejemplo de la colisión de vehículos, pues nuevamente es la respuesta que señala que el estudiante de mayor masa ejercerá una mayor fuerza, a pesar de reconocer que el otro también deberá ejercer una en el par acción-reacción. Por otro lado, la alternativa A de la pregunta en el *FMCE* corresponde a que ninguno de los estudiantes ejerce una fuerza sobre el otro. Es interesante la aparición de este concepto erróneo, pues podría indicar una comprensión nueva de la idea del principio de la superposición de fuerzas, sumado a una aplicación errónea del mismo en conjunto con la segunda ley de Newton; en otras palabras, podría deberse a un grupo de estudiantes entendiendo que, como un estudiante está realizando un empuje sobre el otro, ambos están ejerciendo dos fuerzas en sentidos opuestos de igual magnitud, y por lo tanto, la fuerza neta que se ejerce en el sistema sería cero.

A pesar de que esta aparición de un nuevo concepto erróneo es en baja cantidad, no deja de ser llamativo y, en algunos casos, preocupante. Podría ser de utilidad analizar con más

tiempo los casos de estos estudiantes para determinar el motivo de que esta idea alternativa haya surgido una vez pasado el contenido de la asignatura, cuando su presencia previa al mismo era tan reducida.

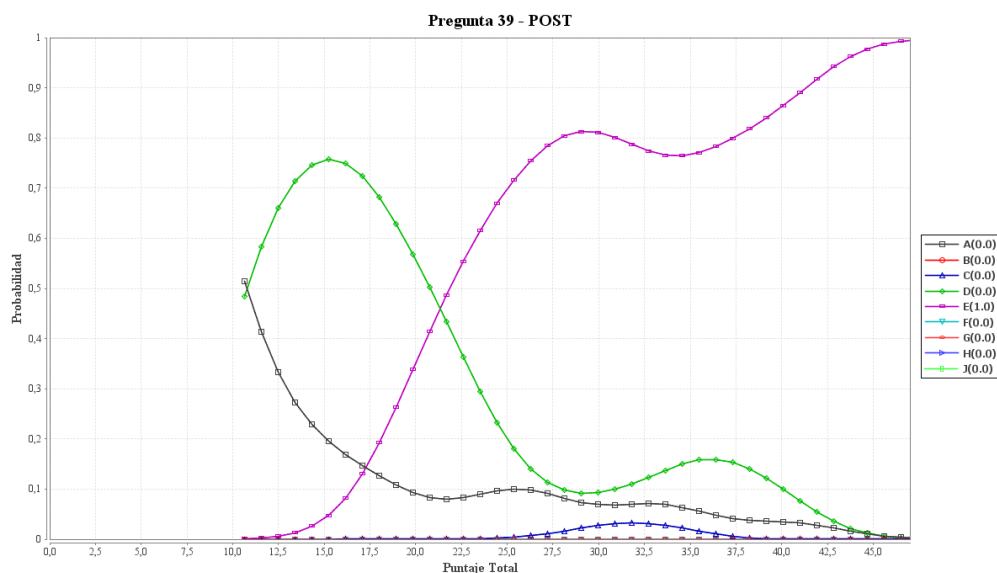


Figura 17: Gráfico Pregunta 39 POST FMCE 2013-2

4.3.- FCI 2012 v/s 2024: A doce años de diferencia

El próximo análisis por realizar, corresponde a la comparación de la misma evaluación con más de una década de diferencia (y una pandemia con educación remota incluida entremedio). Para esto, al igual que en casos previos, se realizó un análisis con *IRT* de un parámetro inicial, que fue utilizado como base para detectar preguntas con patrones de dificultad interesantes; de manera más específica, preguntas con errores conceptuales importantes o cuya evolución entre la evaluación PRE y POST no son lo esperando, mostrando un aumento de dificultad o una disminución de la misma muy baja.

En una primera instancia es interesante el hecho de que los resultados del *FCI* 2012 muestran, en general, mejores resultados tanto en la evaluación PRE como POST en los estudiantes, alcanzando más estudiantes un puntaje máximo y en las evaluaciones POST sobre todo alcanzando un valor medio mayor. Las preguntas puntuales para analizar, además, corresponden con las 1, 17 y 23.

Comenzando con las figuras 18 y 19, es notorio de inmediato que para el grupo del semestre 2012-2 que los valores de puntaje total de los estudiantes cubren valores más extendidos, alcanzando un número de estudiantes suficientes el valor máximo de puntaje como para que la regresión lineal sin parámetros ocupe todo el espacio del eje x de las gráficas, mientras que para los datos de 2024-2 no alcanzan a los 25 puntos totales en la evaluación.

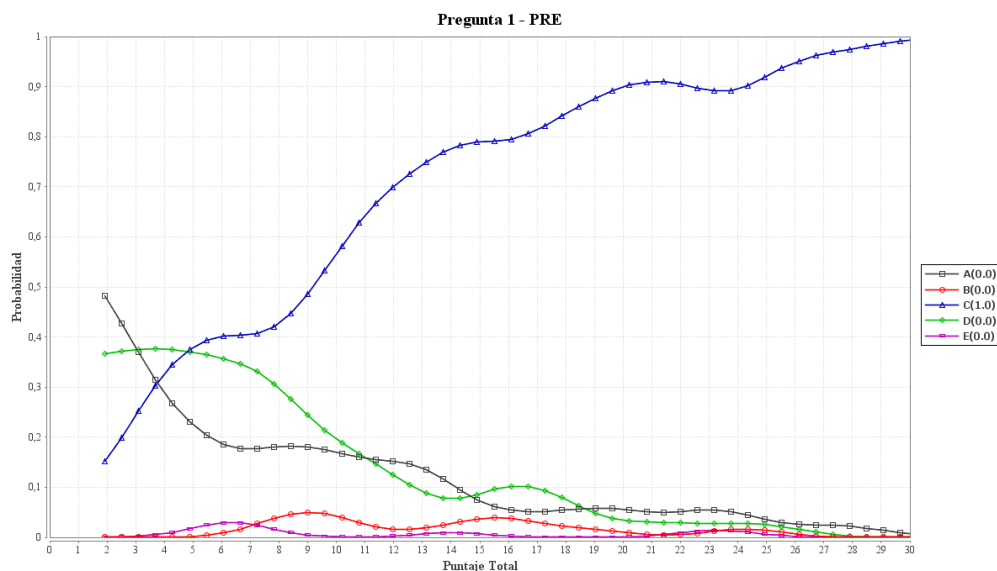


Figura 18: Gráfico Pregunta 1 PRE FCI 2012-2

Más allá de este detalle, las tendencias de los estudiantes siguen los mismos patrones de errores conceptuales principales, siendo las alternativas A y D las predominantes para ambos grupos, aunque es claro que para la generación del año 2024 el error conceptual de la alternativa D resulta más llamativo, y que todas las demás alternativas erróneas tienen una probabilidad ligeramente más alta de ser seleccionadas. También es claro que la respuesta correcta comienza a ganar presencia entre los estudiantes del año 2012 mucho antes en lo que a nivel de habilidad respecta, teniendo una posibilidad de al menos un cincuenta por ciento de ser elegida ya desde los 10 puntos en adelante, en comparación con el umbral cercano a los 16 puntos para la generación del año 2024.

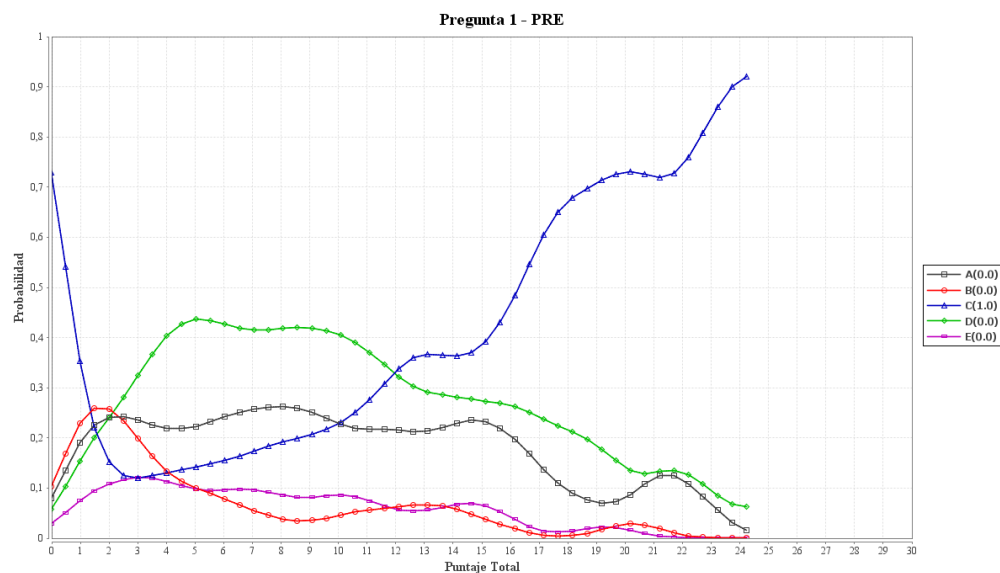


Figura 19: Gráfico Pregunta 1 PRE FCI 2024-2

En cuanto a las respuestas después de terminada asignatura se encuentra un patrón similar. Los resultados en la *Figura 20* presentan esta vez datos que cubren menos del eje x de las variables, pero a favor de un mayor nivel de habilidad promedio, mientras que para la *Figura 21* la cantidad de estudiantes con puntaje máximo en la evaluación permiten que la regresión lineal genere curvas que cubren todo el espectro de valores, pero sigue existiendo un volumen de estudiantes suficiente con puntajes bajos para que las curvas de respuestas se extiendan por toda la gráfica.

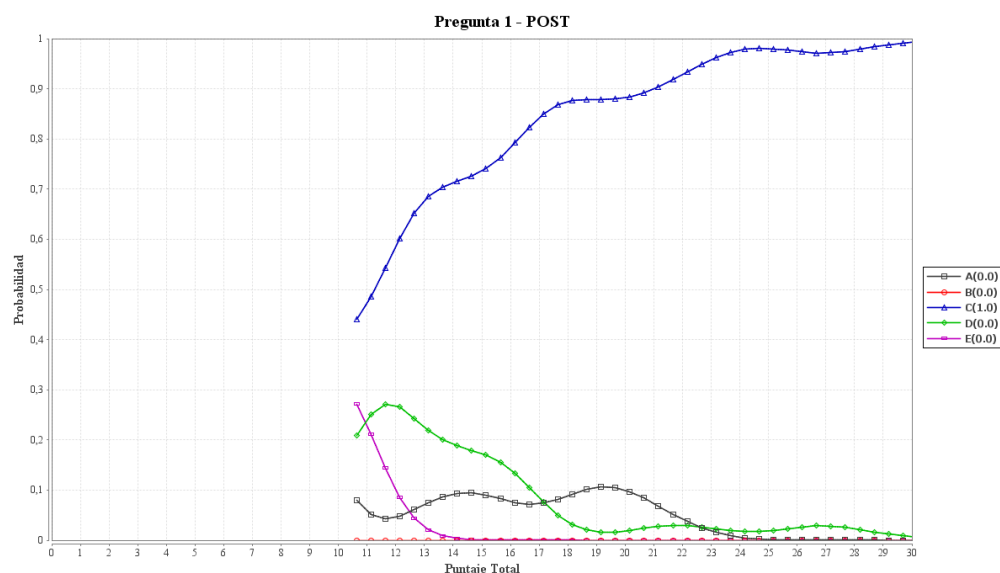


Figura 20: Gráfico Pregunta 1 POST FCI 2012-2

En lo que a conceptos erróneos respecta también se evidencia un comportamiento curioso; en primer lugar, la curva de la respuesta correcta es mucho más gradual en la *Figura 21*, mostrando en lugar de una mejora en comparación con la evaluación PRE, solamente el hecho de que la presencia de más estudiantes con mejores niveles de habilidad causa que la curva siga el comportamiento esperado debido a la existencia de más datos en esta zona de la gráfica, pero no que la alternativa C tenga más presencia en valores de puntaje máximo menores, o que la curva de la misma sea creciente de manera más abrupta. Por otro lado, mientras que en las dos gráficas del año 2012 se notan los mismos errores conceptuales, pero en menor medida en el POST que en el PRE, para el año 2024 la alternativa errónea A gana más fuerza en el grupo de estudiantes que ya terminaron la asignatura.

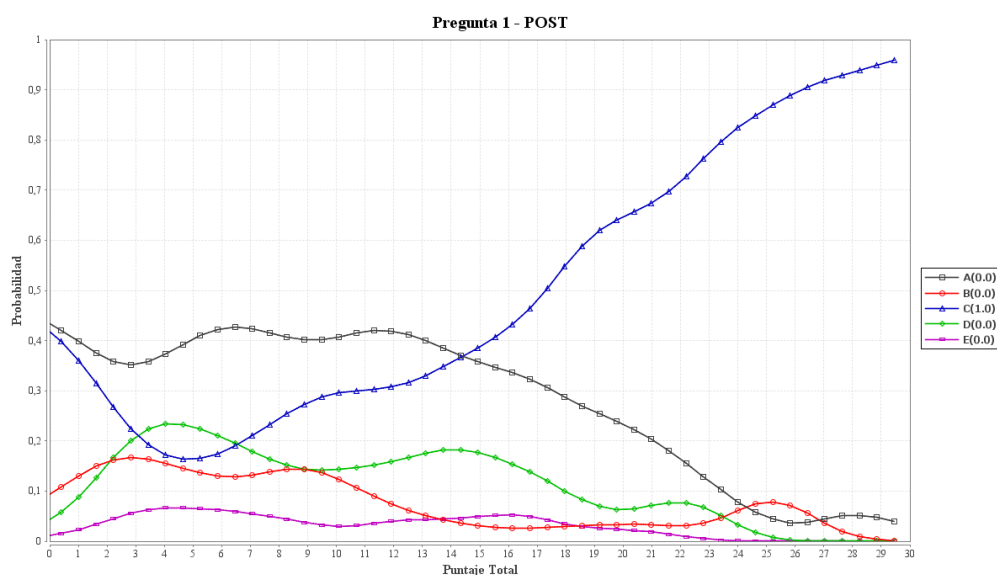


Figura 21: Gráfico Pregunta 1 POST FCI 2024-2

En la pregunta 17 aparecen resultados ligeramente más cercanos a lo esperado, pero no por esto menos llamativos. En la *Figura 22* se observa que la generación del año 2012 tuvo una gran tendencia, en todos los niveles de habilidad menos en los más altos, a caer en la alternativa A, de manera muy similar a la generación 2024. Sin embargo, en la *Figura 23* se evidencia, no solo que la tendencia a otras alternativas incorrectas es menor, sino, además, que la curva de la alternativa correcta es ligeramente mayor en los valores más bajos de puntaje total en la evaluación.

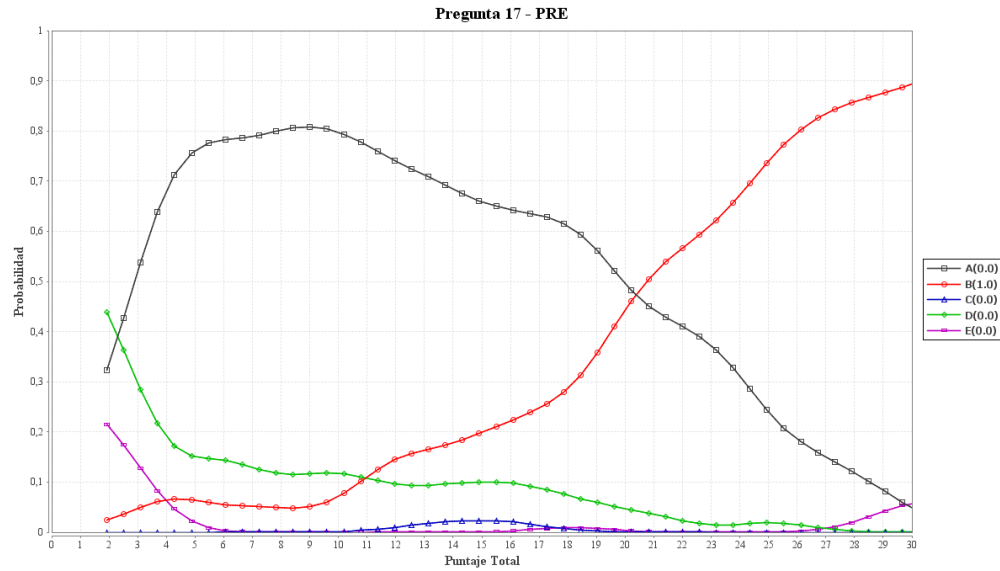


Figura 22: Gráfico Pregunta 17 PRE FCI 2012-2

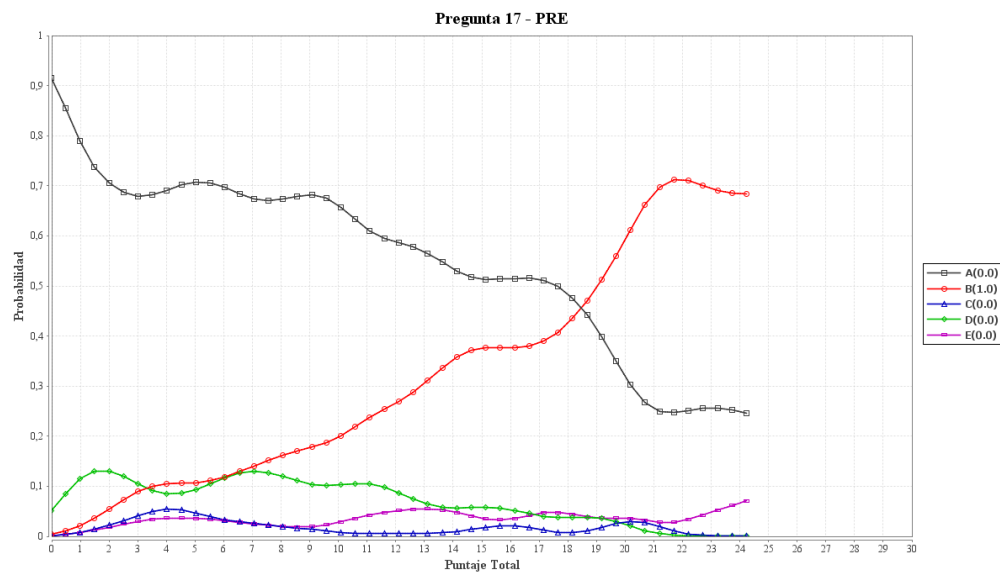


Figura 23: Gráfico Pregunta 17 PRE FCI 2024-2

Esto contrasta de manera interesante con las figuras 24 y 25, donde notamos que, si bien ambas generaciones presentan el mismo punto crítico donde concepto erróneo y alternativa correcta se intercambian, la presencia de una nueva idea alternativa se hace notar en los valores intermedios de nivel de habilidad en la generación 2012, mientras que la 2024 muestra una nueva tendencia mucho menor para niveles de habilidad relativamente altos.

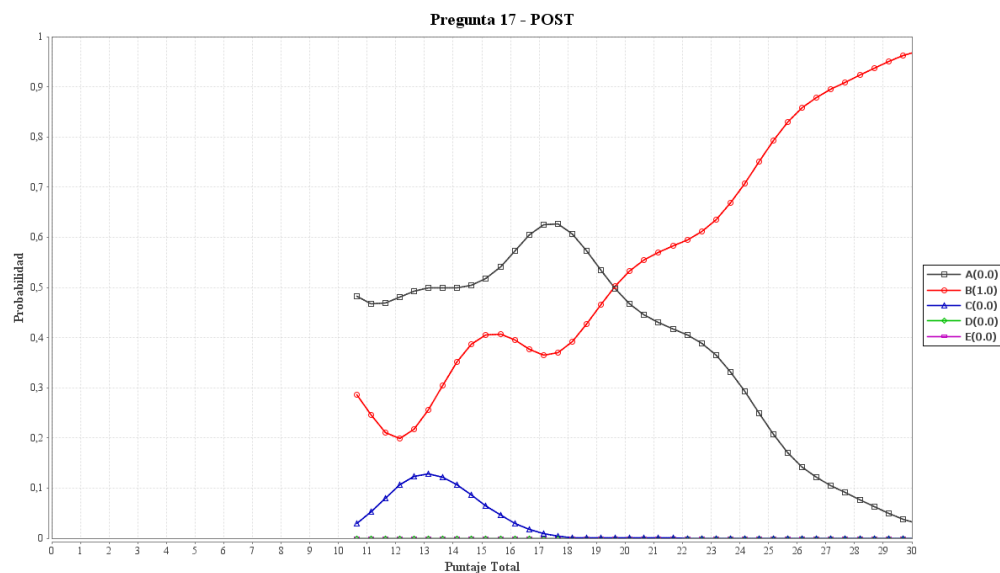


Figura 24: Gráfico Preguntar 17 POST FCI 2012-2

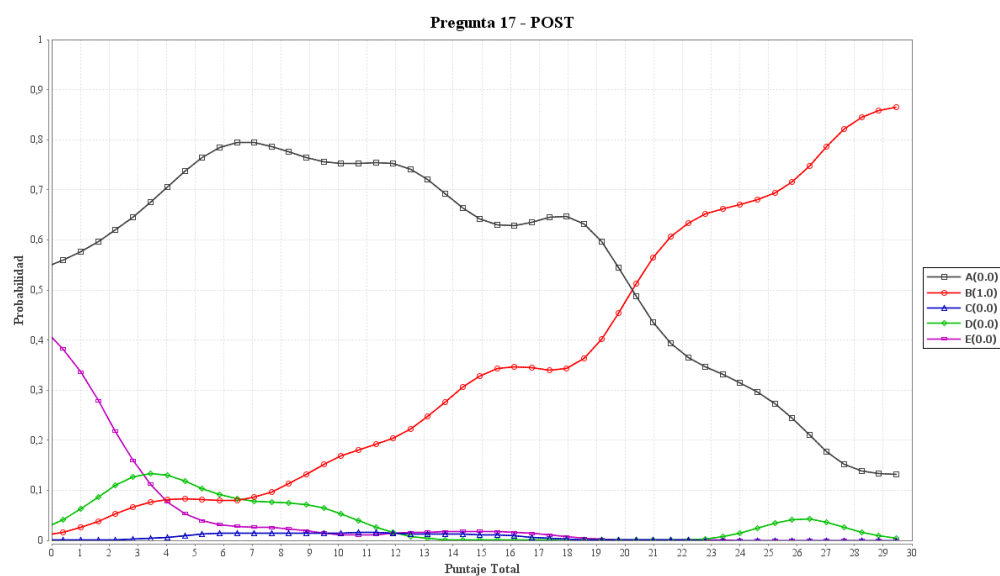


Figura 25: Gráfico Preguntar 17 POST FCI 2024-2

Tras haber visto ya un caso donde la generación 2012 parece afrontar mejor los problemas del FCI que la 2024 y uno donde el comportamiento de ambos grupos de estudiantes se mantiene dentro de lo esperado, solo falta analizar el último caso. En la Figura 26 se nota de inmediato la fuerte presencia de una idea alternativa en la forma de la respuesta C, que se muestra de manera prominente a lo largo de todos los niveles de habilidad de los estudiantes en la generación del año 2012.

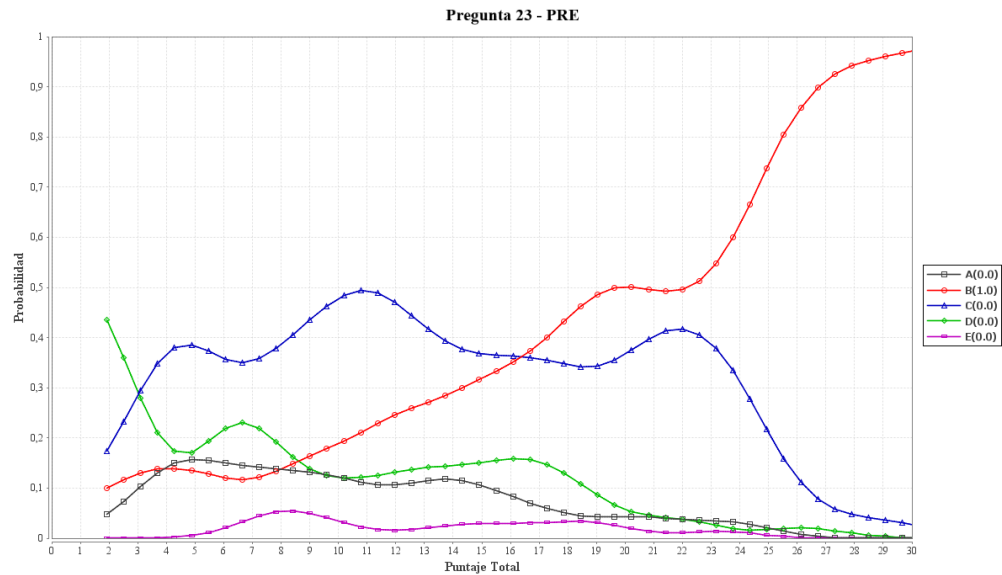


Figura 26: Gráfico Pregunta 23 PRE FCI 2012-2

Por otro lado, en la *Figura 27* se ve que la generación 2024, que si bien presenta un similar comportamiento en lo que a la respuesta correcta (B) respecta, tienden a responder con una tendencia ligeramente menor la alternativa C, aunque también en sus valores de habilidad menores presentan con algo más de fuerza una tendencia a la alternativa errónea D.

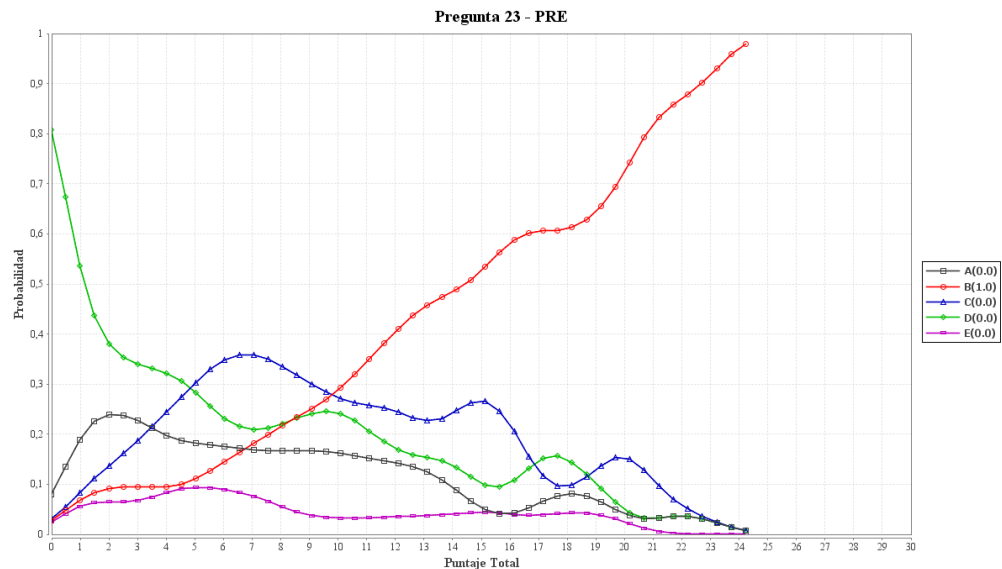


Figura 27: Gráfico Pregunta 23 PRE FCI 2024-2

Aquí es donde se percibe un cambio respecto al comportamiento en las preguntas anteriores, pues en la *Figura 28* sigue siendo evidente la tendencia del grupo del año 2012 a

caer en la misma idea alternativa de la alternativa C, a pesar de que los valores de puntaje total de los estudiantes haya causado un desplazamiento general de las curvas hacia los valores mayores de habilidad. Por otro lado, en la *Figura 29* se nota de inmediato que la tendencia por la alternativa C es mucho menor, no alcanzando esta una probabilidad de ser elegida mayor a un treinta por ciento. Esto, por supuesto, no viene sin matices; la alternativa D para este año de estudiantes no se desvanece con la rapidez que en el año 2012, y si bien ninguna de las alternativas incorrectas alcanza valores tan altos como en la generación anterior, sí ocurre que todas mantienen cierta presencia no nula a lo largo de todos los valores de nivel de habilidad de los estudiantes.

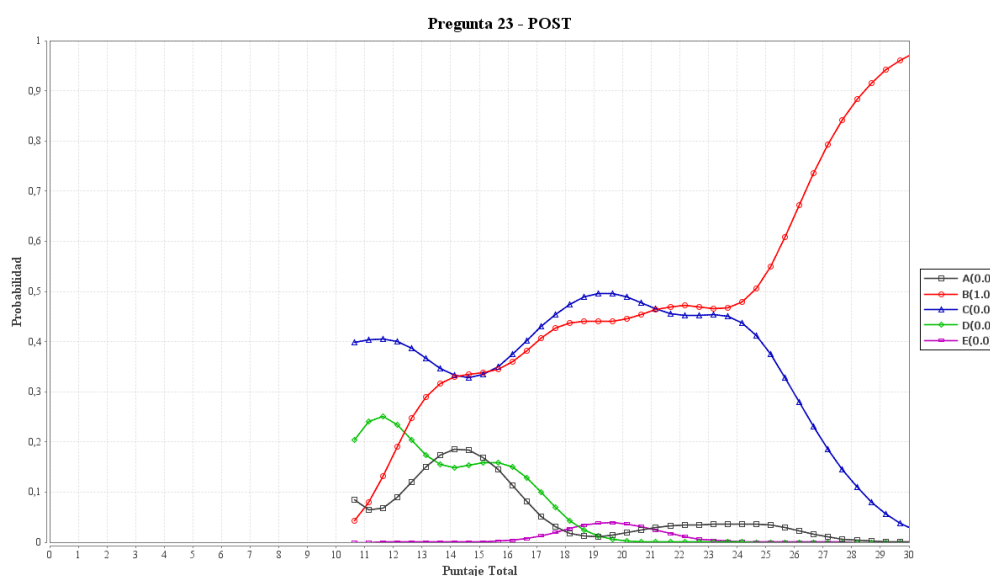


Figura 28: Gráfico Pregunta 23 POST FCI 2012-2

Además, comparando ambas curvas para la alternativa correcta, es bastante evidente que el crecimiento de la misma para el año 2024 es más marcado, superando esta el 50% de probabilidades de ser elegida a los 15 puntos totales en la evaluación, en comparación con el punto crítico de 24 puntos para la generación 2012.

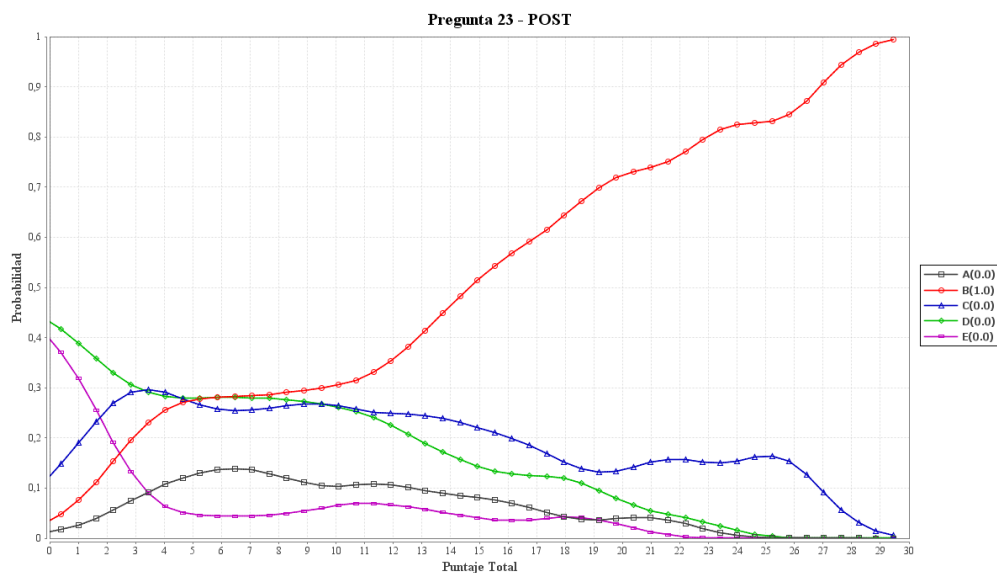


Figura 29: Gráfico Pregunta 23 POST FCI 2024-2

4.4.- FMCE 2014: Una Retroalimentación Más Precisa

Fue mencionada con antelación la aplicación de jMetrik para realizar un estilo de análisis más allá de lo hecho hasta este punto: la generación de una tabla de datos fuera del software que comparara las respuestas marcadas por los estudiantes en la evaluación una vez terminada la asignatura, pero graficadas en función de sus puntajes al inicio de la misma.

Este análisis producirá tres gráficas por pregunta, y por lo mismo, se reducirá el número de preguntas estudiadas a una, con el fin de permitir profundizar más en cada una. Iniciando con la elección de la pregunta tras aplicar a la evaluación un análisis con modelo de Rasch, destacan, de entre otros comportamientos curiosos en las tablas de datos, la preguntas 6 de una rendición del FMCE el año 2014.

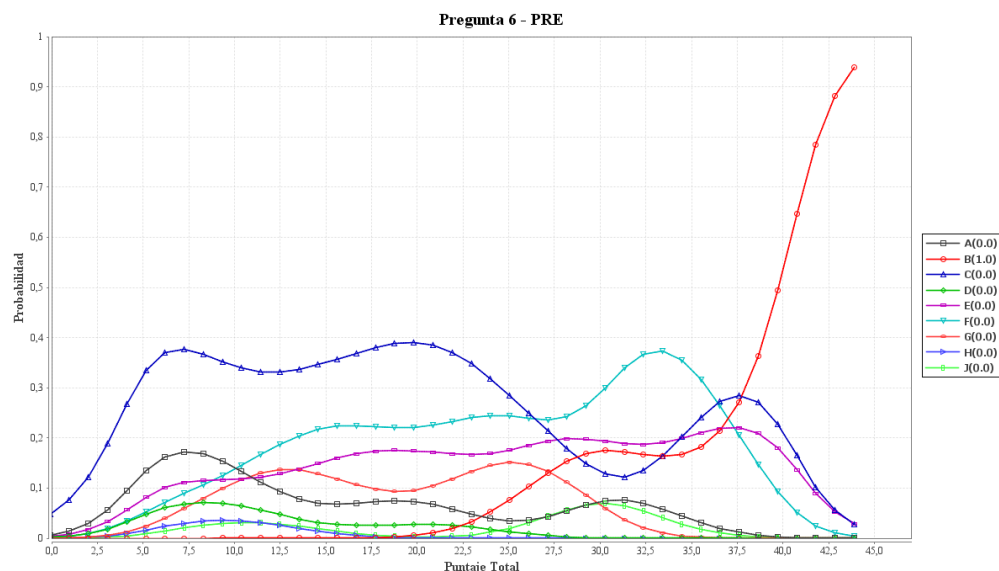


Figura 30: Gráfico Pregunta 6 PRE FMCE 2014-1

Empezando con la *Figura 30*, tenemos una situación donde la dificultad de la pregunta es bastante alta, y por un motivo más bien obvio: además de la presencia de una alternativa errónea que tiene mucha presencia a lo largo de todos los niveles de habilidad (C), y de una segunda respuesta errónea que cerca de los puntajes totales más altos tiene una aparición prominente (F), casi todas las posibles respuestas de la pregunta tienen cierta posibilidad de ser elegidas por los estudiantes de todos los niveles de habilidad de la muestra estudiada, y además, el punto crítico para el cual la posibilidad de que los estudiantes respondan la alternativa correcta más de la mitad de las veces se encuentra por sobre los 40 puntos.

Al estudiar entonces los resultados POST en la *Figura 31*, los avances no parecen muy prometedores. La respuesta correcta, incluso en su punto más alto, no alcanza el 70% de ser elegida incluso por los estudiantes con el mayor número de respuestas correctas en la evaluación, y las curvas para las respuestas alternativas, si bien más reguladas y con una tendencia menos clara a una o dos respuestas puntuales, siguen estando presentes de manera regular para todos los calores de habilidad con un rango de porcentajes de entre diez y treinta por ciento.

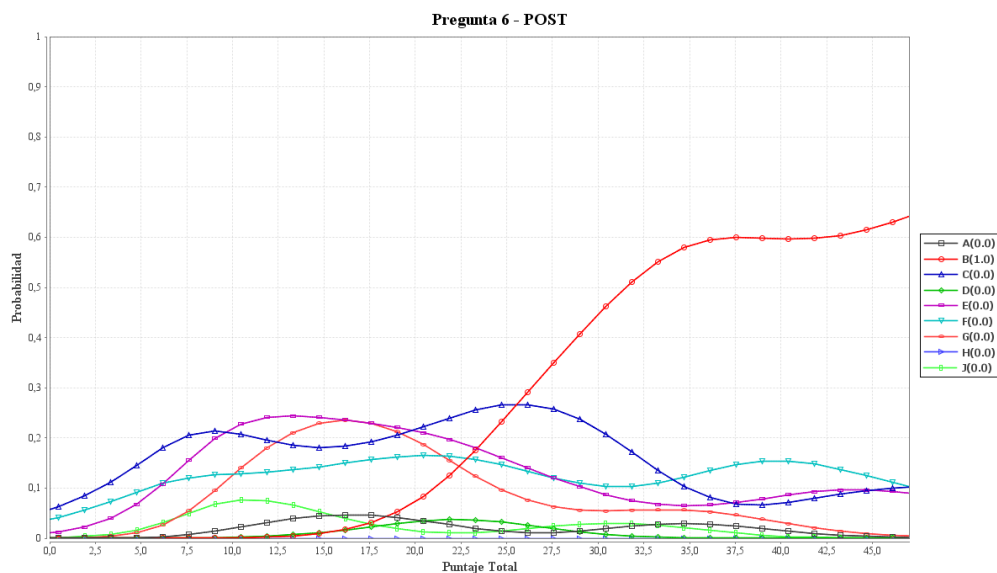


Figura 31: Gráfico Pregunta 6 POST FMCE 2014-1

Este análisis, que indicaría en primera instancia un bajo índice de corrección de ideas alternativas y conceptos erróneos en los estudiantes a lo largo de la asignatura, puede verse sesgado al tomar los datos en bruto de antes y después del semestre, y aquí es donde se deben aplicar los resultados estudiados en la *Figura 32*. Al considerar los puntajes de los estudiantes *antes* de cursar la asignatura, pero considerando sus respuestas después, se tendrá un seguimiento más cercano de los mismos grupos de estudiantes y el cambio en sus respuestas en la evaluación POST, y de inmediato es notorio que, aquellos que en un principio respondieron un total de entre 27 y 40 respuestas correctas, que antes seguían encontrándose muy por debajo de la probabilidad del 50% de responder de manera correcta, ahora estarán muy por encima de la media, al momento de responder correctamente la pregunta. Sin ser esto toda la información a obtener, también se nota que, a diferencia de la gráfica anterior, los estudiantes que en un principio caían bajo todos los errores de las demás alternativas en casi todos los niveles de habilidad, ahora caen en estos mismos principalmente para valores de habilidad por debajo de los 30 puntos totales en la evaluación.

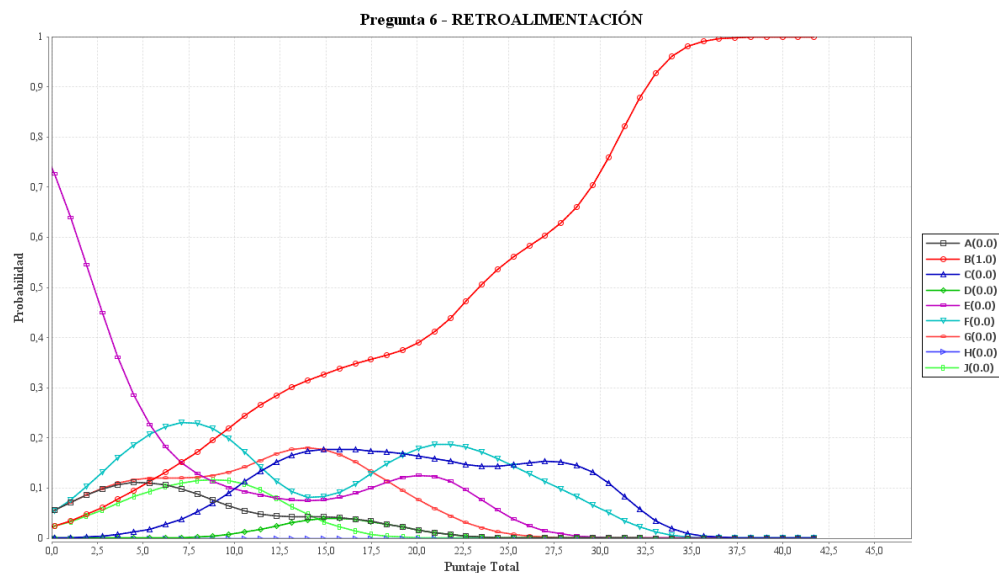


Figura 32: Gráfico Pregunta 6 Retroalimentación FMCE 2014-1

Esta diferencia de análisis tiene muchos matices y consideraciones, pero el principal de todos es que, a la hora de construir manualmente la tabla de datos que entrega como resultado la *Figura 32*, se deben tener en cuenta única y exclusivamente los estudiantes que rindieron tanto la evaluación PRE como la evaluación POST, ignorando y eliminando de los datos obtenidos a quienes hayan cursado la asignatura sin haber dado la evaluación inicial, los estudiantes que se hayan retirado de la asignatura a medio camino, o cualquier otra excepción que cause datos en la tabla de respuestas POST que no puedan ser relacionados con los puntajes PRE de cada estudiante; para esto también es importante registrar, enumerar o etiquetar a los estudiantes de manera correcta, para facilitar el seguimiento del grupo completo de estudio en ambas etapas y así poder crear la tabla de datos constituida de ambos grupos.

5.- Conclusiones

La manera en que un tutor enfrenta los vacíos y contradicciones en el conocimiento de sus estudiantes sigue siendo un área de trabajo arduo y cuya solución no fue el objetivo de este estudio. La enseñanza y el aprendizaje son un área muy amplia cuyas posibles soluciones no pueden ser consideradas sencillas, y áreas como la epistemología llevan siglos buscando perfeccionar el cómo se consigue y transmite el conocimiento.

El análisis de datos de evaluaciones del estilo del *FCI* o el *FMCE* ha demostrado ser una herramienta muy útil para muchas áreas de la educación. El estudio de las evaluaciones y su efectividad para sacar a la luz preconceptos erróneos clave en el entendimiento de los estudiantes es una herramienta útil para la mejora en el proceso de enseñanza y aprendizaje. La comparación de los conocimientos de los estudiantes al iniciar y al finalizar un proceso educativo apoya la evaluación de los planes de enseñanza y la manera en la que los profesores comparten sus conocimientos, buscando áreas donde hacer mejoras y reformas cuando sea adecuado. La consideración del paso de los años y el efecto de hitos históricos importantes en cómo los estudiantes conciben la educación misma.

Todas estas aristas se hicieron notar a lo largo del presente estudio, y son un importante primer paso que podría llevar a hacer cambios en evaluaciones, planes de enseñanza y paradigmas en cómo se imparten ciertas asignaturas. Sobre todo con las comparaciones entre los años 2012 y 2024 de la misma evaluación surgen más preguntas que respuestas; la educación jamás volverá a ser la misma tras los fatídicos años de pandemia, y eso se nota en los resultados.

De aquí en adelante, se debe buscar utilizar estas potentes herramientas analíticas con sabiduría, pues las evaluaciones de este estilo existen, y fueron diseñadas, para mucho más que para obtener datos de notas promedio y seguir adelante. Los datos están para utilizarse, y como se pudo evidenciar a lo largo de estas páginas, tienen mucho potencial. Lo único que se necesita es saber sacarle provecho.

6.- Bibliografía

- Adams, W., & Wieman, C. (2010). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science*, 33(9), 1289-1312.
- Ausubel, D. P. (1961). In Defense of Verbal Learning. *Educational Theory*, 11(1), 15–25.
- Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics - Physics Education Research*, 5(2).
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141–158.
- Livingston, S. A. (June de 2020). *Basic Concepts of Item Response Theory: A Nonmathematical Introduction*. Obtenido de ETS: https://www.ets.org/research/policy_research_reports/publications/report/2020/kbxx.html
- Mestre, J. P., & Docktor, J. L. (2021). The Formation of Concepts and How to Fix Broken Ones. En *The Science of Learning Physics: Cognitive Strategies for improving instruction*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Meyer, J. P. (2014). *Applied measurement with jMetrik*. New York: Routledge/Taylor and Francis Group.
- Morris, G. A., Branum-Martin, L., Harshman, N., Baker, S. D., Mazur, E., Dutta, S., . . . McCauley, V. (2006). Testing the test: Item response curves and test quality. *American Journal of Physics*, 74(5), 449–453.
- Shuell, T. J. (1992). Designing Instructional Computing Systems for Meaningful Learning. En M. Jones, P. H. Winne, & C. Tubman, *Adaptive Learning Environments: Foundations and frontiers*. (págs. 19-54). Berlin.: Springer-Verlag.
- Smith, T. I., Gray, K. A., Louis, K. J., Ricci, B. J., & Wright, N. J. (2017). Showing the dynamics of student thinking as measured by the FMCE. *Physics Education Research Conference Proceedings*, 380–383.
- Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of Active Learning Laboratory and lecture curricula. *American Journal of Physics*, 66(4), 338–352.

- Thornton, R. K., Kuhl, D., Cummings, K., & Marx, J. (2009). Comparing the force and motion conceptual evaluation and the Force Concept Inventory. . *Physical Review Special Topics - Physics Education Research*, 5(1).
- Zavala, G., Alarcon, H., Henderson, C., Sabella, M., & Hsu, L. (2008). Evaluation of instruction using the conceptual survey of electricity and magnetism in Mexico. *AIP Conference Proceedings*, 231–234.