



Benemérita Universidad Autónoma de Puebla  
Facultad de Ciencias de la Electrónica

**Estudio del uso y aceptación de una herramienta en línea para la  
evaluación automática de conceptos básicos de física basados en  
preguntas de respuesta corta en una universidad privada de Puebla**

Tesis

Presentada para obtener el grado de  
Doctor en Sistemas y Ambientes Educativos

Presenta

Julián Gerardo Torres Kauffman

Director

Dr. Iván Olmos Pineda

Codirector

Dr. Juan Manuel González Calleros

Evaluadores externos

Dra. Rosario Rogel-Salazar

Universidad Autónoma del Estado de México

Dr. Alfonso Infante Moro

Universidad de Huelva

Junio 2024

## **Dedicatoria**

A Flor Estela

A Don Roger

A Caro

## **Agradecimientos**

Quiero expresar mi agradecimiento al comité tutorial, compuesto por el Dr. Iván Olmos Pineda y el Dr. Juan Manuel González Calleros, por su dedicación y compromiso en guiar el desarrollo de este trabajo.

Extiendo mi gratitud a todas las doctoras y doctores miembros del núcleo académico del DSAE por su constante apoyo y orientación.

Agradezco de manera especial a la Dra. Josefina Guerrero García, la Dra. Eugenia Erica Vera Cervantes y el Dr. Alfredo García Suárez por su tiempo y aportaciones fundamentales que han brindado a este trabajo.

Asimismo, quiero agradecer a los evaluadores externos, la Dra. Rosario Rogel-Salazar y el Dr. Alfonso Infante, por la revisión y valiosa retroalimentación.

Finalmente, agradezco al Consejo Nacional de Humanidades, Ciencias y Tecnología (CONAHCYT) por la beca número 769631, que me fue otorgada para la realización de este trabajo.

## Resumen

El rápido desarrollo de las Tecnologías de la Información y la Comunicación (TIC) abrió nuevas posibilidades para las prácticas de enseñanza y evaluación en la educación superior, entre estas se encuentra la evaluación basada en computadora (*Computer Based Assessment*, CBA). Ya que el desarrollo efectivo de esta depende de la aceptación de los estudiantes, en el presente estudio se analizaron los constructos que afectan la intención de utilizar la CBA mediante el Modelo de Aceptación de Evaluación Basada en Computadora (*Computer Based Assessment Acceptance Model*, CBAAM) en una universidad privada de Puebla, México. La metodología fue cuantitativa y se dividió en dos fases: 1) desarrollo e implementación de un sistema de evaluación automática utilizando dos formatos de preguntas, uno de respuesta corta y otro de opción múltiple, 2) prueba de la medición y el modelo estructural del instrumento CBAAM. La muestra fue de 84 estudiantes del primer semestre de ingeniería la cual se dividió aleatoriamente en grupo de control y grupo experimental.

Para analizar el modelo de adopción de la evaluación basada en computadora, en este trabajo, se utilizó el enfoque de modelado de ecuaciones estructurales (*Structural Equation Modeling*, SEM) basado en mínimos cuadrados parciales (*Partial Least Squares*, PLS), donde se probó el modelo de medición y estructural. Los resultados indican una diferenciación en la aceptación de estos dos formatos de preguntas, respaldada por el Modelo de Aceptación de Evaluación Basado en Computadora (CBAAM), que revela una preferencia notable hacia el formato de respuesta corta, sugiriendo que este formato no solo facilita una evaluación más profunda de la comprensión conceptual, sino que también mejora la percepción lúdica de la evaluación por parte de los y las estudiantes.

Adicionalmente, los hallazgos indican que la percepción lúdica tiene un efecto directo sobre el uso de la CBA, mientras que las condiciones facilitadoras, la autoeficacia informática, la facilidad de uso percibida, la expectativa de meta, la influencia social y el contenido solo tienen efectos indirectos. El modelo de aceptación estudiado para el formato de respuesta corta explica aproximadamente 77% de la variación de la intención

de uso, y para el formato de opción múltiple se obtuvo un valor de 47%. Se recomienda investigar otras variables que afectan el propósito de usar y aplicar el modelo en otros contextos para mayor confirmación.

**Palabras clave:** aceptación de la tecnología; educación superior; evaluación basada en computadora; evaluación electrónica.

## **Abstract**

The rapid development of information and communication technologies (ICT) has opened new possibilities for teaching and assessment practices in higher education, among which is computer-based assessment (CBA). Since the effective development of this depends on student acceptance, this study analyzed the constructs affecting the intention to use CBA through the computer-based assessment acceptance model (CBAAM) at a private university in Mexico. The methodology was quantitative and divided into two phases: 1) development and implementation of an automatic assessment system using two question formats, one short-answer and one multiple-choice, and 2) testing the measurement and structural model of the CBAAM instrument. The sample consisted of 84 first-semester engineering students which were randomly divided into a control group and an experimental group.

To analyze the adoption model of computer-based assessment, this work used the structural equation modeling (SEM) approach based on partial least squares (PLS), where the measurement and structural model were tested. The results indicate a differentiation in the acceptance of these two question formats, supported by the Computer-Based Assessment Acceptance Model (CBAAM), revealing a notable preference for the short-answer format. This suggests that this format not only facilitates a deeper evaluation of conceptual understanding but also improves the playful perception of the assessment by students.

Furthermore, the findings indicate that playful perception has a direct effect on the use of CBA, while facilitating conditions, computer self-efficacy, perceived ease of use, goal expectancy, social influence, and content only have indirect effects. The acceptance model studied for the short-answer format explains approximately 77% of the variation in the intention to use, and for the multiple-choice format, a value of 47% was obtained. It is recommended to investigate other variables affecting the intention to use and apply the model in other contexts for further confirmation.

**Keywords:** technology acceptance; higher education; computer-based assessment; electronic assessment.

## CONTENIDO

ÍNDICE DE TABLAS .....	11
ÍNDICE DE FIGURAS .....	13
CAPÍTULO 1. INTRODUCCIÓN.....	14
1.1 Problema de investigación.....	14
1.2 Objetivos.....	18
1.2.1 Objetivo general .....	18
1.2.2 Objetivos específicos .....	18
1.3 Pregunta de investigación .....	18
1.4 Hipótesis .....	19
1.5 Justificación .....	20
1.5.1 Contexto y Situación Actual del Problema .....	20
1.5.2 Evaluación actual de conceptos de la física.....	23
1.6 Delimitación del objeto de estudio .....	25
1.7 Alcance del estudio.....	28
1.7.1 Delimitaciones.....	28
1.7.2 Limitaciones .....	29
1.8 Paradigma de la investigación .....	29
1.9 Línea de investigación .....	30
1.10 Conclusiones del capítulo.....	31
1.11 Organización de la tesis .....	31
CAPÍTULO 2. ESTADO DEL ARTE .....	33
2.1 Una descripción general de los inventarios de conceptos en física.....	35
2.2 Preguntas de opción múltiple .....	37
2.3 Evaluación automática de preguntas de respuesta libre .....	43
2.4 Modelos de aceptación de tecnología .....	48
2.5 Modelos de aceptación de evaluaciones electrónicas .....	51
2.6 Conclusiones sobre el estado del arte.....	54
CAPÍTULO 3. MARCO TEÓRICO.....	60
3.1 Escribir para aprender .....	62
3.2 Evaluación formativa .....	63
3.3 Respuesta construida .....	64
3.4 Evaluación escrita.....	66
3.5 Comprensión conceptual de la física .....	67



3.6 El inventario de conceptos sobre fuerza.....	68
3.7 Evaluación asistida por computadora o e-assessment.....	71
3.8 Sistemas de evaluación electrónica.....	71
3.9 Procesamiento del Lenguaje Natural.....	73
3.10 Expresiones regulares.....	74
3.11 La aceptación de tecnología.....	74
3.12 Contexto de la investigación.....	79
3.13 Aceptación de la evaluación basada en computadora.....	80
3.14 Modelo de investigación CBAAM e hipótesis.....	82
3.15 Conclusiones.....	90
CAPÍTULO 4. DISEÑO METODOLÓGICO.....	92
4.1 Introducción.....	92
4.2 Filosofía de investigación.....	92
4.3 Perspectiva de Investigación del Presente Estudio.....	93
4.4 Enfoque y alcance de la investigación.....	94
4.5 Diseño de investigación.....	95
4.6 Población, muestra y muestreo.....	96
4.7 Recolección de los datos cuantitativos.....	100
4.8 Instrumentos aplicados en la investigación.....	100
4.9 Confiabilidad de los instrumentos.....	101
4.9.1 Instrumento de evaluación automática de conceptos básicos de física.....	101
4.9.2 Modelo de Aceptación de Evaluación Basado en Computadora.....	105
4.10 Procedimientos para la recolección de datos.....	107
4.10.1 Estudio piloto.....	107
4.10.2 Resultados preliminares de la prueba piloto.....	108
4.10.3 Estudio principal.....	109
4.11 Procedimiento Estudio Principal.....	110
4.12 Análisis de los datos.....	111
4.13 Técnicas estadísticas.....	111
4.14 Modelado de ecuaciones estructurales.....	113
4.15 Elección de PLS-SEM para la estimación del modelo de este estudio.....	113
4.16 Procedimiento para analizar los datos.....	114
4.17 Modelo estructural.....	116
4.17.1 Las variables exógenas.....	116

4.17.2 Las variables endógenas .....	117
4.18 El modelo de medición .....	118
4.19 Escala de medición reflectiva .....	118
4.20 Escala de medición formativa .....	119
4.21 Evaluación del resultado PLS-SEM del modelo.....	119
4.22 Consideraciones éticas.....	120
4.23 Conclusiones .....	124
CAPÍTULO 5. RESULTADOS .....	125
5.1 Modelo estructural .....	126
5.2 Etapa A: Evaluación del modelo reflexivo.....	128
5.2.1 Primero. Cargas del modelo exterior y fiabilidad del indicador.....	129
5.2.2 Segundo. Fiabilidad de consistencia interna .....	133
5.2.3 Tercero. Validez convergente .....	134
5.2.4 Cuatro. Validez discriminante.....	135
5.3 Etapa B: Evaluación del modelo estructural .....	139
5.3.1 Primero. Prueba del modelo estructural para detectar posibles problemas de colinealidad .....	139
5.3.2 Segundo. Evaluación de la capacidad del modelo para predecir las variables latentes endógenas.....	141
5.3.3 Tercero. Comprobar la significancia y relevancia de las rutas estructurales .....	144
5.4 Discusión .....	162
5.4.1 Contribuciones significativas.....	166
5.4.2 Similitudes y diferencias con estudios previos .....	167
5.4.3 Limitaciones, Impacto y Recomendaciones .....	169
5.4.4 Aprendizajes y Reflexiones finales.....	171
CAPÍTULO 6. CONCLUSIONES.....	174
Referencias .....	183
Apéndice A: Traducción del Modelo de Aceptación de Evaluación Basada en.....	204
Apéndice B. Instrumento de evaluación en formato de respuesta corta .....	206
Apéndice C: Instrumento de evaluación en formato de opción múltiple .....	233
Apéndice D: Resultados de la prueba piloto sobre la confiabilidad del CBAAM .....	254
Apéndice E: Variables del modelo CBAAM.....	255

## ÍNDICE DE TABLAS

Tabla 1. Estudios relacionados con la aceptación de las evaluaciones electrónicas ....	52
Tabla 2. Línea del tiempo de los modelos y teorías incluidos en la UTAUT.....	76
Tabla 3. Constructos y vínculos causales desarrollados en estudios previos.....	81
Tabla 4. Características del positivismo.....	94
Tabla 5. Descripción de las variables de investigación.....	96
Tabla 6. Cargas de modelo externo (grupo experimental).....	129
Tabla 7. Cargas de modelo externo (grupo control).....	130
Tabla 8. Confiabilidad de constructo y validez del modelo (grupo experimental).....	131
Tabla 9. Confiabilidad de constructo y validez del modelo (grupo control).....	132
Tabla 10. Validez discriminante utilizando el criterio de Fornell-Lacker (grupo experimental).....	136
Tabla 11. Validez discriminante utilizando el criterio de Fornell-Lacker (grupo control).....	136
Tabla 12. Validez discriminante usando cargas cruzadas (grupo experimental).....	137
Tabla 13. Validez discriminante usando cargas cruzadas (grupo control).....	138
Tabla 14. Valores VIF del modelo interno (grupo experimental).....	140
Tabla 15. Valores VIF del modelo interno (grupo control).....	140
Tabla 16. Valores R <sup>2</sup> de las variables latentes endógenas (grupo experimental).....	142
Tabla 17. Valores R <sup>2</sup> de las variables latentes endógenas (grupo control).....	142
Tabla 18. Redundancia con validación cruzada de variables latentes (Q <sup>2</sup> ) (grupo experimental).....	144
Tabla 19. Redundancia con validación cruzada de variables latentes (Q <sup>2</sup> ) (grupo control).....	144
Tabla 20. Estadísticos T de coeficientes de trayectoria [modelo estructural (interno)] (grupo experimental).....	146
Tabla 21. Estadísticos T de coeficientes de trayectoria [modelo estructural (interno)] (grupo control).....	148
Tabla 22. Estadísticos T y valores P de las cargas del modelo externo (grupo experimental).....	149

Tabla 23. Estadísticos T y valores P de las cargas del modelo externo (grupo control). .....	150
Tabla 24. Efectos directos, indirectos y totales (grupo experimental). .....	151
Tabla 25. Efectos directos, indirectos y totales (grupo control). .....	152
Tabla 26. Prueba de significancia de los coeficientes de trayectoria del modelo estructural (grupo experimental). .....	155
Tabla 27. Prueba de significancia de los coeficientes de trayectoria del modelo estructural (grupo control). .....	157
Tabla 28. Resultados de las pruebas de hipótesis .....	165

## ÍNDICE DE FIGURAS

Figura 1. Esquema del marco teórico. Elaboración propia.....	61
Figura 2. Concepto básico subyacente a los modelos de aceptación y uso de la tecnología. Fuente: tomada y traducida de Venkatesh, Morris y Davis et al., 2003. ....	76
Figura 3. UTAUT. Fuente: Venkatesh, Morris y Davis et al., 2003.....	79
Figura 4. Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM) de Terzis y Economides (2011).....	89
Figura 5. El ‘diagrama cebolla’. Fuente: Saunders, Lewis y Thornhill, (2018).....	95
Figura 6. Resumen metodología. ....	97
Figura 7. Esquema del experimento.....	98
Figura 8. Ejemplo de una pregunta de respuesta corta sobre. Fuente: Elaboración propia. ....	103
Figura 9. Modelo de investigación (CBAAM) de Terzis y Economides (2011). ....	117
Figura 10. Escala de Medición Reflectiva: La variable latente (X) se mide mediante cuatro variables de medición VM-1, VM-2, VM-3 y VM-4 de forma reflexiva. ....	118
Figura 11. Escala de Medición Formativa: La variable latente (Y) se mide mediante cuatro variables de medición VM-1, VM-2, VM-3 y VM-4 de manera formativa. ....	119
Figura 12. Etapas de evaluación de PLS-SEM. Fuente: Sarstedt et al. (2014).....	121
Figura 13. Resultados PLS-SEM del modelo analizado para el grupo experimental. Las variables latentes se muestran en círculos y los moderadores (variables de medición) se muestran en cuadros amarillos. ....	127
Figura 14. Resultados PLS-SEM del modelo analizado para el grupo control. Las variables latentes se muestran en círculos y los moderadores (variables de medición) se muestran en cuadros amarillos. ....	128
Figura 15. Coeficientes de ruta del modelo CBAAM aplicado a este estudio (grupo experimental). ....	154
Figura 16. Coeficientes de ruta del modelo CBAAM aplicado a este estudio (grupo control). ....	155
Figura 17. Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM) de Terzis y Economides (2011).....	162

## **CAPÍTULO 1. INTRODUCCIÓN**

### **1.1 Problema de investigación**

En la era moderna, todos los aspectos de nuestra vida se ven fuertemente afectados por la tecnología, particularmente su integración en la vida privada y profesional de las personas. La aceptación o rechazo por parte de los usuarios de las aplicaciones tecnológicas en diferentes campos, incluida la educación, ha llamado mucho la atención de los investigadores (Granic & Marangunic, 2019; Marangunic & Granic, 2015). Aunque han existido muchos modelos de investigación en este campo, el Modelo de Aceptación de Tecnología (*Technology Acceptance Model*, TAM), introducido por Davis (1989), se convirtió en uno de los modelos más críticos relacionados con los factores que inciden en la adopción por parte de los usuarios en el uso de la tecnología. Además, se ha enfatizado la importancia que tiene la aplicación de la tecnología en las actividades de enseñanza y aprendizaje (Davis, 2011; Davis et al., 2011; Scherer et al., 2019). Entre estas actividades, la evaluación es uno de los factores clave en la práctica educativa.

La evaluación electrónica (*e-assessment*), con nombres alternativos tales como evaluación basada en computadora (EBC), evaluación digital, o evaluación en línea (Kundu & Bej, 2020), consiste en todo el proceso desde el diseño de las tareas hasta archivar los resultados. Se puede utilizar con fines formativos o sumativos (Appiah & Van Toner, 2018). Las instituciones educativas pueden decidir usar herramientas de evaluación ya integradas en un sistema de gestión de aprendizaje (*Learning Management System*, LMS) o usar herramientas de evaluación separadas (Conole & Warburton, 2005). Si bien plataformas como Moodle, Blackboard o Canvas se pueden utilizar tanto en evaluación sumativa como evaluación formativa, otras aplicaciones como Google Forms, Quizizz, Mentimeter, o Kahoot son las preferidas en la evaluación formativa. Con el uso de esas aplicaciones se pueden crear juegos, concursos, debates, encuestas y evaluaciones para los y las estudiantes. Las instituciones educativas y docentes tienen la ventaja de poder adaptar y personalizar estas herramientas a las necesidades específicas de su contexto educativo, según lo señalan Appiah y Van Toner (2018).

El *Joint Information Systems Committee* (JISC, 2006) afirma que existe una suposición de que las actividades de evaluación electrónica comprenden principalmente preguntas de opción múltiple (POM) y no se pueden utilizar para evaluar habilidades de pensamiento de orden cognitivo superior. Además, investigaciones posteriores han planteado dudas acerca de la confiabilidad y eficacia de la evaluación electrónica (Conole & Warburton, 2005), específicamente, se examina si las actividades de evaluación electrónica pueden evaluar el aprendizaje de orden superior y si los profesores disponen del tiempo necesario para desarrollar evaluaciones efectivas con el apoyo de la tecnología (Hepplestone & Helm, 2003). Gibbs (2006) también agrega que existe la preocupación de que las actividades de evaluación electrónica fomenten un enfoque de aprendizaje superficial (Biggs & Tang, 2011), y que cualquier implementación de la evaluación electrónica debe considerar esto.

Sin embargo, Farrell y Rushby (2016), el JISC (2006) y Jordan (2011) cuestionan los supuestos anteriores al decir que, si las POM son correctamente diseñadas, aún pueden evaluar habilidades de orden superior. Varios autores como Crisp (2011), Craven (2009), el JISC (2006) y Howarth (2015) coinciden en que la evaluación electrónica es más que solo POM, ya que los wikis, los blogs, las autoevaluaciones, las revisiones por pares, las preguntas de escenarios, el software de simulación, los juegos de roles y observaciones pueden utilizarse en evaluaciones electrónicas para valorar las habilidades de pensamiento de orden superior de los estudiantes, así como sus habilidades prácticas.

Winkley (2010) concuerda en que las actividades de evaluación electrónica ofrecen el potencial para impulsar a los estudiantes a adoptar un enfoque de aprendizaje profundo, según lo descrito por Biggs y Tang (2011), incentivándolos a comprometerse de manera significativa con las tareas asignadas.

Un profesor puede decidir crear una evaluación fácil que requiere menos tiempo o esfuerzo para calificar, o invertir tiempo para crear una evaluación que requiera un pensamiento de orden superior del estudiante. Por lo tanto, se cree que las actividades de evaluación electrónica tienen la capacidad de fomentar un enfoque de aprendizaje

profundo (Biggs & Tang, 2011) y evaluar el pensamiento de orden superior. Sin embargo, esto solo es posible si el profesor tiene la experiencia y la formación para crear preguntas o tareas de alta calidad que no solo requerirán que los estudiantes recuerden conceptos, sino también aplicar esos conceptos, analizar información, evaluar resultados y crear nuevas ideas o soluciones. Por lo tanto, los profesores deben ser capaces de proporcionar tareas que mejoren el aprendizaje de los estudiantes y de esta manera mantener altas sus expectativas.

Según Crisp (2011), no es difícil construir POM. Sin embargo, generar POM bien contruidos implica un esfuerzo considerable. Las POM son fáciles de calificar por medio del software de computadora apropiado y luego proporcionar retroalimentación instantánea. Sin embargo, con respecto a las preguntas cortas, es relativamente difícil calificar las respuestas de los estudiantes automáticamente porque el profesor (evaluador) debe tomar algunas decisiones en términos de respuestas imprevistas que pueden considerarse aceptables.

Si un profesor no considera todas las posibles respuestas de los estudiantes a un pregunta o tarea, el software de la computadora podría calificar la mayoría de las respuestas de los estudiantes como incorrectas. La evaluación electrónica es apoyada por Craven (2009); quien destaca que en una "evaluación objetiva o convergente", cuando el esquema de calificación se basa en una sola respuesta correcta, la calificación automática es muy eficaz. Por otra parte, en una "evaluación subjetiva o divergente" en la que pueden aceptarse respuestas diferentes, la calificación automática puede ser muy difícil. En cuanto a la evaluación subjetiva, las tecnologías se incorporan en las evaluaciones electrónicas para facilitar el proceso de evaluación, pero ese proceso no está automatizado.

Con el surgimiento de las tecnologías educativas basadas en la informática, la minería de datos y las técnicas de Procesamiento del Lenguaje Natural (PLN) se han aplicado en el dominio de educación, especialmente en la evaluación educativa automática (Lachner et al., 2017). Sobre la base de conocimientos multidisciplinares de la



informática, la lingüística y la minería de datos educativos, se han desarrollado una variedad de sistemas o herramientas basados en computadora para automatizar el proceso de calificar y proporcionar retroalimentación para satisfacer las necesidades de diferentes contextos de evaluación escrita.

Dentro del espectro de materias enseñadas, la física se presenta como una disciplina que, por su naturaleza cuantitativa y conceptual, podría beneficiarse enormemente de estos métodos de evaluación computarizados. En este sentido, los inventarios de conceptos, herramientas que buscan evaluar la comprensión de los estudiantes sobre ideas fundamentales en física, se presentan como candidatos ideales para ser administrados y evaluados mediante plataformas digitales. Sin embargo, se destaca que los sistemas actuales de evaluación basados en computadora se centran principalmente en preguntas de opción múltiple, careciendo de un sistema adecuado para calificar ensayos y respuestas cortas. Esta limitación representa un desafío significativo, ya que la evaluación integral del conocimiento de los estudiantes requiere herramientas que puedan manejar diversos tipos de respuestas.

La implementación de estas tecnologías en el ámbito educativo no está exenta de desafíos. Es fundamental entender cómo los estudiantes perciben y aceptan estas herramientas tecnológicas en su entorno de aprendizaje. Aquí es donde los modelos de aceptación de tecnología entran en juego. Estos modelos nos proporcionan un marco teórico para analizar y comprender los factores que influyen en la aceptación, adopción y uso continuo de tecnologías. El análisis de estos factores es crucial para mejorar la implementación de plataformas digitales que puedan evaluar tanto preguntas de opción múltiple y respuestas cortas, logrando así un sistema de evaluación más completo y efectivo.

En este contexto, la motivación de esta investigación es explorar un enfoque de evaluación automática de preguntas de respuesta corta en un curso de física introductoria, utilizando un inventario de conceptos y analizar los factores que afectan la

aceptación de los estudiantes al usar este instrumento educativo. Dicho lo anterior, los objetivos que se plantean son los siguientes:

## **1.2 Objetivos**

### **1.2.1 Objetivo general**

Analizar el uso y la aceptación de una herramienta de evaluación automática de conceptos básicos de física basados en preguntas de respuesta corta, en estudiantes de una universidad privada de Puebla.

### **1.2.2 Objetivos específicos**

1. Identificar las técnicas y herramientas empleadas actualmente para la evaluación automática de respuestas cortas.
2. Implementar en una herramienta tecno-pedagógica las estrategias de evaluación automática identificadas para respuestas cortas.
3. Adaptar una escala de medición para evaluar la Aceptación de la Evaluación Basada en Computadora
4. Analizar los factores que afectan la intención de los estudiantes al utilizar la herramienta tecno-educativa para la evaluación automática de preguntas de respuesta corta.

Hay que mencionar, además que la pregunta de investigación que dio origen a los objetivos es la siguiente:

### **1.3 Pregunta de investigación**

¿Cuáles son los factores que influyen en el uso y aceptación de una herramienta de evaluación automática de conceptos básicos de física basados en preguntas de respuesta corta en estudiantes de una universidad privada de Puebla?

## 1.4 Hipótesis

Para atender el problema identificado y responder a la pregunta de investigación, es importante destacar que en este trabajo se utilizó un modelo de aceptación de tecnología para predecir y explicar el comportamiento de las personas en relación con la adopción de tecnología. Las hipótesis en este modelo se plantean para identificar y medir los factores que influyen en la aceptación de una tecnología.

En el contexto de esta investigación sobre la herramienta para la evaluación automática de conceptos básicos de física basados en preguntas de respuesta corta, sostenemos que existen relaciones significativas entre múltiples variables que influyen en la adopción y uso exitoso de esta tecnología en los alumnos. Estas relaciones se explorarán a través de una serie de 15 hipótesis relacionales basadas en el Modelo de Aceptación de Evaluación Basada en Computadora (Terzis & Economides, 2011) y que abarcan diferentes aspectos de la percepción, actitudes y comportamientos de los usuarios en relación con esta tecnología de evaluación:

H1: La percepción lúdica tendrá un efecto positivo en la intención de utilizar la Evaluación Basada en Computadora (EBC)

H2: La utilidad percibida tendrá un efecto positivo en la intención de utilizar la EBC

H3: La utilidad percibida tendrá un efecto positivo en la percepción lúdica

H4: La facilidad de uso percibida tendrá un efecto positivo en la intención de utilizar la EBC

H5: La facilidad de uso percibida tendrá un efecto positivo en la utilidad percibida

H6: La facilidad de uso percibida tendrá un efecto positivo en la percepción lúdica

H7: La autoeficacia informática tendrá un efecto positivo en la facilidad de uso percibida

H8: La influencia social tendrá un efecto positivo en la utilidad percibida

H9: Las condiciones facilitadoras tendrán un efecto positivo en la facilidad de uso percibida

H10: La expectativa de meta tendrá un efecto positivo en la utilidad percibida

H11: La expectativa de meta tendrá un efecto positivo en la percepción lúdica

H12: El contenido tendrá un efecto positivo en la utilidad percibida

H13: El contenido tendrá un efecto positivo en la percepción lúdica

H14: El contenido tendrá un efecto positivo en la expectativa de meta

H15: El contenido tendrá un efecto positivo en la intención de uso de la EBC

Es necesario recalcar que el Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM) se basa en modelos previos de aceptación tecnológica como el Modelo de Aceptación de Tecnología (*Technology Acceptance Model*, TAM) (Davis, 1989), la Teoría del Comportamiento Planificado (*Theory of Planned Behavior*, TPB) (Ajzen, 1991), y la Teoría Unificada de Aceptación y Uso de la Tecnología (*Unified Theory of Acceptance and Use of Technology*, UTAUT) (Venkatesh et al., 2003). Por lo tanto, las 15 hipótesis planteadas en este modelo se fundamentan en estas teorías y en la revisión de la literatura. Estas teorías sugieren que la percepción individual sobre la utilidad y facilidad de uso de una tecnología influye en su actitud hacia ella y, por consiguiente, en su intención de usarla. Además, se consideran factores contextuales y personales que pueden modificar estas percepciones y actitudes, como experiencias previas, expectativas y el contexto cultural y social del usuario.

Al considerar las percepciones de los usuarios y abordar las barreras identificadas a través de estos modelos, las instituciones educativas pueden optimizar el uso de la tecnología para mejorar la experiencia de aprendizaje de los estudiantes y el desempeño educativo en general (Granic & Marangunic, 2019).

## **1.5 Justificación**

### **1.5.1 Contexto y Situación Actual del Problema**

La evaluación de conceptos básicos de física dentro del ámbito educativo enfrenta retos considerables en cuanto a su eficacia y profundidad. Tradicionalmente, las evaluaciones se han apoyado en formatos que privilegian la memorización por encima del entendimiento conceptual y la capacidad de aplicar conocimientos. Estas metodologías, que incluyen cuestionarios de opción múltiple y preguntas de verdadero o falso, han restringido significativamente la capacidad de los educadores para evaluar con precisión

el nivel de comprensión y la aptitud de los estudiantes para emplear conceptos físicos en situaciones nuevas. La literatura académica refleja esta preocupación, señalando que los métodos de evaluación convencionales no logran capturar la complejidad del aprendizaje significativo en física (Mazur, 1997; Wieman, 2014).

En este contexto, surge la necesidad de explorar alternativas de evaluación que promuevan un aprendizaje más profundo y aplicado. Los avances en tecnología educativa ofrecen la posibilidad de desarrollar herramientas de evaluación automática que puedan evaluar respuestas abiertas, brindando una oportunidad para medir la comprensión conceptual de manera más efectiva. Sin embargo, la adopción de estas herramientas enfrenta barreras tanto en términos de desarrollo tecnológico como de aceptación por parte de la comunidad educativa (Beichner, 2014).

Por ejemplo, el sistema de gestión de aprendizaje Moodle fue lanzado por primera vez en 2002 y su sistema de cuestionarios ha estado en constante desarrollo desde su primera versión en 2003 (Hunt, 2012). Hunt identifica alrededor de treinta tipos diferentes de preguntas disponibles dentro de Moodle, pero aun así, una encuesta ad hoc de más de 50,000,000 preguntas de alrededor de 2,500 sitios de Moodle encontró que aproximadamente el 90% de las preguntas en uso eran preguntas de respuesta seleccionada, es decir, tipos de preguntas como opción múltiple o arrastrar y soltar donde se presentan opciones para que un estudiante elija, en contraste con "respuesta construida" donde los estudiantes escriben su propia respuesta.

El estado actual de la evaluación en física revela, por lo tanto, un campo en transición, buscando equilibrar la necesidad de evaluaciones rigurosas y eficientes con el deseo de fomentar un aprendizaje más integrado y aplicado. Este equilibrio es crucial para preparar a los estudiantes no solo para superar exámenes sino para enfrentar desafíos reales en el campo de la física, promoviendo así una educación más relevante y significativa. Este enfoque ha captado un interés particular en la universidad privada donde se llevó a cabo este estudio, especialmente en la escuela de ingeniería, que desde 2019 ha adoptado exámenes argumentativos que enfatizan las explicaciones detalladas

a preguntas específicas. No obstante, el desafío predominante con este formato de examen radica en que su evaluación se realiza manualmente por los docentes, lo cual plantea una oportunidad significativa para la innovación y mejora.

En esta investigación también se considera que, la adopción de la evaluación basada en computadora (EBC) enfrenta el reto significativo de la escasez de investigaciones que busquen identificar todos los constructos conductuales relacionados con este proceso. Aunque existen múltiples estudios que se basan en distintas perspectivas teóricas sobre la innovación tecnológica, ninguno hasta la fecha ha integrado y perfeccionado los modelos prominentes de adopción de innovación específicamente para la EBC. A pesar del considerable esfuerzo y enfoque en la investigación de la adopción de tecnologías de la información y comunicación, persisten vacíos significativos en este ámbito. Un ejemplo concreto es el estudio de Terzis y Economides (2011), que destaca como una de las escasas investigaciones enfocadas en formular un modelo para la adopción de la EBC.

Así mismo, Imtiaz y Maarop (2014) destacaron que, si bien numerosos estudios en el campo educativo se han centrado en el aprendizaje electrónico (e-learning), hay escasas investigaciones acerca de la EBC. Estos autores también mencionaron que la primera investigación sobre la adopción de evaluaciones electrónicas data de 2011, lo que sugiere que aún falta exploración en este ámbito. Esta situación refleja una brecha investigativa en el tema, evidenciando la relevancia de este estudio a escalas tanto nacional como global.

Dentro del marco de este estudio, la EBC se refieren a evaluaciones que se desarrollan de una manera que permite a los docentes crear, organizar, supervisar, administrar y reportar sobre estas evaluaciones. El uso de una computadora siempre es intrínseco a este tipo de evaluaciones. El factor clave para las EBC es que la computadora califica o evalúa las respuestas proporcionadas por los estudiantes de forma automática (Conole y Warburton, 2005). Una de las aportaciones adicionales de este trabajo es el énfasis en el formato de pregunta de respuesta corta (PRC) en lugar del tradicional formato de

pregunta de opción múltiple (POM). Aquí radican varias razones fundamentales para esta elección:

1. Profundidad de Respuesta: El formato PRC permite a los estudiantes expresar sus respuestas en sus propias palabras, ofreciendo una visión más profunda y matizada de su entendimiento. En lugar de simplemente reconocer una respuesta correcta entre opciones predeterminadas, el estudiante debe construir y articular su respuesta, lo que puede reflejar una comprensión más genuina del material.

2. Disminuye Conjeturas: En las POM, los estudiantes a menudo pueden eliminar respuestas claramente incorrectas y adivinar entre las restantes. Esto puede inflar artificialmente las puntuaciones y no reflejar una verdadera comprensión. El formato PRC minimiza la posibilidad de conjeturas acertadas.

3. Adaptabilidad a la Tecnología: Las herramientas de procesamiento de lenguaje natural (PLN) y las técnicas de análisis automático de texto han avanzado significativamente. Esto hace que la evaluación automática de respuestas cortas sea más viable, abriendo nuevas puertas para la EBC.

4. Modelo de Aceptación de Tecnología: Investigar cómo los usuarios perciben y aceptan el formato PRC en una EBC utilizando un modelo de aceptación de tecnología proporcionará información valiosa sobre cómo diseñar e implementar sistemas de evaluación más efectivos y amigables para el usuario.

### **1.5.2 Evaluación actual de conceptos de la física**

La comprensión conceptual se refiere a la capacidad de aplicar el conocimiento en una variedad de contextos. La evaluación formativa es crucial para que ocurra la comprensión conceptual (Bell & Cowie, 2001). La comprensión conceptual requiere que los estudiantes sean conscientes de sus concepciones y conceptos erróneos, y cómo hacer modificaciones a su pensamiento, a través de retroalimentación formativa. A menudo,

los estudiantes recurren a la memorización de hechos en lugar de la comprensión conceptual.

Tal como se utiliza actualmente en la investigación sobre enseñanza de la física y áreas temáticas similares, un inventario de conceptos está diseñado principalmente como un instrumento para aprender sobre la comprensión de los estudiantes de un área temática específica (Smith & Tanner, 2010). En su forma actual, la mayoría de los inventarios de conceptos hacen uso del tipo de pregunta de opción múltiple. Las preguntas de opción múltiple son un tipo de pregunta de respuesta seleccionada, lo que significa que las posibles respuestas se le dan al estudiante como una lista de opciones (Jordan, 2013). Este formato restringe la cantidad de formas en que los estudiantes pueden responder las preguntas y mostrar sus aciertos y errores.

Por el contrario, más información sobre la comprensión conceptual del estudiante estaría disponible si los estudiantes escribieran sus propias respuestas en un formato de pregunta de respuesta construida (Jordan, 2013), y esto podría lograrse mediante el uso de preguntas de respuesta corta (PRC) (Rebello & Zollman, 2004). A su vez, esto proporcionaría información útil a los profesores de física, en particular con respecto a los conceptos erróneos, y así permitirles centrar su instrucción en estas áreas.

Hacer uso de PRC permite a los estudiantes dar una gama mucho más amplia de respuestas de lo que podrían con preguntas de opción múltiple, pero para estas PRC, las respuestas toman mucho más tiempo para calificar a mano que sus contrapartes de opción múltiple. Por lo tanto, se considera pertinente el desarrollo de la presente investigación, donde se explora el uso de preguntas en un formato de respuesta corta y su correspondiente calificación automática. Esto es posible, en principio, mediante tecnologías como las utilizadas en las preguntas de coincidencia de patrones (*Pattern Match*) en *OpenMark*. *OpenMark* es un sistema de evaluación asistido por computadora (CAA) desarrollado por la Open University, basado en el aprendizaje asistido por



computadora (Butcher & Jordan, 2010). Este sistema sirvió de base para el desarrollo del tipo de pregunta *Pattern Match*<sup>1</sup> en el motor de preguntas de Moodle.

Dado que la investigación empírica sobre esta tecnología es escasa, será útil comenzar adaptando un inventario de conceptos de física ya bien establecido en una versión de formato de respuesta libre. Un inventario de conceptos adecuado para esta tarea es el Inventario de conceptos de fuerza (*Force Concept Inventory*, FCI) (Hestenes et al., 1992), que es el primero y el más conocido de los inventarios de conceptos de física (Smith & Tanner, 2010). La versión tradicional de opción múltiple del FCI ya ha sido ampliamente utilizada y probada, como lo resumen Scott y Schumayer (2017), y está ampliamente aceptado como un instrumento educativo válido y confiable.

## 1.6 Delimitación del objeto de estudio

Para justificar la necesidad de herramientas de evaluación automática en la enseñanza de conceptos básicos de física, es esencial identificar áreas específicas de mejora basadas en datos y literatura existentes. Por ejemplo, estudios han demostrado que la calificación manual de respuestas abiertas puede ser inconsistente debido a sesgos personales, además de ser un proceso lento y laborioso. Un vacío importante es la falta de herramientas automatizadas que puedan evaluar eficazmente la creatividad y el razonamiento complejo en respuestas cortas sin sacrificar la objetividad. Este problema subraya la necesidad de desarrollar sistemas de evaluación que combinen técnicas de procesamiento de lenguaje natural y aprendizaje automático para proporcionar evaluaciones precisas y justas, abordando así las limitaciones de tiempo y subjetividad (Zawacki-Richter et al., 2019; Hussein et al., 2019).

Por otro lado, la implementación de preguntas de respuesta corta, cuando están bien diseñadas y aplicadas, puede ofrecer una alternativa valiosa al permitir a los estudiantes

---

<sup>1</sup> *Pattern Match*: es un tipo de pregunta dentro del motor de preguntas de Moodle que permite crear preguntas de respuesta libre.

expresar su comprensión de forma más completa y reflexiva. Sin embargo, la poca disponibilidad de herramientas automatizadas para evaluar este tipo de respuestas plantea un desafío significativo en términos de eficiencia y escalabilidad de la evaluación (del Gobbo, et al., 2023). La necesidad de desarrollar soluciones tecnológicas de fácil uso que permitan la evaluación automática de respuestas cortas es, por lo tanto, un área crítica de mejora que puede tener un impacto sustancial en la enseñanza y el aprendizaje de la física.

En este contexto, esta investigación se articula en las siguientes dos etapas:

1. Desarrollo e implementación del sistema de evaluación automática utilizando el formato de pregunta de respuesta corta.
2. Análisis de los factores que afectan la intención de los estudiantes al utilizar el sistema de evaluación automática de preguntas de respuesta corta mediante un Modelo de Aceptación de Evaluación Basada en Computadora.

En relación con la primera etapa, cabe destacar el aspecto de que el sistema fue implementado en el sistema gestor de aprendizaje MOODLE a través de un plugin llamado *Pattern Match* (Butcher & Jordan, 2010), dirigido a usuarios sin experiencia en PLN. La principal fortaleza de esta tecnología es que proporciona una herramienta de creación de preguntas que puede ser utilizada por un autor de preguntas sin conocimientos de PNL. El software de evaluación de texto libre de *Pattern Match* comprende el propio motor de calificación, proporcionado como un servicio web, y una herramienta de autoría, utilizada para configurar las reglas de calificación para cada pregunta.

La interfaz de usuario de *Pattern Match* (Butcher & Jordan, 2010) fue diseñada para simplificar la tarea de generar plantillas, protegiendo al usuario de las complejidades del PLN, permitiéndoles enfocarse en aspectos clave como la definición de respuestas modelo y la selección de palabras clave pertinentes para cada una. Dicha interfaz está ideada para profesionales del ámbito educativo, como docentes, quienes pueden aprovecharla sin necesidad de ser expertos en PLN.

Para la creación del formato de las preguntas de respuesta corta con evaluación automática se tienen las siguientes dos consideraciones:

1. El punto de partida es hacer una pregunta que sea posible calificar con “precisión” (lo que se considera bien o mal).
2. Las respuestas no deben tener más de 20 palabras, sin contar los espacios.

En cuanto a la segunda etapa de esta investigación, el objetivo principal es analizar los factores que determinan la aceptación, por parte de los estudiantes de primer semestre de ingeniería, de una herramienta computacional destinada a evaluar de forma automática conceptos básicos de física. Dicha herramienta utiliza preguntas de respuesta corta, en contraste con las preguntas de opción múltiple. Para guiar este análisis, nos apoyaremos en el Modelo de Aceptación de Evaluación Basada en Computadora (*Computer Based Assessment Acceptance Model*, CBAAM) de Terzis y Economides (2011). Es importante destacar que este modelo fue desarrollado específicamente para evaluaciones que contienen preguntas en formato de opción múltiple.

Imtiaz y Maarop (2014) señalan que la evaluación electrónica representa el futuro de las evaluaciones educativas, pero es un campo que ha recibido poca atención investigativa. La relevancia de explorar este tema radica en la necesidad de comprender los factores que influyen en la aceptación y eficacia de estas tecnologías, con el fin de promover prácticas de evaluación más innovadoras y eficientes (Kiryakova, 2021).

Con los resultados de la investigación se pretende contribuir al conocimiento sobre la adopción y aplicación efectiva de la Evaluación Basada en Computadora (EBC) utilizando el formato de preguntas de respuesta corta, una herramienta pedagógica que, aunque es prometedora, todavía no ha sido explorada en profundidad. Al aplicar un modelo de aceptación de tecnología, no sólo investigaremos las percepciones, actitudes y posibles barreras de los usuarios hacia la EBC, sino que también proporcionaremos

información relevante sobre cómo mejorar la aceptabilidad y la integración de estas herramientas en ambientes educativos. Esperamos que nuestros hallazgos sirvan como una base para el diseño e implementación de futuras plataformas de EBC, y que además contribuyan al diálogo académico en torno a la adaptación y uso óptimo de las tecnologías educativas.

## **1.7 Alcance del estudio**

Este estudio es una aportación significativa a la literatura existente sobre la adopción de la evaluación electrónica en la educación superior, utilizando un enfoque de diseño cuantitativo. La investigación se realizó implementando un sistema de evaluación automática utilizando el formato de pregunta de respuesta corta y, posteriormente se aplicó un modelo de adopción para la evaluación electrónica en el que se analizaron los constructos que afectan la intención de uso por parte de los estudiantes. No obstante, como en cualquier otro estudio de investigación limitado por el alcance y el tiempo, se tienen las siguientes delimitaciones y limitaciones:

### **1.7.1 Delimitaciones**

1. Enfoque en primer semestre: el estudio se centra exclusivamente en estudiantes de primer semestre de ingeniería, lo cual brinda una mirada detallada a las etapas iniciales del aprendizaje universitario, pero limita la generalización de los resultados a otros niveles académicos o disciplinas.
2. Contexto de una Universidad Privada: La elección de una sola institución para el estudio ayuda a controlar variables contextuales, pero también restringe la extensión de las conclusiones a otras universidades con diferentes culturas académicas o recursos.
3. Competencia Digital: Se asumió que los estudiantes poseen las habilidades básicas necesarias para manejar una computadora. Esta suposición podría no reflejar la realidad

de todos los participantes, afectando la interacción con la herramienta de evaluación y sus resultados.

Además, es fundamental destacar que en este estudio no se toman en cuenta los problemas de aprendizaje o cualquier tipo de discapacidad de los participantes. Este aspecto delimita aún más el alcance de la investigación, ya que no se evalúa cómo estas variables podrían influir en la adopción y uso de la herramienta de evaluación automática, lo cual podría ser relevante para entender completamente su accesibilidad y eficacia en un contexto educativo diverso.

### **1.7.2 Limitaciones**

1. Tamaño de la Muestra: Aunque 84 sujetos pueden proporcionar datos significativos, el tamaño limitado de la muestra puede afectar la capacidad de generalizar los resultados a una población más amplia.

2. Respuestas de Autoinforme: El uso de herramientas de evaluación basadas en respuestas de autoinforme puede introducir sesgos. Por ejemplo, los participantes pueden elegir no responder honestamente; esto podría ser para presentarse a sí mismos de una manera favorable. Además, los participantes pueden exagerar sus logros, habilidades o experiencias.

### **1.8 Paradigma de la investigación**

Esta investigación adopta un enfoque cuantitativo, de tipo cuasi-experimental y de alcance explicativo. De acuerdo con Hernández-Sampieri y Mendoza Torres (2018), el enfoque cuantitativo es un conjunto de procesos secuencial y probatorio que, a partir de una idea, va acotándose hasta derivar en objetivos, preguntas de investigación e hipótesis.

Dentro de los objetivos, pregunta de investigación e hipótesis se observa la necesidad de analizar los constructos que afectan la intención de utilizar una herramienta

tecnológica para calificar de forma automática las respuestas en formato de texto libre proporcionadas por los alumnos y que se relacionan con conceptos básicos de física. La recolección y análisis de datos se dividió en dos fases: primero, los alumnos respondieron a la evaluación basada en computadora. Al finalizar el examen, los participantes respondieron el cuestionario CBAAM. La segunda fase, consistió en analizar el modelo de adopción de la evaluación basada en computadora, en este trabajo se utilizó el enfoque de modelado de ecuaciones estructurales (SEM) basado en mínimos cuadrados parciales (PLS). Este enfoque es particularmente adecuado para modelos exploratorios y complejos donde las relaciones entre variables latentes son de interés primordial. La implementación de PLS-SEM se realizó siguiendo un conjunto de reglas y pautas establecidas para garantizar la precisión en la estimación estadística del modelo, según lo sugieren Hair et al. (2016).

## **1.9 Línea de investigación**

El proyecto de tesis se sitúa dentro de la línea de investigación del programa doctoral que corresponde a: Desarrollo Tecnológico e Innovación en los Sistemas y Ambientes Educativos. Esto debido a que en este trabajo se tienen dos etapas principales: la primera es la implementación de un instrumento educativo que se puede utilizar para evaluar la comprensión de los estudiantes de conceptos básicos en física con tecnología mejorada, en particular la calificación automática de preguntas de respuesta corta. La segunda es la aplicación de un modelo de Aceptación de Evaluación Basada en Computadora, con el fin de comprender y predecir cómo los estudiantes adoptan y utilizan herramientas de evaluación digital.

Este trabajo implica aplicar herramientas y técnicas del Procesamiento del Lenguaje Natural (PLN). Como resultado de la revisión de literatura, para el desarrollo de esta investigación se eligió utilizar una tecnología desarrollada en la comunidad de software libre y que se integra al sistema gestor de aprendizaje MOODLE a través de un plugin llamado Pattern Match (Butcher & Jordan, 2010). Con esta integración, ya se pueden construir preguntas en formato de respuesta corta y con el potencial de calificarse de forma automática. Esta tecnología está basada en un algoritmo de coincidencia de

palabras que busca palabras clave y tiene en cuenta sinónimos y errores ortográficos, sin embargo, no hace uso de la gramática o la sintaxis en sus algoritmos.

### **1.10 Conclusiones del capítulo**

Este capítulo introduce la investigación destacando la necesidad de métodos de evaluación innovadores en la enseñanza de la física, frente a los enfoques tradicionales que priorizan la memorización sobre la comprensión conceptual. Además, subraya el potencial que representan las evaluaciones basadas en computadora (EBC) para ofrecer una evaluación más eficiente y variada, aunque señala los retos en su adopción y desarrollo tecnológico. Al centrarse en el uso de preguntas de respuesta corta (PRC) y adoptar el modelo CBAAM, el capítulo identifica un vacío en la investigación sobre la aceptación de la EBC y argumenta la importancia de este estudio tanto a nivel nacional como global, proyectando una contribución significativa al avance de las prácticas educativas en física y la integración de tecnologías innovadoras en la educación.

### **1.11 Organización de la tesis**

La estructura de esta tesis se compone de seis capítulos, cada uno dedicado a explorar elementos clave del tema de investigación. La organización de los capítulos es la siguiente:

El capítulo 1 presenta el contexto y la importancia del estudio. Se detallan los objetivos de la investigación, las preguntas de investigación y las hipótesis planteadas. Además, se describen la justificación del estudio y su relevancia académica y práctica.

El capítulo 2 sobre el estado del arte ofrece una visión general del estado actual de la investigación en cuatro áreas fundamentales: el uso de Inventarios de Conceptos en Física para la evaluación del entendimiento de los estudiantes sobre conceptos físicos esenciales, la aplicación de Evaluación Automática tanto en preguntas de opción múltiple como de respuesta corta, los Modelos de Aceptación de Tecnología que ilustran los factores detrás de la adopción de innovaciones tecnológicas, y los Modelos de

Aceptación de Evaluaciones Electrónicas. Este capítulo sintetiza investigaciones previas para contextualizar la integración de la tecnología en la evaluación de la física, subrayando tanto las potencialidades como los desafíos asociados.

El capítulo 3 detalla las teorías fundamentales, modelos, conceptos y estrategias que fundamentan la presente investigación, facilitando una comprensión más profunda de los retos y ventajas de la evaluación basada en computadora. Se pone especial énfasis en el uso de preguntas de respuesta corta y se examina un Modelo de Aceptación de Evaluación Basada en Computadora, todo ello en el contexto del entorno educativo actual.

El capítulo 4 explica la filosofía de investigación, el enfoque, el diseño y los métodos de análisis de datos elegidos. Se enfatiza la justificación de estas elecciones metodológicas. Además, se describe el proceso de selección de la muestra, las consideraciones éticas y el procedimiento de recolección de datos, proporcionando una base firme para los resultados y discusiones subsiguientes.

El capítulo 5 presenta y analiza los resultados obtenidos de la investigación, utilizando los métodos de análisis de datos detallados en el capítulo metodológico. Se examina la relevancia de estos hallazgos dentro del marco teórico descrito en capítulos previos, evaluando su consonancia o discrepancia con teorías y estudios existentes. Se exploran las implicaciones de estos resultados para la práctica educativa y la integración tecnológica en la enseñanza de la física, resaltando las ventajas y retos encontrados

El capítulo 6 resume las conclusiones de la investigación, destacando cómo los hallazgos responden a las preguntas planteadas y sus implicaciones teóricas y prácticas, especialmente en el ámbito de la evaluación educativa basada en computadora. Reflexiona sobre las limitaciones del estudio y propone áreas para futura investigación, subrayando la relevancia de estos resultados para mejorar la práctica educativa y la adopción tecnológica en la evaluación.



## **CAPÍTULO 2. ESTADO DEL ARTE**

Para proporcionar una visión general de la investigación actual e ilustrar el campo de la tecnología disponible para ayudar al desarrollo de prácticas eficaces de evaluación electrónica en la educación superior se utilizó una estrategia metodológica de tipo bibliográfico, que consistió en la búsqueda de información en bases de datos académicas. Este capítulo busca explorar y analizar el estado actual de la investigación y desarrollo en tres campos de estudio que se relacionan con el objetivo de este trabajo: la Evaluación basada en Computadora, la Evaluación Automática de Textos Cortos y los Modelos de Aceptación de Tecnologías.

Los descriptores utilizados para realizar la búsqueda de los artículos pertinentes para el análisis fueron: evaluación electrónica (“e-assesment”, “Computer-Based Assessment (CBA)”), evaluación automática de textos cortos (“Automated Short Answer Grading (ASAG)”, modelos de aceptación de tecnología (“Technology Adoption Models”). Estos fueron considerados en español e inglés, y la búsqueda se realizó a través de las bases de datos Scopus y Web of Science. La literatura sobre evaluación automática es vasta y en la última década aumentó considerablemente. Este campo de investigación es amplio y multidisciplinario, por lo tanto, para los objetivos de este trabajo se identificaron las cuatro trayectorias de investigación más relevantes para el contexto de este estudio. La intención de esta descripción general es recorrer estas diferentes comunidades de investigación y construir puentes entre ellas para motivar y construir un marco conceptual.

La revisión de literatura se concentra en las siguientes cuatro áreas:

- Inventario de conceptos en física
- Evaluación automática de preguntas de opción múltiple y de respuesta libre.
- Modelos de aceptación de tecnología
- Modelos de aceptación de evaluaciones electrónicas

Analizar y sintetizar la literatura existente sobre estas cuatro áreas es relevante de diversas maneras, ya que apoya: (1) la identificación de temas de investigación actualmente poco explorados; (2) la selección de modelos de adopción tecnológica apropiados para esta investigación; y (3) la formación de metodología alrededor de la pregunta y objetivos concretos de la investigación.

A partir de los contenidos que se exploraron, se realizó una depuración de la información, donde se eligieron un total de 64 artículos, publicados entre 2000 y 2023 por equipos de investigación y universidades. Los criterios de exclusión fueron: Artículos repetidos y publicaciones que no sean artículos. En cuanto a los criterios de inclusión se consideraron los siguientes elementos:

- Publicaciones de 2000 a 2023
- Artículos que traten aspectos empíricos y específicos de la evaluación electrónica.
- Artículos que se refieran a la educación superior

En este contexto, cabe mencionar que dentro de las líneas de investigación correspondientes al programa doctoral, nuestro trabajo corresponde a la temática de Desarrollo Tecnológico e Innovación en los Sistemas y Ambientes Educativos. Dado que se utilizan muchos términos para referirse a la integración de las tecnologías de información y comunicación en las prácticas de evaluación, es indispensable conceptualizar este proceso.

En la literatura sobre evaluación electrónica (e-assessment), autores como Jamil et al. (2012), Al-Qdah y Ababneh (2017), Jordan (2013), Bukie (2014) y Timmis et al. (2016), han identificado un amplio espectro de términos que generalmente se consideran sinónimos en este campo. Estos términos abarcan desde métodos y tecnologías específicas de evaluación hasta diferentes formatos de presentación. La clasificación de estos términos puede ayudar a comprender mejor su uso y aplicación en contextos educativos:

### ***Tecnologías y Métodos de Evaluación:***

- Evaluación Basada en Computadora
- Tecnología para la Evaluación
- Pruebas Basadas en Computadora
- Evaluación Asistida por Computadora
- Pruebas Asistidas por Computadora

***Mejoras y Habilitación por Tecnología:***

- Evaluación Mejorada por la Tecnología
- Evaluación Habilitada por la Tecnología

***Formatos de Presentación de Evaluación:***

- Evaluación Informatizada
- Pruebas Informatizadas
- Evaluación Basada en la Web
- Evaluación Electrónica
- Pruebas Electrónicas
- Evaluación en Línea

La comprensión de estos términos es crucial para la investigación y la práctica en la educación superior, ya que refleja la diversidad y la riqueza de opciones disponibles para la evaluación electrónica. La selección de uno sobre otro puede depender de varios factores, incluidos los objetivos de aprendizaje específicos, la infraestructura tecnológica disponible y las preferencias tanto de los profesores como de los estudiantes.

A continuación, se presentan un resumen sobre los hallazgos de la revisión de literatura que involucra los cuatro campos del conocimiento abarcados por esta investigación.

**2.1 Una descripción general de los inventarios de conceptos en física**

Hay muchos inventarios de conceptos que se utilizan en física y astronomía. Un inventario de conceptos es un instrumento de opción múltiple de nivel investigación, que está diseñado para evaluar la comprensión conceptual de un tema en particular (Lindell et al., 2007), con un mínimo contenido matemático. El inventario se basa en una serie de

conceptos clave del tema (Jorion et al., 2015). Para cada reactivo en el inventario de conceptos, hay una serie de respuestas, que contienen una respuesta correcta y otros distractores de respuestas incorrectas. Las respuestas distractoras están diseñadas para corresponder a los conceptos erróneos más comunes en los estudiantes (Sadler et al., 2010). Dick-Pérez et al. (2016) observaron que la clave para la construcción de un inventario de conceptos efectivo radica en la selección de distractores apropiados. En su trabajo, el diseño de estos distractores también resultó ser uno de los aspectos más difíciles de construir en su inventario de conceptos.

El objetivo de un inventario de conceptos puede variar (Smith & Tanner, 2010). Sin embargo, la mayoría de los inventarios están diseñados con la idea de probar la efectividad de estrategias de enseñanza (Porter et al., 2014). Para esto, se les pide a los estudiantes que contesten el inventario de conceptos antes de la instrucción, el cual se conoce en la literatura como el pre-test, y luego se les aplica de nuevo después de la instrucción, lo que se conoce en la literatura como post-test. Los puntajes previos y posteriores a la prueba de todos los estudiantes se comparan para medir la efectividad de los métodos de enseñanza utilizados por el instructor (Bailey et al., 2012). Los resultados de esta comparación se pueden utilizar para orientar una instrucción futura al abordar los conceptos erróneos que parecen haber permanecido. Este aspecto de orientación pedagógica de los inventarios de conceptos los convierte en instrumentos populares para usar en materias STEM (*Science, Technology, Engineering, Arts, and Mathematics*; Ciencia, Tecnología, Ingeniería, Arte y Matemáticas). Además, ser capaz de probar cuantitativamente la efectividad de las intervenciones de enseñanza hace que su uso sea atractivo para los instructores.

Existen más de 60 inventarios de conceptos en física y astronomía que abarcan una amplia variedad de temas y niveles educativos (Madsen et al., 2017). Estos inventarios han sido desarrollados mediante un proceso riguroso que incluye la identificación de ideas de los estudiantes, pruebas con expertos y revisión continua. La mayoría de los inventarios son de opción múltiple y se utilizan tanto como pruebas previas como posteriores para evaluar el aprendizaje de los estudiantes. Se pueden encontrar detalles

sobre cada inventario, incluyendo su contenido, nivel de validación y guías de implementación, en el sitio web *PhysPort.org*.

El inventario de conceptos más utilizado en física es el FCI de Hestenes et al. (1992), el cual ha tenido un impacto significativo en la mejora de la educación en física durante las últimas dos décadas. El FCI ha sido administrado a miles de estudiantes en todo el mundo y ha demostrado que los métodos de enseñanza basados en la investigación, como la participación interactiva, conducen a mejoras dramáticas en el aprendizaje de los estudiantes. Este inventario se destaca por incluir las ideas cotidianas y el lenguaje natural de los estudiantes en las opciones de respuesta, lo que requiere que los estudiantes tengan una comprensión sofisticada de los conceptos para tener un buen desempeño en la prueba.

Como se mencionó en el capítulo uno, el objetivo de la primera etapa de esta investigación es desarrollar una evaluación basada en computadora utilizando el formato de preguntas de respuesta corta. Para ello, se ha seleccionado el FCI de Hestenes et al. (1992), considerado el más adecuado para los propósitos de esta investigación debido a su validez y relevancia comprobadas en estudios previos.

## **2.2 Preguntas de opción múltiple**

Para evaluar la comprensión conceptual de la física, se emplean comúnmente inventarios de conceptos que utilizan el formato de pregunta de opción múltiple. Las primeras preguntas de opción múltiple se utilizaron para probar reclutas del ejército estadounidense en la Primera Guerra Mundial (Mathews, 2006), mientras que las preguntas de opción múltiple se utilizaron por primera vez en el contexto educativo a finales del siglo XX (Bacon, 2003). Una ventaja de las preguntas de opción múltiple es que son versátiles, lo que significa que pueden ser utilizadas para probar una amplia gama de temas (Bull & McKenna, 2003); por otro lado, las preguntas de respuesta libre son ventajosas porque se pueden usar para probar contenido preciso (Betts et al., 2009; Ferrao, 2010). Debido a la forma en que las preguntas de opción múltiple están

configuradas, es difícil hacer uso de ellas para evaluar resultados de aprendizaje más avanzados (Conole & Warburton, 2005), aunque ha habido algunos esfuerzos para hacerlo (Itza-Ortiz et al., 2003; Gwinnett & Cassella, 2011).

La principal motivación para usar preguntas de opción múltiple sobre otros tipos de preguntas ha sido que estas se pueden evaluar rápidamente, lo que hace una diferencia práctica para los académicos que tienen grupos numerosos para evaluar (Woodford & Bancroft, 2005). Sin embargo, los mismos autores afirman que la calidad de las preguntas de opción múltiple debe ser alta para justificar su uso, lo que requiere que tales preguntas sean cuidadosamente diseñadas y probadas.

Un defecto común identificado con las preguntas de opción múltiple es que a veces es posible descifrar la respuesta correcta analizando cada una de las opciones de respuesta para buscar la opción menos plausible, lo que significa que tales preguntas no están evaluando las habilidades que se supone que deben (Sangwin, 2013). Adicionalmente, es posible que los estudiantes respondan correctamente a preguntas de opción múltiple adivinando; cuando esto sucede, el estudiante no está demostrando la comprensión prevista (Crisp, 2007). Un argumento en contra de esta crítica establece la poca probabilidad de que un estudiante pueda aprobar una prueba completa de opción múltiple a través de adivinanzas solamente (Downing, 2003); sin embargo, es lógico pensar que las conjeturas exitosas y bien ubicadas en preguntas de opción múltiple pueden hacer una diferencia en la calificación otorgada a un estudiante en un caso límite (Burton, 2005).

Además, hay diferentes procesos cognitivos involucrados en responder preguntas de respuesta libre y preguntas de opción múltiple (Nicol, 2007). Con esto en mente, hay maneras en que se puede aumentar la eficacia de las preguntas de opción múltiple, por ejemplo, desarrollando opciones de distracción plausibles (Dick-Perez et al., 2016), o considerando cuidadosamente el orden en que se hacen las preguntas (Gray et al., 2002). Para asegurar que las preguntas sean capaces de probar la comprensión para la que están diseñadas, es importante preparar preguntas de alta calidad,

independientemente de si son de opción múltiple o respuesta libre (Bull & McKenna, 2003).

Simon y Snowdon (2014) compararon preguntas de opción múltiple y de respuesta libre en el contexto de la evaluación de las habilidades de codificación. A los estudiantes se les dio una selección de múltiples preguntas de elección y de respuesta libre basadas en la interpretación del código, y se encontró que los estudiantes, en general, encontraron que las preguntas de respuesta libre eran las más difíciles de los dos tipos de preguntas. Los autores señalaron que, con preguntas de opción múltiple, el estudiante necesita elegir de una lista de ideas construida por otra persona, y también puede emplear estrategias de adivinación y eliminación para llegar a la respuesta; por otro lado, los autores postularon que las preguntas de respuesta libre requieren que los estudiantes piensen más profundamente, ya que se les exige que construyan sus propias respuestas, y que además se puede obtener información sobre el pensamiento de los estudiantes a partir de las respuestas en formato libre de los estudiantes como un resultado de esto.

Una comparación de la respuesta de los usuarios a preguntas de opción múltiple y de respuesta libre fue realizada por Bridgeman (1991) al usar el Examen General de Registro de Graduados (GREGT), una prueba estandarizada realizada por los solicitantes a las escuelas de posgrado de Estados Unidos, como base para el estudio. A un conjunto de estudiantes se les dieron versiones de opción múltiple de algunas de las preguntas del GREGT para responder, mientras que a un conjunto de control de estudiantes se les dieron versiones de respuesta libre de las mismas preguntas para responder. Las respuestas correctas e incorrectas de la versión de opción múltiple de las preguntas se compararon con las respuestas dadas a la versión de respuesta libre de las preguntas, y se encontró que las puntuaciones en las preguntas de opción múltiple y las versiones de respuesta libre correspondientes eran similares. Bridgeman concluyó que, bajo ciertas circunstancias, las preguntas de opción múltiple y las contrapartes de respuesta libre se pueden usar indistintamente para evaluar la comprensión del estudiante.

Otro estudio que compara preguntas de opción múltiple y de respuesta libre fue llevado a cabo por Funk y Dickson (2011), cuyo objetivo era investigar el efecto de estos tipos de preguntas en el rendimiento de los estudiantes. Con este fin, los autores crearon versiones de opción múltiple y respuesta libre de un conjunto de preguntas del examen de psicología. Estas preguntas fueron divididas y dadas a cincuenta estudiantes para tratar de la siguiente manera: veinticinco estudiantes hicieron 10 preguntas de respuesta libre antes de hacer una prueba de selección múltiple de 50 preguntas; los otros veinticinco hicieron 10 preguntas de respuesta libre después de hacer una prueba de selección múltiple de 50 preguntas. En ambos casos, el rendimiento en las versiones de opción múltiple de las preguntas fue mayor que en las versiones de respuesta libre, lo que indica que los estudiantes encontraron que las variantes de respuesta libre de las preguntas eran más difíciles que sus contrapartes de opción múltiple. Este hallazgo concuerda con el trabajo previo de Hudson (2010), quien encontró que las preguntas de opción múltiple fueron más fáciles de responder que las preguntas de respuesta libre para estudiantes de química de secundaria.

Woodford y Bancroft (2005) señalaron que la aparente dificultad de una pregunta de opción múltiple puede revelar fallas en su construcción: si una pregunta parece fácil, entonces sus distractores pueden ser ineficaces; si parece difícil, entonces los distractores pueden ser engañosos, o la redacción de la pregunta puede ser ambigua. Cualquiera que sea la dificultad que revele la pregunta, se deben tomar las medidas adecuadas para mejorarla si hay un problema, como se describe en las prioridades de diseño de Woodford y Bancroft. En el nivel de prueba, el número de preguntas en la prueba en general, así como el orden de estas preguntas, afectará la eficacia de la prueba. En el nivel de las preguntas, el número de opciones dadas a la pregunta de opción múltiple, el orden de estas opciones, la redacción de la pregunta en sí y la redacción de los elementos de distracción afectan a la eficacia de la pregunta. Woodford y Bancroft argumentaron que cambiar estos factores a nivel de prueba y a nivel de preguntas puede conducir a preguntas de elección múltiple de mejor calidad.



Woodford y Bancroft propusieron que, siempre que se apliquen sus prioridades de diseño, las preguntas de opción múltiple se pueden usar para probar un nivel más profundo de comprensión, pero los ejemplos que se dan se refieren a cuestiones matemáticas. Sin embargo, se puede argumentar que las preguntas de opción múltiple, sin importar cuán refinadas o bien diseñadas, pueden ser inadecuadas para probar la comprensión conceptual (Conole & Warburton, 2005), que es más descriptiva y menos procedimental.

Shuhidan et al. (2010) dieron cuestionarios a 66 instructores de programación para investigar sus perspectivas de usar preguntas de opción múltiple para evaluar a sus estudiantes. Más de un tercio de los instructores confiaban en el uso de preguntas de elección múltiple de alta calidad para evaluar los resultados de aprendizaje de los estudiantes. Los instructores elegirían usar preguntas de opción múltiple con sus estudiantes porque pueden ser contestadas por estudiantes de bajo desempeño, son adecuadas para la revisión y porque preparan a los estudiantes para preguntas de opción múltiple en otros cursos.

Por otro lado, algunos instructores fueron desalentados a usar preguntas de opción múltiple porque sentían que eran demasiado fáciles, y que animaban a los estudiantes a adivinar. De estos instructores, algunos preferían el uso de preguntas en formato de respuesta libre, tanto de ensayo como de respuestas cortas, para poner a prueba la comprensión de los y las estudiantes. Según el estudio, se encontró que más del 90% de los instructores preferían utilizar preguntas de opción múltiple en sus evaluaciones sumativas, mientras que alrededor del 10% de los encuestados no apoyaban el uso de este formato y preferían preguntas en formato de respuesta libre.

También parece haber diferencias entre la forma en que los hombres y las mujeres perciben y se desempeñan en las preguntas de elección múltiple, aunque las razones de esto no están claras (Ben-Shakter & Sinaí, 1991). Gipps y Murphy (1994) encontraron que las niñas de 15 años preferían responder preguntas de respuesta libre, mientras que los niños de 15 años preferían responder preguntas de opción múltiple. Ben-Shakter y

Sinai (1991) encontraron que los hombres superaban a las mujeres en las preguntas de opción múltiple tanto para el noveno grado como para los grupos de edad de ingreso a la universidad.

Ben-Shakter y Sinaí no creían que las diferencias en los patrones de adivinación entre hombres y mujeres fueran las únicas responsables de las diferencias observadas en el rendimiento en las preguntas de elección múltiple. Esto difiere del hallazgo de Richardson y O'Shea (2013), que encontraron que los hombres y las mujeres eran tan propensos como el uno al otro a adivinar una pregunta correcta cuando lo intentaron, pero que los hombres eran más propensos a hacer un intento de adivinar en primer lugar. Tales discrepancias en la literatura ponen de relieve la escasa comprensión de las diferencias demográficas en las respuestas a los reactivos de la evaluación.

Por su parte, Puthiaparampil y Rahman (2020) examinaron la viabilidad de las preguntas de respuesta muy corta (PRMC) como una alternativa a las preguntas de opción múltiple (POM) en evaluaciones de escuelas de medicina. Señalaron que las POM tienen limitaciones para evaluar la competencia de los estudiantes y promover un aprendizaje significativo. En contraste, las PRMC se presentan como una opción más efectiva para evaluar el conocimiento factual y fomentar un pensamiento más profundo.

El estudio comparó el rendimiento de los estudiantes en pruebas de POM y PRMC, encontrando que las PRMC mostraron mejores resultados psicométricos y un rendimiento más equilibrado entre los estudiantes. Además, se recopilaron opiniones de estudiantes y académicos, quienes expresaron una clara preferencia por las PRMC. Se destacó que el cambio de POM a PRMC transferiría la responsabilidad de construir respuestas de los profesores a los estudiantes, lo que podría mejorar el proceso de aprendizaje.

Asimismo, se mencionó que las PRMC pueden ser calificadas de manera automática, lo que podría reducir el tiempo dedicado a la corrección manual y beneficiar tanto a los docentes como a los estudiantes. Las respuestas de PRMC pueden ser evaluadas por

cualquier persona con la ayuda de respuestas modelo, y se planteó la posibilidad de que la tecnología informática facilite aún más esta tarea en el futuro. En resumen, se concluyó que las PRMC son una alternativa viable a las pruebas de opción múltiple y son ampliamente aceptadas en las facultades de medicina.

La literatura presentada en esta sección ha sugerido que las preguntas de opción múltiple son prácticas y sencillas, pero no son idóneas para evaluar niveles avanzados de aprendizaje, como el entendimiento conceptual. Además, los alumnos pueden utilizar técnicas de eliminación o simplemente adivinar para acertar en este tipo de preguntas, sin realmente comprender el tema en cuestión. Estas preguntas, al ofrecer opciones previamente establecidas, no motivan a los estudiantes a emplear su propio razonamiento para responder.

Un formato de preguntas de respuesta libre incentivaría a los alumnos a reflexionar en profundidad y sus respuestas darían a los docentes una visión del proceso cognitivo del estudiante. No obstante, calificar estas respuestas no es sencillo y demanda mucho tiempo para quienes las evalúan. Si existiera un método para calificar automáticamente estas respuestas, podría ser una alternativa a las ampliamente utilizadas preguntas de opción múltiple. La evaluación automática del formato de pregunta de respuesta libre es el eje principal de la siguiente sección.

### **2.3 Evaluación automática de preguntas de respuesta libre**

Stödberg (2012) reportó que para el período 2004-2009, de 76 artículos, 29 se referían a la evaluación formativa y 17 a la evaluación electrónica sumativa y formativa. Su revisión reveló que la mayoría de las tareas de evaluación electrónica se construyeron sobre las preguntas cerradas (opción múltiple). Por su parte, Jordan y Mitchell (2009) proporcionan evidencia para ir más allá de las preguntas de opción múltiple y usar preguntas abiertas con la mejora de la tecnología. Ellos sugieren que las preguntas abiertas son adecuadas para una retroalimentación computarizada, "si las respuestas correctas se pueden dar en frases cortas o simples y hay una diferencia entre correcto e incorrecto, entonces las respuestas son claras" (p.382).

Pérez-Marín et al. (2009) revisaron la literatura de la época sobre la evaluación automática de preguntas de respuesta libre. Revisaron el estado del campo en ese momento, en lugar de tratar de introducir nuevas ideas y técnicas, señalaron similitudes en los enfoques desarrollados por diferentes grupos, así como discutir las formas en que diferentes autores habían intentado evaluar y validar sus métodos. Los autores indicaron como resumen que el campo se había desarrollado de manera considerable en el poco tiempo que había existido como una línea de estudio viable, pero que todavía había mucho por hacer antes de desarrollar un sistema de calificación automatizado ideal para las preguntas de respuesta libre. Sin embargo, tomaron el uso generalizado de esquemas de evaluación automatizados como un indicador de que dichos sistemas estaban en el camino correcto para el desarrollo de un sistema de calificación ideal.

Algunos de los primeros intentos de desarrollar la evaluación automatizada de preguntas con respuestas libres hicieron uso del algoritmo *Bilingual Evaluation Understudy* (BLEU), que fue desarrollado en IBM (Papineni et al., 2001). Inicialmente desarrollado como un algoritmo para permitir que las máquinas traduzcan entre diferentes idiomas, fue adoptado por otros para realizar diferentes tareas. Pérez et al. (2004a; 2004b) aplicaron el algoritmo BLEU para calificar las respuestas breves escritas por los estudiantes. Esta versión del algoritmo estuvo basada en la identificación de palabras clave para calificar, y no empleó ninguna otra construcción más allá de esto.

Noorbehbahani y Karden (2011) usaron otra versión modificada del BLEU algoritmo para calificar las preguntas de respuesta libre. Para construir las reglas de la calificación automatizada, los autores hicieron uso de respuestas libres de estudiantes reales. Su práctica se basaba en construir un banco de respuestas de referencia, y luego las respuestas se calificaban en función de la similitud a estos bancos. Dado que este era el procedimiento estándar para construir reglas efectivas, se requerían muchas respuestas por parte de los estudiantes, y estas necesitaban ser calificadas manualmente para producir el banco inicial de respuestas de referencia para entrenar. Sin embargo, si bien

la calificación automática funcionó, este paso de la calificación manual solo se necesitaba hacerse una vez.

Leacock y Choddorow (2003) desarrollaron un motor de puntuación para calificar preguntas de respuesta libre, al que llamaron C-rater. Los autores señalaron que hay muchas maneras en que un concepto puede expresarse en lenguaje natural, por lo que el motor de puntuación C-rater necesitaba ser capaz de decir cuándo se expresaba un concepto y cuándo no. Para abordar este problema, los autores consideraron que las respuestas correctas eran versiones parafraseadas de una respuesta correcta modelo; por lo tanto, diseñaron el C-rater para ser un reconocedor de paráfrasis, en lugar de ser una herramienta de reconocimiento de palabras.

Sukkarieh et al. (2003) señalaron que muchos tipos de exámenes académicos tienen preguntas que requieren respuestas libres de unas pocas palabras o frases. Estas preguntas pueden tener palabras de instrucción correspondientes a: indicar, sugerir, describir y explicar. Hay muchas preguntas de este tipo, y no requieren una gran cantidad de tensión cognitiva para evaluar, pero toman tiempo para calificar. Como resultado, los autores se dieron cuenta de que un sistema automatizado que pudiera calificar total o parcialmente tales respuestas sería útil desde el punto de vista de la eficiencia en la evaluación. Sin embargo, en ese momento, se pensaba que tal sistema sería demasiado difícil de desarrollar completamente.

Pulman y Sukkarieh (2005) continuaron con su trabajo de 2003, tratando de utilizar métodos de lingüística computacional para calificar respuestas cortas y respuestas libres. El trabajo de 2003 hizo uso de un proceso de extracción de información (IE), el cual requiere reglas de calificación que deben ser escritas a mano. Los autores señalaron que esto es problemático porque requiere que los autores de las preguntas tengan experiencia tanto en el dominio del tema y en lingüística computacional, la última de las cuales la mayoría de los profesores en las disciplinas que no son del área de computación no tendrían.

Mitchell et al. (2002) idearon un software basado en la lingüística computacional para probar y calificar automáticamente las respuestas cortas de texto libre. El software resultante *AutoMark*, desarrollado por *Intelligent Assessment Technologies Ltd.* (IAT), operado para hacer coincidir las respuestas de texto libre con las plantillas de oraciones. Jordan y Mitchell (2009) hicieron uso de este software para investigar el proceso de desarrollo de reglas de calificación para preguntas de respuesta libre. Butcher y Jordan (2010) se basaron en este trabajo comparando el enfoque basado en la lingüística computacional del software IAT con dos sistemas basados en algoritmos; estos dos sistemas fueron el tipo de pregunta *OpenMark Pattern Match* (Butcher, 2008) y *Regular Expressions*. *Pattern Match* es un algoritmo de concordancia de palabras que busca palabras clave y tiene en cuenta sinónimos y faltas de ortografía, mientras que *Regular Expressions* es un algoritmo de búsqueda de cadenas; no utiliza la gramática ni la sintaxis en sus algoritmos. Se encontró que *Pattern Match* y *Regular Expressions* eran capaces de funcionar a un nivel similar al del software basado en la lingüística computacional de IAT.

Klein et al. (2011) intentaron desarrollar un sistema de calificación automatizado basado en los principios de Análisis Semántico Latente (LSA) y agrupando datos mediante técnicas de clusterización. Los autores querían que el sistema se utilizara en un esquema de evaluación sumativa, por lo que uno de sus principales objetivos era ser lo más preciso posible. El sistema se basa en un algoritmo que está diseñado para calificar respuestas de texto libre que son mucho más cortas que las respuestas de la extensión de un ensayo y, como resultado, se centró en el contenido en lugar de la estructura. El sistema necesitaba ser configurado variando los valores de los parámetros que controlan la calificación automatizada para dar resultados de calificación más precisos. No había ninguna forma viable de automatizar este proceso, por lo que se llevó a cabo de forma manual y requirió mucho tiempo. Con la configuración correcta, el sistema fue capaz de calificar consistentemente cuando se compara con las calificaciones correspondientes dadas por evaluadores humanos.

Zehner et al. (2016) trataron de desarrollar un esquema de calificación automatizado utilizando la clusterización y técnicas de aprendizaje automático. Señalaron que era importante que al usar cualquier esquema de aprendizaje automático se tratara de evitar entrenar al algoritmo con un conjunto de datos demasiado particular, ya que esto puede conducir a sobreajuste por el cual el algoritmo puede operar bien en el conjunto de datos a partir del cual se entrenó, pero no puede operar de manera efectiva en otros conjuntos de datos. Los autores proponen que una vez que se desarrolle el esquema de calificación automatizada, puede integrarse en otros esquemas, de modo que cualquier costo que se involucre en su desarrollo se puede equilibrar rápidamente con los beneficios.

Hussein et al. (2019) revisaron los métodos utilizados para calificar automáticamente ensayos y preguntas de respuesta corta y encontraron que a pesar de las mejoras implementadas en los sistemas de evaluación automática de ensayos, persisten tres desafíos principales: la falta de la percepción del evaluador como persona, el riesgo de que los sistemas puedan ser engañados para otorgar una puntuación más alta o más baja de la que merece un ensayo, y la limitada capacidad para evaluar la creatividad de las ideas y proposiciones. Se destaca que muchas técnicas se han centrado en abordar los dos primeros desafíos, mientras que la evaluación de la creatividad en la escritura humana sigue siendo un desafío pendiente que requiere más investigación.

En una revisión de la literatura reciente realizada por Ramesh y Sanampudi (2022) sobre sistemas automatizados de puntuación de ensayos, se abordó la importancia de la evaluación en el sistema educativo y la necesidad de encontrar alternativas al proceso de evaluación manual debido a su complejidad y falta de fiabilidad. A medida que la tecnología avanza, se ha observado un cambio hacia la automatización de la evaluación en el ámbito educativo. Los sistemas actuales se han centrado principalmente en la evaluación de preguntas de opción múltiple, dejando un vacío en la evaluación efectiva de ensayos y respuestas cortas.

Los desafíos en este campo incluyen la necesidad de evaluar no solo la gramática y la estructura del texto, sino también la relevancia del contenido, el desarrollo de ideas, la

cohesión y la coherencia del ensayo. Estos aspectos más complejos de la evaluación requieren el uso de técnicas avanzadas de inteligencia artificial y aprendizaje automático. En resumen, la revisión destaca la necesidad de avanzar en el desarrollo de sistemas automatizados de puntuación de ensayos y respuestas cortas para abordar las limitaciones de la evaluación manual y mejorar la eficiencia y fiabilidad del proceso de evaluación en el ámbito educativo.

Sintetizando, la literatura revisada ilustró que la investigación enfocada a calificar automáticamente las preguntas de respuesta libre sigue siendo en general un problema abierto, a pesar de los avances publicados.

## **2.4 Modelos de aceptación de tecnología**

Los modelos de aceptación de los sistemas de aprendizaje se basaron en investigaciones anteriores sobre la aceptación de los sistemas de información. Hay nueve modelos principales en el campo de la aceptación de las Tecnologías de la Información y la Comunicación (TIC). Estos modelos se han utilizado en numerosos estudios anteriores. Cada modelo trató de explicar los determinantes de la aceptación de las TIC y especialmente el comportamiento de uso. A continuación, se describe cada modelo y las variables propuestas como principales precursores de la aceptación de las TIC.

El primer modelo es la Teoría de la Acción Razonada (TRA) (Fishbein & Ajzen, 1975). TRA fue antepasado de los modelos de aceptación de las Tecnologías de la Información (TI). Las actitudes y las normas subjetivas son los dos constructos principales en la TRA. Basado en TRA, el Modelo de Aceptación de Tecnología (TAM) fue desarrollado para predecir la aceptación de TI mediante el uso de Utilidad Percibida y Facilidad de Uso Percibida (Davis, 1989). TAM es el modelo más popular y se ha utilizado en numerosos estudios sobre la aceptación de la tecnología.



En paralelo, el Modelo Motivacional (*Motivational Model*, MM) fue empleado por Davis, Bagozzi y Warshaw (1992) utilizando los constructos de la Motivación Extrínseca y la Motivación Intrínseca. Otro modelo importante fue la Teoría del Comportamiento Planificado (TPB) (Ajzen, 1991). TPB también se basa en TRA. TPB añadió el constructo de Control de Comportamiento Percibido en TRA. TAM y TPB se combinaron en un modelo híbrido llamado TAM y TPB combinados (*Combined TAM and TPB*, C-TAM-TPB) (Taylor y Todd, 1995).

En 2000, TAM2 fue sugerido mediante la adición del constructo de la Norma Subjetiva en el TAM original (Venkatesh & Davis, 2000). Otros modelos importantes en el campo de la aceptación de TI son la Teoría Cognitiva Social (*Social Cognitive Theory*, SCT) (Bandura, 1986; Compeau & Higgins, 1995), el Modelo de Utilización de PC (*Model of PC Utilization*, MPCU) (Thompson, Higgins, & Howell, 1991; Triandis, 1977) y la Teoría de Difusión de la Innovación (*Innovation Diffusion Theory*, IDT) (Moore & Benbasat, 1991; Rogers, 2003). Finalmente, Venkatesh, Morris, Davis, & Davis (2003) propusieron un modelo unificado, llamado la Teoría Unificada de la Aceptación y el Uso de la Tecnología (UTAUT), que surge de la necesidad de crear un referente teórico común que retomara los constructos que otras teorías y modelos habían comprobado que eran útiles en la evaluación de la aceptación de la tecnología y daban cuenta de una evolución en la explicación del fenómeno. El modelo UTAUT se ha utilizado para la investigación sobre el uso intencional y conductual de las tecnologías (Wedlock & Trahan, 2019). Así como para investigar la aceptación de las tecnologías educativas desarrolladas recientemente (Thongsri et al., 2019).

La aceptación de la tecnología puede definirse como "la voluntad del usuario de emplear la tecnología para las tareas para las que está diseñada" (Teo, 2011). Los investigadores creen que la aceptación es necesaria para adoptar tecnologías educativas (Thongsri et al., 2019). Según Kaldi et al., (2008), el término adopción se utiliza para describir el proceso en el que una organización investiga, considera y decide utilizar un nuevo sistema en la organización, mientras que la aceptación se refiere a los cambios en un individuo "actitudes, percepciones y acciones que los llevan a probar nuevas prácticas,

actividades o innovaciones que son diferentes de sus rutinas normales" (Kaldi et al., 2008, p.38). En la mayoría de los artículos que se refieren a la aceptación y adopción de la tecnología, estos dos términos se utilizan indistintamente. En este trabajo la aceptación del término será utilizada según Davis (1989) como la intención del usuario de utilizar y continuar utilizando un determinado producto de tecnología de la información.

Con el fin de proporcionar una visión general del estado actual de la investigación sobre la adopción de las TIC, Williams et al. (2009) realizaron una revisión sistemática y exhaustiva de 345 artículos que aparecieron en 19 revistas revisadas por pares diferentes durante un período de 22 años. Descubrieron que estos documentos han utilizado 51 teorías. El Modelo de Aceptación de Tecnología (TAM) (Davis, 1989) resultó ser la teoría más popular ya que se utilizó en 88 (29% del total) publicaciones, seguido por la teoría de Difusión de la Innovación (DOI) (Rogers, 2003), que fue utilizado en 49 (16,3% del total) publicaciones. El tercer modelo más popular utilizado fue la Teoría del comportamiento planificado (TPB) (Ajzen, 1991), ya que se utilizó en 17 estudios, seguido de la Teoría de la acción razonada (TRA) (Fishbein & Ajzen, 1975), que se utilizó en ocho estudios solamente. Otras 48 teorías y 182 constructos teóricos se registraron a partir de una variedad de otros estudios. En otro estudio, Korpelainen (2011) afirmó que las teorías de adopción más influyentes que se utilizaron para explicar la intención de los individuos de adoptar las TIC son: TAM, TRA, TPB y la UTAUT.

Wedlock y Trahan (2019) revisaron la historia del modelo de UTAUT desde la cual se concluye que se necesita investigación para evaluar la aceptación y adopción de la tecnología en temas específicos y entornos de aprendizaje virtual más recientes. Reportan que es importante investigar las variables independientes que influyen en las percepciones y el comportamiento posterior de los alumnos y que por lo tanto, la investigación de la aceptación es necesaria para alcanzar el objetivo de utilizar la tecnología de aprendizaje en línea, es decir, para ser eficaz y garantizar el logro óptimo del estudiante.

Adicionalmente, Imtiaz y Maarop (2014) afirmaron que la mayoría de los estudios de adopción de tecnología en el área de educación han sido sobre e-learning y muy pocos sobre e-assessment. También señalaron que el estudio más antiguo sobre la adopción de evaluaciones electrónicas se realizó en el año 2011, lo que muestra que la adopción y difusión de evaluaciones electrónicas aún no se ha estudiado lo suficiente. Además, esto revela una falta de investigación en esta área, lo que demuestra una necesidad de investigación tanto a nivel nacional como internacional.

## **2.5 Modelos de aceptación de evaluaciones electrónicas**

Varios modelos teóricos se han usado para explicar la aceptación de diversos sistemas de e-learning. En gran parte de la investigación, se eligieron la facilidad de uso percibida y la utilidad percibida para explicar la aceptación y adopción del aprendizaje electrónico. Ong et al. (2004), por ejemplo, emplearon el modelo de aceptación de la tecnología para explicar las decisiones de los estudiantes de ingeniería de aceptar el aprendizaje electrónico. También propusieron un nuevo constructo que es la credibilidad percibida, para explorar la aplicabilidad de TAM en el contexto de su estudio. Además, muchos investigadores desarrollaron modelos teóricos basados en relaciones causales para explicar la satisfacción del alumno (Terzis y Economides, 2011).

Algunos investigadores como Ong y Lai (2006) proporcionaron evidencias del efecto de la utilidad percibida sobre la intención conductual de utilizar un sistema de e-learning. Utilizando la misma línea de pensamiento, Terzis y Economides (2011) sugieren que “si la Evaluación basada en Computadora (EBC) es útil para el alumno, ayudará a aumentar su concentración y probablemente su disfrute”. Estos autores creen que la utilidad percibida tiene un efecto en la intención conductual de los estudiantes de utilizar la EBC; sin embargo, existe un vacío en la literatura académica relacionada con los modelos de adopción de evaluación electrónica. Afirmaron: "Aunque existen estudios previos sobre la aceptación de los sistemas de gestión del aprendizaje (LMS), no hubo ningún estudio previo sobre la aceptación de la EBC". Además, Imtiaz y Maarop (2014) declararon que la adopción de la evaluación electrónica es un área que los investigadores han dejado

de lado; por lo tanto, se deberían realizar más investigaciones en el campo de la evaluación electrónica.

Ante esta necesidad, se realizó una revisión en fuentes pertinentes para establecer lo que se conoce acerca de la aceptación de la evaluación electrónica y hasta qué punto se ha investigado dicho tema. Los principales hallazgos corresponden al trabajo de los autores Terzis y Economides mismos que se muestran en la tabla 1.

**Tabla 1.** *Estudios relacionados con la aceptación de las evaluaciones electrónicas*

<b>Autores</b>	<b>Nombre del estudio</b>
Terzis, V., y Economides, A. A. (2011)	The Acceptance and Use of Computer Based Assessment (CBAAM)
Terzis, V., Moridis, C. N., y Economides, A. A. (2012a)	The Effect of Emotional Feedback on Behavioral Intention to Use Computer Based Assessment
Terzis, V., Moridis, C. N., y Economides, A. A. (2012b)	How student's personality traits affect Computer Based Assessment Acceptance: Integrating BFI with CBAAM
Terzis, V., Moridis, C. N., y Economides, A. A. (2013a)	Continuance acceptance of computer based assessment through the integration of user's expectations and perceptions
Terzis, V., Moridis, C. N., y Economides, A. A. (2013b)	Measuring instant emotions based on facial expressions during computer-based assessment
Papamitsiou, Z. K., Terzis, V., y Economides, A. A. (2014)	Temporal learning analytics for computer based testing
Maqableh, M. (2015)	The Acceptance and Use of Computer Based Assessment in Higher Education
Nikou, S. A., & Economides, A. A. (2017)	Mobile-based assessment: Investigating the factors that influence behavioral intention to use

Fuente: *elaboración propia.*

Actualmente, parece existir escasa esperanza en desarrollar una teoría general que, basándose en modelos existentes, logre abarcar tanto la explicación como la predicción de la aceptación del usuario. Además, debería ofrecer métodos para asegurar que el

proceso de diseño resulte en un producto aceptado por los usuarios. Por ello, la estrategia de integrar diversas teorías mediante la creación de un inventario amplio y completo, que utilice los constructos de las teorías relacionadas con la adopción tecnológica, representa una metodología prometedora para identificar los factores más determinantes en la adopción de evaluaciones electrónicas.

El modelo CBAAM fue construido sobre modelos de aceptación previos como: Modelo de Aceptación de Tecnología (TAM), Teoría del Comportamiento Planificado (TPB), y Teoría Unificada de la Aceptación y Uso de la Tecnología (UTAUT). Las limitaciones de cada teoría se resumen a continuación:

### ***Teoría del Comportamiento Planificado (TPB)***

Las principales carencias de la Teoría del Comportamiento Planificado (TPB) se deben a su insuficiente atención a las variables emocionales, como los estados anímicos, el miedo, la ansiedad, las amenazas, y tanto las emociones positivas como negativas, las cuales son consideradas de forma superficial por el TPB (Conner & Armitage, 1998). Este modelo ha sido parcialmente sustituido por el Modelo de Aceptación Tecnológica (TAM) debido a su incapacidad para abordar aspectos como la percepción de la facilidad de uso y la utilidad. Esto indica la necesidad de integrar otros constructos al TPB para comprender adecuadamente cómo se adoptan las evaluaciones electrónicas.

### ***Modelo de Aceptación Tecnológica (TAM)***

Taylor y Todd (1995) señalan como principal deficiencia del Modelo de Aceptación Tecnológica (TAM) su capacidad limitada para explicar el modo en que el diseño y la implementación pueden impactar el uso. Bagozzi (2007) añade que los adoptantes pueden considerar diversos factores que afectan sus intenciones y decisiones, mientras que el TAM solo contempla una cantidad restringida de estos factores para influir en el comportamiento. Por tanto, se destaca la necesidad de complementar el TAM con otros constructos para una comprensión adecuada sobre la adopción de evaluaciones

electrónicas. En este contexto, se espera que una teoría integradora como la UTAUT proporcione una mayor claridad en la adopción de tecnologías de evaluación electrónica.

### ***Teoría Unificada de la Aceptación y Uso de la Tecnología (UTAUT)***

El principal desafío de la Teoría Unificada de Aceptación y Uso de la Tecnología (UTAUT) radica en su complejidad, derivada de la inclusión de un elevado número de variables independientes, alcanzando un total de 41. Bagozzi (2007, p. 245) critica esta característica, destacando que el modelo utiliza estas numerosas variables para anticipar tanto las intenciones como el comportamiento, lo que ha llevado a una situación de confusión en el estudio de la adopción de tecnologías de la información y comunicación (TIC) con dicho marco. En resumen, la utilización integral de todos los constructos de UTAUT para comprender la adopción de evaluaciones electrónicas podría complicar excesivamente su aplicación práctica.

Debido a las limitaciones inherentes a cada modelo de aceptación tecnológica, la estrategia de combinar modelos previos ha emergido como un enfoque común en este ámbito de estudio, buscando aprovechar las fortalezas de cada uno para obtener una comprensión más completa de la adopción tecnológica. Cabe destacar que, al revisar la literatura, no se encontraron instrumentos con reportes de confiabilidad y validez para medir la aceptación y la intención de uso por parte de estudiantes universitarios mexicanos de un sistema de evaluación basada en computadora (EBC). Por ello, se optó por adaptar el modelo CBAAM al contexto de este estudio.

## **2.6 Conclusiones sobre el estado del arte**

Considerando la complejidad y la profundidad de los cuatro ejes temáticos de esta revisión de literatura, las conclusiones que se presentan a continuación buscan conectar los hallazgos con la pregunta central y los objetivos de la investigación, proponiendo una visión integradora que reconoce tanto los desafíos como las oportunidades que estas intersecciones representan para la innovación educativa en física.

La revisión de los inventarios de conceptos en física subraya su valor como herramientas pedagógicas clave para evaluar y mejorar la comprensión conceptual de los estudiantes, utilizando distractores diseñados para identificar y corregir conceptos erróneos. Aunque son fundamentales para evaluar la efectividad de las estrategias de enseñanza y guiar la instrucción futura, existe un amplio margen para la innovación en el diseño de preguntas. La exploración de nuevos formatos de preguntas debe enfocarse en ampliar las metodologías de evaluación para incluir elementos interactivos, simulaciones y respuestas abiertas que capturen mejor la comprensión del estudiante. Además, se sugiere el desarrollo de herramientas que permitan evaluaciones más dinámicas y adaptativas, utilizando tecnologías avanzadas para una evaluación más personalizada. La innovación en el diseño de preguntas y el uso de nuevas tecnologías prometen no solo mejorar la precisión y eficacia de estos inventarios, sino también enriquecer la experiencia educativa, preparando a los estudiantes para los desafíos reales.

Por otra parte, la revisión de la literatura sobre el formato preguntas ha identificado la versatilidad de las preguntas de opción múltiple como un recurso valioso para abordar una amplia gama de temas. Sin embargo, estas preguntas enfrentan limitaciones al evaluar comprensiones conceptuales profundas, dado que pueden permitir a los estudiantes adivinar o emplear estrategias de eliminación sin demostrar un verdadero entendimiento. Esto subraya la necesidad de desarrollar herramientas tecnopedagógicas que puedan equilibrar la eficiencia de las preguntas de opción múltiple con la profundidad de evaluación que ofrecen las preguntas de respuesta libre.

Las evaluaciones electrónicas se han utilizado tradicionalmente para tareas que se centran en probar la adquisición de conocimientos declarativos (o saber “qué”). Tales tareas requieren que el estudiante seleccione una respuesta predeterminada basada en la memoria de hechos como, por ejemplo, los tipos de preguntas de opción múltiple y sus diferentes variaciones (Bruyn et al., 2011).

Estas preguntas han sido populares porque son rápidas de escribir y se construyen fácilmente en sistemas comunes de gestión del aprendizaje (LMS) utilizados

principalmente en centros de enseñanza superior. Sin embargo, las habilidades cognitivas donde los estudiantes tienen que aplicar sus habilidades analíticas, creativas y constructivas no se pueden evaluar a través de pruebas de opción múltiple y formas equivalentes de artículos de evaluación básica (Majchrzak & Usener, 2011).

Esta limitante asociada a las funcionalidades de evaluación de los LMS's es lo que ha motivado principalmente al campo de la inteligencia artificial a proponer soluciones más robustas para evaluar preguntas que involucren respuestas en lenguaje natural. En este punto es donde precisamente hemos encontrado un vacío en la literatura, se menciona escasamente cómo los enfoques de la inteligencia artificial (IA) pueden integrarse con los sistemas LMS y su uso en evaluación electrónica (e-assessment) en la educación superior.

La evaluación automática de preguntas de respuesta libre presenta retos significativos, principalmente en el desarrollo de algoritmos capaces de interpretar con precisión la variedad de expresiones lingüísticas que los estudiantes pueden usar para comunicar conceptos complejos. La exploración de técnicas avanzadas de procesamiento del lenguaje natural y de aprendizaje automático, ilustra el potencial para abordar estos desafíos, aunque aún queda trabajo por hacer para lograr una implementación efectiva y fiable en contextos educativos.

A pesar de que se han desarrollado diversos sistemas de evaluación electrónica en los últimos quince años, la mayoría carece de ciertas características deseables que facilitarían una evaluación más integral, abarcando habilidades además de conocimientos. Este vacío representa una oportunidad para aplicar los avances en inteligencia artificial a la evaluación educativa. Una de las contribuciones de este trabajo es precisamente la exploración de cómo integrar soluciones ya existentes en el campo de la inteligencia artificial, con un sistema LMS como por ejemplo MOODLE, que tiene la ventaja de ser una plataforma abierta con la opción de integración de módulos externos, ofreciendo un marco idóneo para tal innovación.



La implementación exitosa de estrategias de evaluación automática para preguntas de respuesta corta dentro de herramientas tecno-pedagógicas demanda no solo un dominio de las tecnologías disponibles, sino también una comprensión de los principios pedagógicos que fundamentan una evaluación efectiva. Al integrar retroalimentación automatizada y personalizada, basada en las respuestas de los estudiantes, se puede aportar un valor adicional que facilite un aprendizaje más adaptativo y personalizado.

El siguiente punto trata sobre la extensa revisión de los modelos de aceptación de tecnología y de evaluaciones electrónicas, la cual presenta una base sólida sobre la que se pueden fundamentar investigaciones enfocadas en adaptar escalas de medición para evaluar la Aceptación de la Evaluación Basada en Computadora (EBC).

La revisión destaca la utilidad de modelos consolidados como el TAM, TPB, y UTAUT para comprender los factores determinantes de la aceptación tecnológica en entornos educativos. Estos modelos subrayan la importancia de constructos como la Utilidad Percibida, la Facilidad de Uso Percibida, las Normas Subjetivas, y el Control de Comportamiento Percibido. Para adaptar una escala de medición que evalúe la aceptación de la EBC, es crucial integrar estos constructos, asegurando que la escala sea capaz de capturar la complejidad del comportamiento de aceptación de los estudiantes hacia las evaluaciones basadas en computadora. La inclusión de constructos emergentes, como la Credibilidad Percibida, podría ofrecer nuevas dimensiones de análisis, reflejando aspectos específicos de la confianza y la seguridad en las evaluaciones electrónicas.

Los estudios revisados indican que, además de los factores tradicionalmente reconocidos en los modelos de aceptación tecnológica, existen variables contextuales y específicas del dominio que pueden influir en la intención de uso de las herramientas tecno-educativas para evaluaciones automáticas. Por ejemplo, la percepción de la utilidad de estas herramientas en el contexto específico de las evaluaciones de respuestas cortas y la facilidad de integración en los procesos de aprendizaje existentes son factores críticos que podrían afectar la aceptación estudiantil.

Se debe agregar que, al estudiar la innovación tecnológica, es difícil elegir un solo modelo. Según Dillon y Morris (1996), cada uno de los modelos actuales de adopción de TIC ofrece algo para nuestra comprensión de la adopción de tecnología y la aceptación del usuario. No es que un modelo sea inequívocamente mejor que el otro, sino que en determinadas circunstancias (por ejemplo, la presencia de influencias sociales), uno puede ser más poderoso que el otro. Por lo tanto, los investigadores deben ser cautelosos al seleccionar la teoría apropiada para un contexto específico (Cheng, 2019).

En este trabajo, se ha revisado la literatura relevante sobre estudios de evaluación electrónica, en relación con los modelos teóricos de adopción de TIC en general y los modelos de adopción de evaluación electrónica, en particular. La adopción y difusión de nuevas innovaciones tecnológicas, como la evaluación basada en computadora, parece estar aún poco explorada. El estudio realizado por Terzis y Economides (2011) es uno de los pocos que se han llevado a cabo con el objetivo de construir un modelo relacionado con la adopción de la evaluación electrónica. Dicho estudio destaca los constructos que afectan la intención conductual de los estudiantes de usar una evaluación basada en computadora con preguntas de opción múltiple. Este modelo fue utilizado en la presente investigación, añadiendo la contribución de su aplicación a una evaluación que contiene formatos de preguntas de respuesta corta.

La adaptación de escalas de medición específicas para la evaluación de la EBC y el análisis detallado de los factores que afectan la intención de uso de herramientas tecnopedagógicas permitirán a los profesores y desarrolladores comprender mejor y abordar las barreras para la adopción de estas tecnologías. Esto, a su vez, facilitará la creación de entornos de aprendizaje que maximicen el potencial educativo de las evaluaciones electrónicas, promoviendo prácticas de evaluación que sean tanto pedagógicamente sólidas como tecnológicamente actuales.

En resumen, este capítulo ha revelado una serie de hallazgos clave que arrojan luz sobre el complejo panorama de la adopción tecnológica en el ámbito educativo, explorando

áreas tan diversas como el inventario de conceptos en física, la evaluación automática y los modelos de aceptación de tecnología. A continuación, se presenta una síntesis de estos hallazgos, vinculándolos directamente con los objetivos planteados, con el objetivo de establecer las bases para una investigación que aspire a llenar los vacíos detectados y contribuir significativamente al cuerpo de conocimiento existente sobre la adopción de tecnologías educativas.

1. En relación con el objetivo de identificar las técnicas y herramientas empleadas actualmente para la evaluación automática de respuestas cortas, se identificó la tecnología utilizada en el tipo de pregunta de coincidencia de patrones (*Pattern Match*) en *OpenMark* (un sistema de evaluación asistido por computadora) (Butcher & Jordan, 2010).

2. En cuanto al objetivo de implementar en una herramienta tecno-pedagógica las estrategias de evaluación automática identificadas para respuestas cortas, el sistema de evaluación fue implementado en el sistema de gestión de aprendizaje MOODLE mediante el plugin *Pattern Match*.

3. Con respecto a la adaptación de una escala de medición para evaluar la Aceptación de la Evaluación Basada en Computadora, se seleccionó el Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM, por sus siglas en inglés) de Terzis y Economides (2011).

4. Finalmente, para analizar los factores que influyen en la intención de los estudiantes de utilizar la herramienta tecno-educativa para la evaluación automática de preguntas de respuesta corta, se aplicó el modelo CBAAM y se empleó el enfoque de modelado de ecuaciones estructurales (SEM) basado en mínimos cuadrados parciales (PLS).

### CAPÍTULO 3. MARCO TEÓRICO

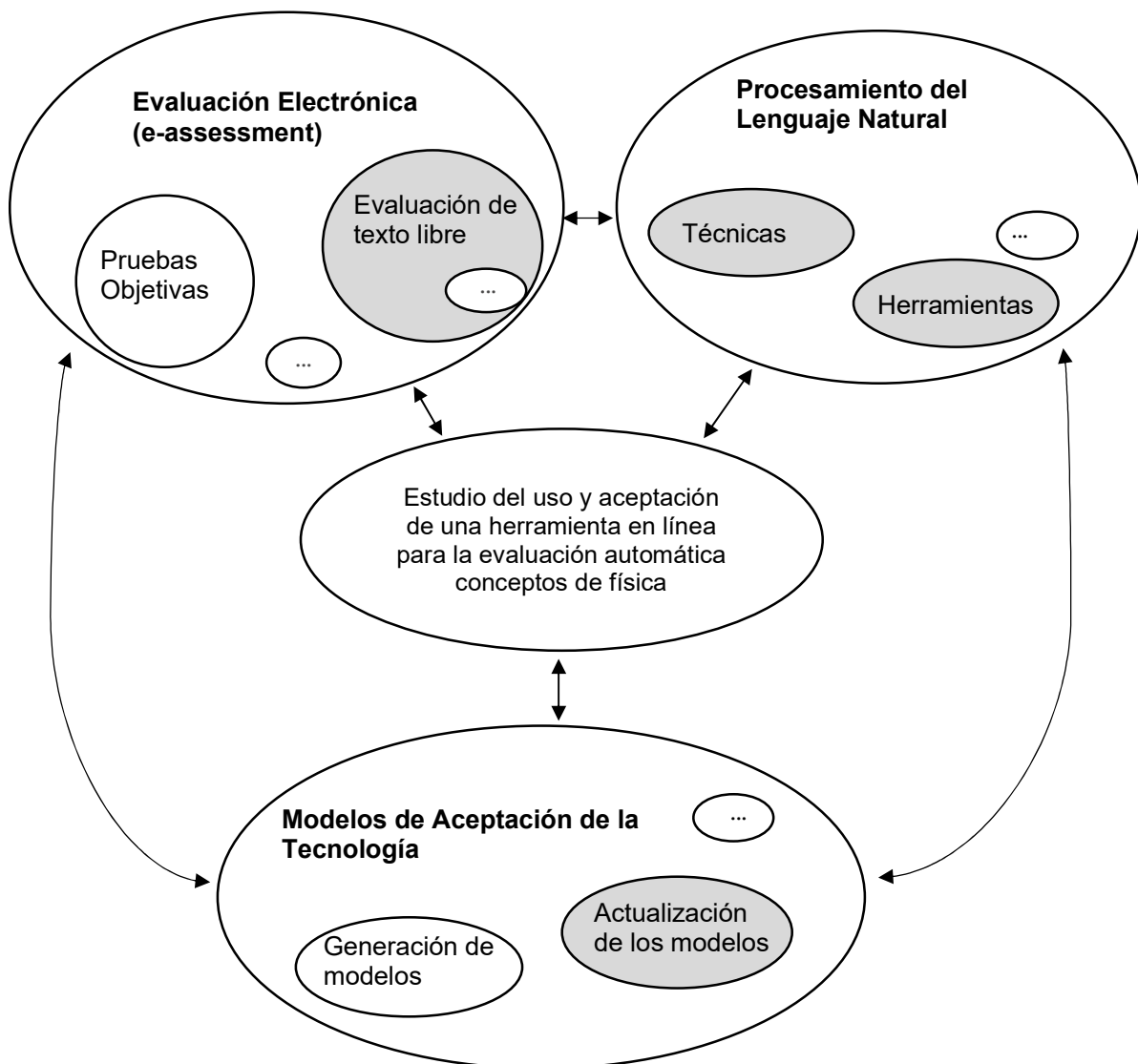
Como se ha mencionado anteriormente, un componente crucial en la evaluación basada en computadora es la estructura y formato de las preguntas. Los inventarios de conceptos en física, por ejemplo, buscan evaluar la comprensión conceptual de los estudiantes en lugar de la memorización. En este contexto, el formato de pregunta de respuesta corta emerge como una herramienta particularmente útil. Este formato no solo permite a los estudiantes expresar su comprensión en sus propias palabras, sino que también facilita una evaluación más cualitativa y profunda de su aprendizaje.

No obstante, la adopción e integración de estos sistemas en entornos educativos no siempre es inmediata. Los modelos de aceptación de tecnología son herramientas teóricas fundamentales para comprender las variables que influyen en la adopción y el uso efectivo de estas tecnologías. Estos modelos permiten identificar y abordar posibles barreras o reticencias por parte de educadores y estudiantes, garantizando así una implementación más fluida y exitosa.

En este contexto, es importante resaltar que, dado el carácter interdisciplinario de los objetivos perseguidos, este trabajo se relaciona con varios campos de estudio, como se muestra en la figura 1. En esta figura, los círculos más sombreados representan los temas abordados en esta investigación, mientras que la nomenclatura "(...)" indica la existencia de otros elementos adicionales que no se enumeran explícitamente. Estos elementos no se muestran por razones de brevedad y porque no son el enfoque principal del diagrama. Esta investigación se relaciona de manera particular con:

- **Inventarios de conceptos**, porque este campo estudia la necesidad de herramientas de diagnóstico que puedan identificar la comprensión conceptual en temas específicos.
- **Evaluación asistida por computadora o e-assessment**, porque este campo estudia cómo usar efectivamente las computadoras para evaluar el aprendizaje de los estudiantes.

- **Procesamiento del lenguaje natural**, porque este campo estudia cómo procesar automáticamente el texto libre y así, desde este campo, se aplican las técnicas para evaluar automáticamente las respuestas de texto libre.
- **Modelos de Aceptación de tecnología**, porque este campo está enfocado en una búsqueda continua para comprender los diversos factores que influyen en la aceptación individual y el uso de tecnología.



**Figura 1.** Esquema del marco teórico.

Fuente: elaboración propia.

En este capítulo se describen las principales teorías, modelos, conceptos y estrategias que dan sustento a la presente investigación y permiten una mejor comprensión de los desafíos y oportunidades presentados por la evaluación basada en computadora en el contexto educativo contemporáneo.

### **3.1 Escribir para aprender**

La investigación educativa destaca la importancia de los métodos de evaluación en el desarrollo del aprendizaje profundo y la comprensión conceptual. En particular, el uso de diferentes formatos de evaluación, como las preguntas de respuesta corta y las preguntas de opción múltiple, ha sido un tema de interés debido a sus implicaciones en el proceso de aprendizaje. La escritura, especialmente en el formato de preguntas de respuesta corta, se ha señalado como una herramienta poderosa para fomentar funciones cognitivas superiores y facilitar una comprensión más profunda del material. En contraste, las preguntas de opción múltiple son útiles para evaluar conocimientos específicos de manera rápida y eficiente, pero pueden no promover el mismo nivel de análisis y síntesis.

La literatura educativa sugiere que la escritura es una tarea en la que las funciones cognitivas superiores, como el análisis y la síntesis, pueden desarrollarse plenamente (Emig, 1977). El acto de escribir ha sido reconocido como una estrategia de aprendizaje de alto impacto en todas las disciplinas (Kuh, 2008). Se ha demostrado que la escritura es eficaz para promover el aprendizaje, la participación y el éxito de los estudiantes en cursos presenciales con matrícula relativamente alta (Reynolds et al., 2012). El proceso de escritura es importante no solo para aprender sobre algo o adquirir conocimiento, sino para construir conocimiento (Rivard, 1994).

La investigación previa (Elbow, 1994) señala que es útil distinguir entre dos objetivos diferentes para la escritura: (1) Aprender a escribir, en el que la escritura es el objetivo normal y convencional para demostrar el aprendizaje; y (2) Escribir para aprender (EPA) (también conocido como escritura para el aprendizaje), que es otro tipo importante de escritura y particularmente eficaz para promover el aprendizaje y la participación en los

materiales del curso. EPA se define como una pedagogía que involucra activamente a estudiantes de diferentes disciplinas en la construcción de su propio conocimiento a través de la escritura (Carter, 2007; Comer et al., 2014; Deane Sorcinelli & Elbow, 1997).

Como estrategia pedagógica, la EPA ha sido ampliamente adoptada para mejorar la adquisición de conocimientos y el desarrollo de habilidades cognitivas en diferentes disciplinas (Reynolds et al., 2012). Según Forsman (1985), EPA es aprender a pensar en un papel que ilustra lo que un estudiante ya sabe y cómo su conocimiento previo encaja con la nueva información que se estudia en un plan de estudios. A través de varias tareas de EPA, los estudiantes pueden pensar a través de conceptos o ideas centrales presentadas en un curso (Forsman, 1985). Como se resume en Reynolds et al. (2012), los tipos de tareas de EPA incluyen resumen, ensayo reflexivo, síntesis, documento de término, artículo corto, escrito en clase, informe de laboratorio. Estas tareas de redacción representan un componente sustancial de la educación de pregrado y posgrado (Morton, 2007), que puede reflejar lo que los estudiantes saben sobre los temas del curso y desarrollar procesos cognitivos de nivel superior que facilitan un aprendizaje significativo. Se argumenta que la importancia de la EPA debe destacarse como una práctica pedagógica significativa y fomentarse más en los cursos masivos abiertos en línea (MOOC) entre las diferentes disciplinas (Comer et al., 2014).

### **3.2 Evaluación formativa**

En educación, las evaluaciones se utilizan a menudo para obtener información sobre el aprendizaje de los estudiantes y logro. La evaluación formativa y la evaluación sumativa son dos enfoques de evaluación ampliamente aceptados. A diferencia de la evaluación sumativa (es decir, evaluación del aprendizaje) que se ocupa de resumir el estado de rendimiento de los estudiantes, la evaluación formativa (es decir, evaluación para el aprendizaje), se considera un proceso, más que una prueba, para monitorear continuamente, proporcionar retroalimentación y responder al progreso del aprendizaje de los estudiantes (Harlen & James, 1997).

La evaluación formativa ocurre durante el proceso de aprendizaje y proporciona retroalimentación tanto al profesor como a los estudiantes (Bell & Cowie, 2001). La retroalimentación a profesores y estudiantes ocurre durante el aprendizaje, no después. Los profesores necesitan conocer las concepciones y conceptos erróneos existentes en los estudiantes para ayudarlos a superarlos. Los estudiantes deben recibir retroalimentación sobre sus concepciones existentes y cómo modificar su pensamiento para que se produzca el aprendizaje. Los estudiantes pueden utilizar esta retroalimentación para determinar qué información necesitan para estudiar más a fondo y qué ajustes en su pensamiento deben realizarse.

Los propósitos de la evaluación formativa son facilitar el aprendizaje de los estudiantes e informar sobre la pedagogía (Bell & Cowie, 2001). El aprendizaje es un proceso adaptativo en el que los esquemas mentales de los estudiantes se reconstruyen basándose en retroalimentación formativa (Chi et al., 1981; Driver, 1989). El desarrollo de este esquema mental es necesario para que un estudiante aplique conceptos en contextos apropiados. A medida que los estudiantes desarrollan experiencia en la materia, se producen cambios correspondientes en la forma en que representan los problemas cognitivamente (Chi et al., 1981).

El esquema mental existente de los estudiantes y sus procesos cognitivos deben tenerse en cuenta para informar sobre la pedagogía. La evaluación formativa permite a los profesores descubrir la eficacia de las actividades de aprendizaje dentro del aula. Ejemplos de evaluación formativa incluyen preguntas con *clickers*, estudios de casos, hojas de trabajo grupales o actividades de pensar en parejas y compartir. A medida que un docente recopila retroalimentación e información de la evaluación formativa sobre el pensamiento y el aprendizaje de los estudiantes, puede hacer ajustes pedagógicos para apoyar aún más el aprendizaje. Por tanto, la evaluación formativa se encuentra en la intersección de la enseñanza y el aprendizaje.

### **3.3 Respuesta construida**



Las preguntas de respuesta construida (RC) son preguntas abiertas, como preguntas de ensayo de respuesta corta, dibujo o un examen oral, en el que los estudiantes utilizan sus propios conocimientos para construir sus respuestas en lugar de elegir entre una lista de opciones, como responder a preguntas de opción múltiple (Martinez, 1991). Aunque no existe una definición universal, las preguntas de respuesta corta se definen típicamente por el tipo de respuesta invocada, es decir, una respuesta que recuerda el conocimiento no declarado explícitamente en la pregunta, dado en el lenguaje natural, con una longitud que puede variar desde una frase corta de pocas palabras hasta varias oraciones (Burrows et al., 2015). Quizás no sea sorprendente que la investigación que estudia la efectividad y las implicaciones instructivas de las preguntas de respuesta corta tenga una larga historia, extendida a través de múltiples dominios y métodos de investigación.

Las preguntas de RC ofrecen ventajas que no ofrecen las preguntas de opción múltiple. Además de la evaluación formativa que proporciona información sobre la comprensión de los estudiantes, las preguntas de RC también identifican conceptos erróneos de los estudiantes (Birenbaum & Tatsuoka, 1987). Las preguntas de RC revelan el pensamiento de los estudiantes al permitirles usar sus diversos elementos de conocimiento para construir sus respuestas escritas (Nehm & Haertig, 2012). Las preguntas de opción múltiple tienen menos valor diagnóstico, mientras que las preguntas de RC tienen más valor diagnóstico y se pueden usar para evaluar las concepciones y conceptos erróneos de los estudiantes (Birenbaum & Tatsuoka, 1987; Martinez, 1999).

Las preguntas de opción múltiple permiten estrategias de adivinación y eliminación de respuestas, mientras que las preguntas de RC requieren una respuesta por escrito que no permite adivinar (Kuechler & Simkin, 2010; Martinez, 1991). Al usar preguntas de RC, un estudiante debe generar conocimiento y organizar información tanto cognitivamente como dentro de su respuesta escrita; por lo tanto, las preguntas de RC pueden ayudar a mejorar las habilidades de pensamiento crítico. Sin embargo, las preguntas sobre RC tienen desventajas. Las limitaciones de recursos de tiempo, dinero y experiencia limitan el uso de preguntas de RC; la retroalimentación formativa a los estudiantes sobre su

aprendizaje puede retrasarse debido a estas limitaciones (Ha et al., 2011). También hay dificultades con la fiabilidad y consistencia de la calificación humana de las preguntas de RC (Ha et al., 2011). Tales inconvenientes en el uso de preguntas de RC para proporcionar información sobre el pensamiento de los estudiantes pueden ser mejorados mediante el uso de calificación automatizada (Martinez, 1999). La evaluación automática ayuda a aliviar las limitaciones de recursos y la inconsistencia de las calificaciones humanas.

### **3.4 Evaluación escrita**

La evaluación escrita es una forma de respuesta construida. La escritura es una herramienta para mejorar las habilidades cognitivas de los estudiantes (Marzano, 1993). Cuando se anima a los estudiantes a escribir, utilizan la metacognición para reflexionar, construir y explorar sus ideas (Glynn & Muth, 1994; Keys, 1999). El aprendizaje de los estudiantes no es necesariamente lineal, sino que implica la exploración de ideas. Por ejemplo, escribir un diario le permite al estudiante explorar su pensamiento sin temor a ser calificado. La redacción informal de un diario anima a los estudiantes a reflexionar sobre sus comprensiones y malentendidos mientras comunican sus ideas (Newell, 2006).

A medida que los estudiantes escriben, pueden organizar hechos distintos en una forma más coherente. Este proceso no significa que el acto de escribir conduzca inevitablemente a una mejor comprensión de la información. Sin embargo, si un estudiante tiene conocimiento previo de un concepto, el acto de escribir y organizar sus ideas ayudará en el proceso de aprendizaje (Newell, 2006). Si un estudiante está desarrollando la comprensión conceptual, una escritura informal puede ayudarle a aclarar ideas (Glynn & Muth, 1994).

La escritura informal y no estructurada permite además a los estudiantes explorar su pensamiento y construir conocimientos (Glynn & Muth, 1994; Keys, 1999; Newell, 2006). Además, hacer que los estudiantes escriban de manera informal mejorará sus habilidades de escritura. Tal dominio de la escritura es importante para la calificación asistida por computadora, ya que los estudiantes con habilidades deficientes de escritura

tienen dificultades para expresar con precisión sus ideas, lo que puede impedir que la calificación asistida por computadora (análisis léxico) reconozca sus conocimientos (Nehm & Haertig, 2012).

### **3.5 Comprensión conceptual de la física**

La comprensión conceptual se refiere a la capacidad de aplicar el conocimiento en una variedad de contextos. La comprensión conceptual requiere que los estudiantes sean conscientes de sus concepciones e ideas erróneas existentes y de cómo realizar modificaciones en su pensamiento, a través de retroalimentación formativa. A menudo, los estudiantes recurren a la memorización de hechos en lugar de a la comprensión conceptual (Michael, 2007). La comprensión conceptual de la física ha sido de interés durante al menos cuatro décadas. El trabajo de Johnstone y Mughol en 1976 y el trabajo de muchos otros autores sirven como punto de partida para investigaciones posteriores sobre la comprensión conceptual de la física. Es relevante señalar que los temas de la mecánica newtoniana (incluyendo el movimiento uniforme, la conservación del momento y las colisiones elásticas e inelásticas) han sido reconocidos como dificultades conceptuales.

Clement (1982) estaba interesado en la comprensión conceptual de los estudiantes sobre la mecánica newtoniana y también en los conceptos erróneos que tenían sobre el tema. En un estudio cualitativo, Clement grabó en vídeo entrevistas con estudiantes en las que resolvieron una variedad de problemas de mecánica newtoniana en un entorno de pensamiento en voz alta. Clement encontró que los estudiantes frecuentemente tenían conceptos erróneos estables asociados con el malentendido común de la relación entre fuerza y aceleración, y que este concepto erróneo surge de la experiencia del día a día. El trabajo de Clement ilustra el uso temprano de un enfoque cualitativo para la investigación de la comprensión de los estudiantes en física, y dichos enfoques todavía se utilizan en la actualidad.

Otro estudio de alrededor de este tiempo fue realizado por White (1983). White planteó problemas de fuerza y movimiento a 40 estudiantes de ciencias de secundaria. Al

analizar las respuestas, White razonó que la mayoría de los estudiantes sólo tenían una comprensión parcial de las ideas newtonianas de fuerza y movimiento. En particular, los estudiantes a menudo no tomaban en cuenta el estado inicial de movimiento de un objeto al responder preguntas sobre su movimiento posterior, y con frecuencia no lograban determinar el efecto que una fuerza externa tendría sobre la velocidad de un objeto. Esta idea está respaldada por los hallazgos de McCloskey (1983), quien también encontró que los estudiantes de todos los niveles hacen uso de la intuición y el pensamiento aristotélico al resolver problemas sobre fuerza y movimiento. Tanto White como McCloskey destacan la prevalencia de estudiantes que utilizan ideas incorrectas al responder problemas de mecánica newtoniana.

### **3.6 El inventario de conceptos sobre fuerza**

Prather (1985) señaló la necesidad de una herramienta de diagnóstico que pudiera identificar la comprensión conceptual de los temas de la física, y esto fue retomado por Halloun y Hestenes (1985) cuando desarrollaron la Prueba de Diagnóstico en Mecánica (*Mechanics Diagnostics Test*, MDT). La MDT fue diseñada para probar las ideas de los estudiantes sobre fenómenos físicos comunes, y evaluó los conceptos cinemáticos de posición, distancia, movimiento, tiempo, velocidad y aceleración; y los conceptos dinámicos de inercia, fuerza, resistencia, vacío y gravedad. Hestenes, Wells y Swackhamer (1992) desarrollaron aún más la MDT en el Inventario de Conceptos de fuerza (*Force Concept Inventory*, FCI), con la idea de que el FCI proporcione un esquema más sistemático y completo de los diversos conceptos erróneos de mecánica newtoniana de los estudiantes.

El FCI evalúa seis nuevos conceptos de mecánica Newtoniana; cinemática, primera ley de Newton, segunda ley de Newton, Tercera ley de Newton, principio de superposición y tipos de fuerzas. La versión final del FCI (Halloun et al., 1995) comprende 30 preguntas de opción múltiple, cada una diseñada para tener un contenido matemático mínimo.

Aunque los problemas con la comprensión conceptual de los estudiantes de la mecánica newtoniana estaban previamente bien documentados, no había manera de investigar estos sin hacer uso de enfoques cualitativos que consumen mucho tiempo, como las

entrevistas. El FCI fue innovador en que proporcionó una forma de aprender sobre los conceptos erróneos de los estudiantes a través de la finalización de una prueba corta. Sin embargo, el FCI fue recibido con una mezcla de elogios y críticas cuando fue lanzado. Una crítica al FCI fue que no medía un solo constructo (Huffman y Heller, 1995). Usando el análisis de factorial, Huffman y Heller discutieron que el FCI era una prueba de dominio de varias situaciones relacionadas con la fuerza, pero no era una prueba del concepto de la fuerza en sí.

Hestenes y Halloun replicaron que el análisis factorial puede no ser apropiado para su aplicación al FCI porque requiere que la puntuación total se desglose en sub-puntuaciones menos significativas. Por lo tanto, argumentaron que la puntuación total en FCI realmente mide la comprensión conceptual de la mecánica newtoniana (Hestenes y Halloun, 1995). El debate entre las dos partes sobre lo que el FCI realmente mide sigue sin resolverse.

Las preocupaciones sobre la efectividad de los distractores del FCI también se han planteado como un posible problema. Con este fin, Rebello y Zollman (2004) investigaron la idea de utilizar preguntas de respuesta libre en un inventario conceptual. Lo hicieron dando a los estudiantes versiones de respuesta libre de cuatro preguntas del FCI. Luego compararon las respuestas de respuesta libre con los correspondientes distractores de opción múltiple, y miraron para ver si los distractores coincidían con las respuestas incorrectas dadas por los estudiantes en las versiones de respuesta libre.

Rebello y Zollman encontraron que los estudiantes a menudo daban respuestas que correspondía a las opciones de opción múltiple dadas en el FCI, pero que también había casos donde los estudiantes dieron respuestas que no correspondían a ninguna de las opciones, indicando diferentes procesos de pensamiento que se utilizan para responder a las preguntas. Rebello y Zollman usaron estas respuestas para crear nuevos distractores para las cuatro preguntas de la FCI, y se encontró que los nuevos distractores eran más efectivos que los anteriores cuando se usaban con otros estudiantes.

El apoyo al FCI provino de Hake (1998), quien argumentó que el uso del FCI junto con los métodos de enseñanza de participación interactiva podrían usarse para facilitar la

instrucción efectiva del material de mecánica newtoniana. Los métodos de participación interactiva difieren de los métodos de enseñanza tradicionales en que hacen uso de una gran cantidad de participación de los estudiantes, aspectos prácticos y retroalimentación.

Un ejemplo de tal método es el método de instrucción entre pares de Mazur (1997), donde se anima a los estudiantes a pensar de forma independiente sobre una pregunta clave y llegar a su propia respuesta, antes de discutirlo con sus compañeros y luego con el instructor. Tales enfoques llevaron a una gran reforma a gran escala en la enseñanza de la física (Reed-Rhoads & Imbrie, 2008), con más de un enfoque en el desarrollo y uso de nuevos métodos de enseñanza de participación interactiva. Estos nuevos métodos podrían, en teoría, evaluarse en cuanto a su efectividad mediante el uso del FCI. Como resultado, el propio FCI se le atribuye parcialmente gran parte de la reforma que ha tenido lugar en la enseñanza de la física en las últimas dos décadas (Scott et al., 2012).

Para 2011, Lasry et al. (2011) estimaron que el FCI había sido realizado más de 100.000 veces por los estudiantes, y todavía se utiliza para probar la eficacia de los métodos de enseñanza (Ding & Caballero, 2014). En general, la comunidad está de acuerdo en que el FCI es capaz de medir algo, pero todavía hay desacuerdos sobre qué puede ser ese algo (Wallace & Bailey, 2010). A pesar de tales preocupaciones, el FCI se usa ampliamente en una variedad de instituciones, y es innegable que el FCI es útil para recopilar grandes cantidades de datos de investigación en enseñanza de la física. Este instrumento cuenta con más de 50 publicaciones que respaldan su validez y fiabilidad.

El FCI sigue siendo una herramienta vigente y ampliamente utilizada en la educación en física, ya que más allá de evaluar solo la comprensión conceptual, promueve la importancia de las estructuras de conocimiento, actitudes y habilidades de resolución de problemas en los estudiantes (Stoen et al., 2020).

Fundamentado en las premisas previamente establecidas, las preguntas que utilizaremos en la presente investigación estarán basadas en el FCI, con la variante que serán evaluadas en el formato de respuesta corta de texto libre y con evaluación automática. Esta adaptación busca capitalizar las ventajas de dicho formato, alentando a los estudiantes a expresar con sus propias palabras los principios físicos en juego y a

exponer sus procesos de pensamiento de manera más detallada, lo cual es fundamental para la evaluación cualitativa de la comprensión conceptual.

### **3.7 Evaluación asistida por computadora o e-assessment**

La evaluación electrónica (e-assessment), según su definición más amplia (JISC, 2007) incluye todo uso de una computadora como parte de cualquier actividad relacionada con la evaluación, ya sea sumativa, formativa o diagnóstica. Así que su alcance incluye la presentación en línea de una tarea para la evaluación por un humano, la evaluación de un portafolio electrónico, retroalimentación entregada por archivos de audio grabados en la computadora y, más comúnmente, los exámenes en línea calificados por computadora.

Jamil et al. (2012), Al-Qdah y Ababneh (2017), Jordan (2013), Bukie (2014), Timmis et al. (2016), mencionan que la evaluación basada en computadora, las pruebas basadas en computadora, la evaluación asistida por computadora, las pruebas asistidas por computadora, las pruebas administradas por computadora, evaluación mejorada por la tecnología, evaluación habilitada por la tecnología, evaluación informatizada, pruebas informatizadas, evaluación basada en la web, examen electrónico, pruebas electrónicas y evaluación en línea, son el conjunto de términos que generalmente se consideran sinónimos de evaluación electrónica dentro de la literatura.

### **3.8 Sistemas de evaluación electrónica**

Los sistemas de evaluación electrónica pueden tener una variedad de formas. Se pueden integrar en un Sistema de Gestión del Aprendizaje (LMS), o puede ser un sistema independiente. No importa el tipo de sistemas que se elija, lo conveniente es que los profesores utilicen criterios definidos para seleccionar el sistema que se va a implementar para la evaluación electrónica. Isaias, Miranda y Pífano, (2017) propusieron un marco compuesto por ocho criterios principales para orientar el análisis y la comparación de estos sistemas: variedad de opciones de diseño, escalabilidad, seguridad, acceso y usabilidad, características de retroalimentación, personalización, costo e interoperabilidad.

Los LMS's, a pesar de no estar diseñados para exámenes electrónicos, tienen funcionalidades que se pueden utilizar para la evaluación (Kuikka, Kitola & Laakso, 2014). El uso de un entorno virtual de aprendizaje (VLE) para la evaluación presenta muchos beneficios como: estar respaldado y desarrollado en servidores seguros y tanto los profesores como los alumnos se sienten cómodos con su uso. Por otro lado, existen preocupaciones en cuanto a su confiabilidad y robustez, ya que se accede a ellos simultáneamente desde diferentes lugares y por un gran número de alumnos (JISC, 2006).

### **Otros sistemas de evaluación electrónica**

Los sistemas que se han desarrollado específicamente para la evaluación electrónica, a pesar de ser más complejos y que requieren más personal y esfuerzos de configuración, son robustos y algunos se enfocan exclusivamente en la evaluación automática de contenido escrito, el cual se puede clasificar en tres grandes categorías: lenguaje natural (p. ej. respuestas cortas y ensayos), texto estructurado (p. ej. matemáticas y códigos fuente) y discurso (p. ej. diálogos). Este tipo de sistemas de evaluación son los que se investigan desde la perspectiva de la inteligencia artificial aplicando herramientas de procesamiento del lenguaje natural y a este campo en la literatura se le conoce como: evaluación automática de respuesta corta (*Automatic Short Answer Grading, ASAG*). (Burrows et al., 2015).

La investigación en la evaluación de respuestas de lenguaje natural con métodos computacionales tiene una historia que se remonta a los primeros trabajos de Page (1966). Desde entonces, la evaluación automática de las respuestas del lenguaje natural se ha convertido en un campo de estudio muy amplio. Además, las técnicas se han ramificado dependiendo del tipo de pregunta, como respuestas cortas versus ensayos. La tecnología de interés todavía está sujeta a cuestiones de investigación abiertas. De hecho, algunas de las ventajas mencionadas no vienen sin problemas. Por ejemplo, el trabajo necesario para crear una solución automatizada a menudo requiere mucho tiempo de desarrollo (Williamson et al. 2012).

Burrows et al. (2015) realizaron un análisis histórico a través de investigaciones en ASAG, dónde revisaron 35 sistemas ASAG e indicaron cómo la investigación se ha



desarrollado desde 1996. Determinaron cinco temas temporales a lo largo de las investigaciones, que son la era del mapeo de conceptos, la era de la extracción de información, la era de los métodos basados en el corpus, la era del aprendizaje automático y la era de evaluación. Las cuatro primeras “eras” comprenden metodologías propias de la inteligencia artificial, la era de la evaluación no está relacionada con un método: utilizan corpus compartidos, también concursos y foros de evaluación entre grupos de investigación de todo el mundo sobre un problema particular por dinero o prestigio. Para el contexto de este trabajo la tecnología empleada está basada en un algoritmo de coincidencia de palabras que busca palabras clave y tiene en cuenta sinónimos y errores ortográficos, sin embargo, no hace uso de la gramática o la sintaxis en sus algoritmos.

### **3.9 Procesamiento del Lenguaje Natural**

El procesamiento de lenguaje natural (PLN) constituye un área extensa y multidisciplinar, que abarca un número casi ilimitado de oraciones y expresiones lingüísticas. Este campo emergió en la década de 1950 como la intersección de la inteligencia artificial y la lingüística. Uno de los retos inherentes al PLN es la inherente ambigüedad del lenguaje humano: muchas palabras presentan múltiples acepciones y las frases pueden adquirir distintos significados según el contexto en el que se enmarquen. Estas complejidades lingüísticas representan un desafío significativo en el desarrollo de software capaz de interpretar y procesar el lenguaje de manera natural y efectiva (Shahbaz et al., 2012).

Las etapas del PNL son (Jurafsky & Martin, 2008):

1. **Morfología:** La morfología se refiere a la forma en que las palabras se construyen a partir de unidades de significados más pequeños.
2. **Sintaxis:** La sintaxis se refiere a cómo se unen las palabras para formar oraciones correctas y qué papel estructural tiene cada palabra.
3. **Semántica:** Semántica se refiere a lo que las palabras significan y cómo estos significados se combinan en oraciones para formar significados de oración.

4. **Pragmática:** La pragmática se refiere a cómo se utilizan las oraciones en diferentes situaciones y cómo el uso afecta a la interpretación de la oración.

5. **Discurso:** El discurso se refiere a cómo las oraciones inmediatamente anteriores afectan la interpretación de la siguiente oración.

### **3.10 Expresiones regulares**

Las expresiones regulares (ER) son una de las herramientas más útiles en informática. ER es un lenguaje formal para especificar la cadena y comúnmente se denomina expresión de búsqueda. EL PNL, como área de la informática, se ha beneficiado enormemente de las expresiones regulares: se utilizan en fonología, morfología, análisis de texto, extracción de información y reconocimiento de voz. Las expresiones regulares se colocan dentro del par de coincidencias. Una expresión regular, o ER, describe cadenas de caracteres (palabras o frases o cualquier texto arbitrario). Es un patrón que coincide con ciertas cadenas y no con otras. Una expresión regular es un conjunto de caracteres que especifican un patrón (Kaur, 2014). Las expresiones regulares son simplemente un conjunto predefinido de reglas de coincidencia basadas en texto que se pueden utilizar para identificar palabras clave o frases estructuradas en la respuesta de un estudiante

El objetivo del trabajo presentado aquí no fue avanzar en el estado del arte en la evaluación automática de respuestas cortas. Más bien, investigar un instrumento educativo como herramienta de evaluación con tecnología mejorada que implica la evaluación del lenguaje natural en las respuestas de los estudiantes y la técnica utilizada en esta implementación fue la de expresiones regulares.

### **3.11 La aceptación de tecnología**

El advenimiento de Internet y los teléfonos inteligentes en el siglo XXI ha hecho que las tecnologías de la información parte indispensable de la vida humana, que estaban disponibles en su mayoría solo para los usuarios de la organización durante finales del siglo XX. La investigación sobre adopción y difusión de tecnología es una corriente madura de exploración dentro de la literatura contemporánea de sistemas de información (SI) y los investigadores de SI están en una búsqueda continua para comprender los

diversos factores que influyen en la aceptación individual y el uso de tecnología de la información emergente (TI) (Hughes et al., 2016, 2017, 2020).

Esta corriente de investigación generalizada ha sido testigo de una variedad de metodologías de investigación que examinan multitud de tecnologías en una variedad de países con la literatura existente que revela numerosas teorías, contextos, unidades de análisis y métodos de investigación (Dwivedi & Williams, 2008; Choudrie & Dwivedi, 2005; Williams et al., 2009). Los diversos contextos de investigación basados en la tecnología, tipo de usuario, ubicación, tiempo de adopción y la tarea realizada dio lugar a muchas teorías que compiten y modelos.

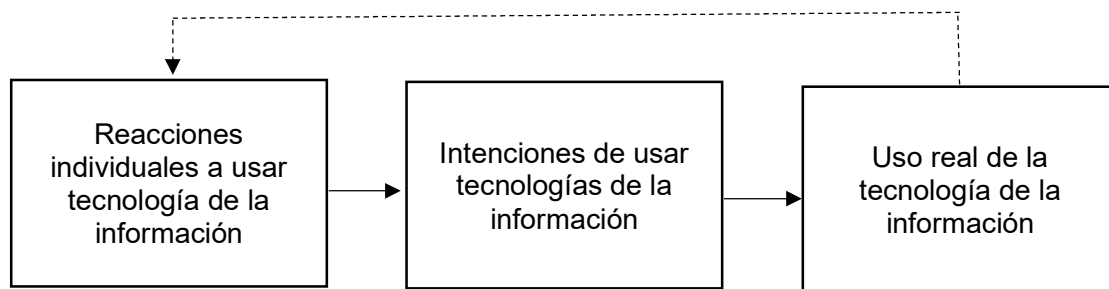
Desde su introducción, el modelo de la Teoría Unificada de Aceptación y Uso de la Tecnología (UTAUT) ha sido ampliamente validado y utilizado como un lente teórico para la investigación de adopción y difusión, observando la intención del usuario y el comportamiento dentro de múltiples contextos. En la actualidad, la inducción de innovaciones tecnológicas ha impulsado a que los investigadores se concentren en examinar los factores y las tasas de adopción y difusión, trayendo esta área de la academia a buen término (Venkatesh, Morris, Davis & Davis, 2003; Williams, Rana & Dwivedi, 2015).

Posteriormente, la existencia de varios modelos tecnológicos llevó a Venkatesh et al. (2003) a unir múltiples teorías en un modelo general para explicar la adopción y el uso de la tecnología. En el campo educativo, estas teorías y modelos se han aplicado a los formatos tradicionales de aprendizaje en línea y a distancia. Actualmente, el aprendizaje mixto o híbrido, junto con los MOOC's o los sistemas de entrega de cursos masivos abiertos en línea, han brindado nuevas oportunidades para examinar empíricamente la efectividad de las modalidades de aprendizaje asistido por tecnología.

El modelo UTAUT fue desarrollado originalmente a través de la combinación de ocho teorías dominantes de tecnología para formar un modelo universalmente aceptado para el uso de la tecnología: la Teoría de la Acción Razonada (*Theory of Reasoned Action*, TRA), el Modelo de Aceptación de Tecnología (TAM), el Modelo Motivacional, la Teoría del Comportamiento Planificado (TPB), un combinación del TBP/TAM, el Modelo de

Utilización de PC (MPCU), la Teoría de la Innovación de la Difusión (IDT) y Teoría Social Cognitiva (SCT) (Williams et al., 2015).

En general, estas teorías y modelos comparten el supuesto de que el ser humano es un ser racional cuyos actos están influenciados por un bagaje de experiencia acumulada, de tal forma que las reacciones individuales sobre el uso de la tecnología (establecidas por factores de diversa índole, como percepciones, actitudes y el contexto social del individuo) determinan la intención que el sujeto tiene de utilizarla y, a su vez, esta intención produce el uso real de la tecnología, es decir, se emite la conducta de uso (figura 2). Esta premisa básica surge de la teoría de la acción razonada (Fishbein & Ajzen, 1975) y se convierte en la base de la mayoría de los modelos y teorías al respecto.



**Figura 2.** Concepto básico subyacente a los modelos de aceptación y uso de la tecnología.

Fuente: tomada y traducida de Venkatesh, Morris y Davis et al., 2003.

La teoría unificada de la aceptación y uso de la tecnología surge de la necesidad de crear un referente teórico común que retomara los constructos que otras teorías y modelos habían comprobado que eran útiles en la evaluación de la aceptación de la tecnología y daban cuenta de una evolución en la explicación del fenómeno. En la Tabla 2 se muestran las ocho teorías y modelos que han abordado el tema a través del tiempo, así como los constructos propuestos por cada uno de ellos.

**Tabla 2.** Línea del tiempo de los modelos y teorías incluidos en la UTAUT

Teorías y modelos	Constructos
Teorías de la difusión de la innovación 1962, Rogers	<ul style="list-style-type: none"> <li>• Ventaja relativa</li> <li>• Facilidad de uso</li> </ul>

	<ul style="list-style-type: none"> <li>• Imagen</li> <li>• Visibilidad</li> <li>• Compatibilidad</li> <li>• Demostrabilidad de los resultados</li> <li>• Uso voluntario</li> </ul>
Teorías de la acción razonada (TRA) 1967, Fishbein, Ajzen	<ul style="list-style-type: none"> <li>• Norma subjetiva</li> <li>• Actitud hacia</li> <li>• La conducta</li> </ul>
Teorías de la conducta planeada (TPB), 1985, Ajzen	<ul style="list-style-type: none"> <li>• Actitud hacia la conducta</li> <li>• Norma subjetiva</li> <li>• Control de la conducta percibida</li> </ul>
Modelo de la aceptación de la tecnología (TAM), 1986, Davis	<ul style="list-style-type: none"> <li>• Utilidad percibida</li> <li>• Facilidad de uso percibida</li> <li>• Norma subjetiva</li> </ul>
Teoría social cognitiva (SCT), 1986, Bandura	<ul style="list-style-type: none"> <li>• Expectativa de resultados del desempeño</li> <li>• Expectativa de resultados personales</li> <li>• Autoeficiencia</li> <li>• Afecto</li> <li>• Ansiedad</li> </ul>
Modelo de utilización de la PC (MPCU), 1991, Thompson	<ul style="list-style-type: none"> <li>• Trabajo</li> <li>• Complejidad</li> <li>• Consecuencias a largo plazo</li> <li>• Afecto hacia el uso</li> <li>• Factores sociales</li> <li>• Condiciones de facilidad</li> </ul>
Combinación TAM y TPB, 1995, Taylor y Todd	<ul style="list-style-type: none"> <li>• Actitud hacia la conducta</li> <li>• Norma subjetiva</li> <li>• Percepción de control sobre la conducta</li> <li>• Utilidad Percibida</li> </ul>
Modelo motivacional (MM), 1997, Vallerand	<ul style="list-style-type: none"> <li>• Motivación extrínseca</li> <li>• Motivación intrínseca</li> </ul>
Teoría unificada de la aceptación y uso de la tecnología (UTAUT), 2003, Venkatesh	<ul style="list-style-type: none"> <li>• Expectativa del funcionamiento</li> <li>• Expectativa del esfuerzo</li> <li>• Influencia social</li> <li>• Condiciones de facilidad</li> </ul>

Fuente: *Venkatesh, Morris & Davis et al., 2003.*

Los autores de esta teoría estudiaron todos los constructos y formularon un modelo unificado que los integró (figura 3). De la totalidad de los constructos, se identificaron cuatro factores principales: expectativa del funcionamiento, expectativa del esfuerzo, influencia social y condiciones de facilidad, los cuales están moderados por el género, la

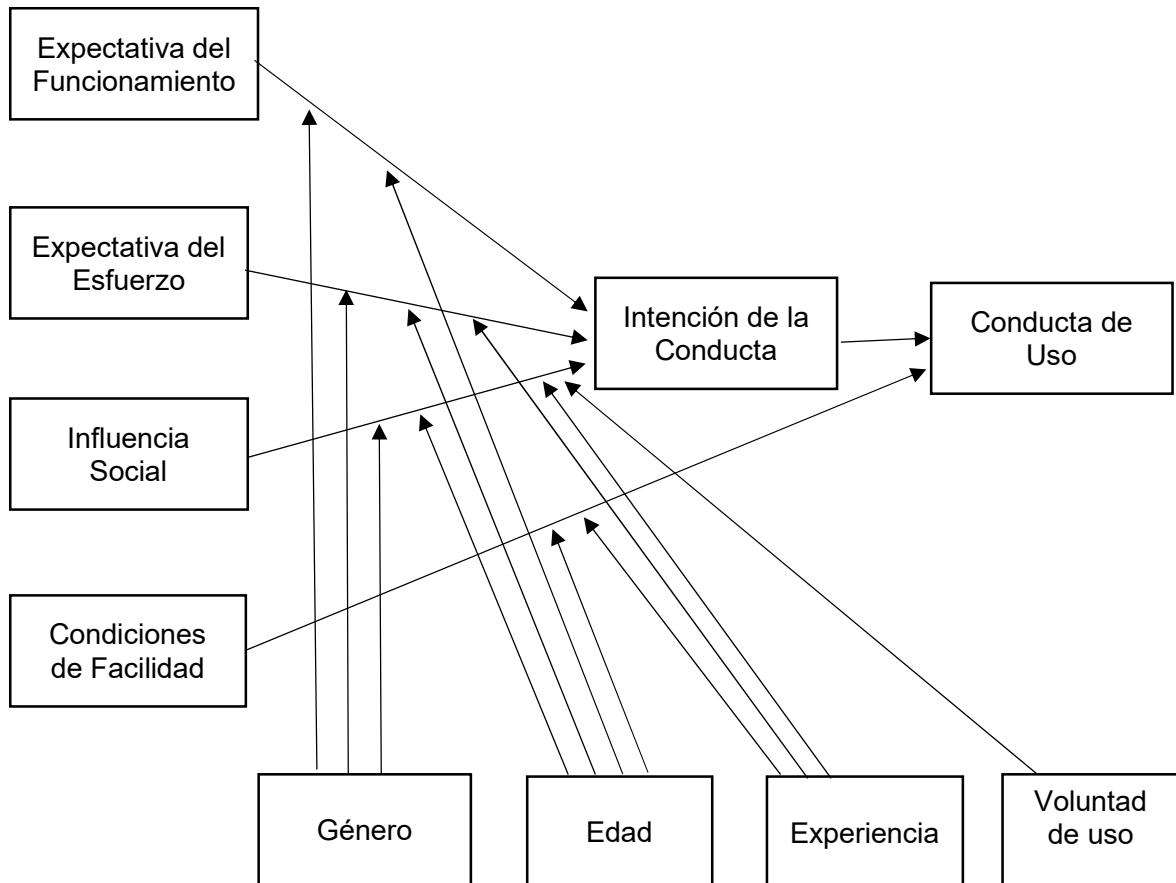
edad, la experiencia y la voluntad de uso; este último hace referencia a si el uso de la tecnología es voluntario o impuesto (Venkatesh, Morris & Davis et al., 2003).

- *Expectativa del funcionamiento*: el grado en que una persona cree que el uso del sistema le ayudará a obtener ganancias en su desempeño. Este factor se encuentra moderado por el género y la edad.
- *Expectativa del esfuerzo*: el grado de facilidad asociado al uso del sistema. Este factor se encuentra moderado por el género, la edad y la experiencia.
- *Influencia social*: el grado en que una persona percibe que otras personas importantes para él creen que debe usar el sistema. Este factor se encuentra moderado por el género, la edad, la experiencia y la voluntad de uso.
- *Condiciones de facilidad*: el grado en que una persona percibe que la estructura organizacional y la infraestructura técnica pueden dar soporte al uso del sistema. Este factor se encuentra moderado por la edad y la experiencia.

Para desarrollar y probar el modelo UTAUT, Venkatesh, et al. (2003) llevó a cabo procedimientos de validación para la escala combinada, lo que dio como resultado un  $R^2$  (coeficiente de determinación) ajustado general del 69 % (este porcentaje se relaciona con la variación de la intención de comportamiento). Este estudio de validación no solo estableció la relevancia del modelo UTAUT, sino también su dominio con respecto a las teorías anteriores. Un estudio desarrollado por Williams et al. (2015) encontró que el modelo UTAUT todavía es ampliamente utilizado hoy en día especialmente en las áreas de gobierno, banca, aprendizaje y comercio, que se apoyan de la tecnología.

El UTAUT sugiere que cuatro constructos principales (expectativa de desempeño, la expectativa de esfuerzo, la influencia social y las condiciones facilitadoras) son determinantes directos de la intención conductual y, en última instancia, del comportamiento, y que estos constructos son a su vez moderados por género, edad, experiencia y voluntariedad de uso (Venkatesh et al., 2003). Se argumenta que al examinar la presencia de cada uno de estos constructos en un contexto del “mundo real”, los investigadores y los profesionales podrán evaluar la intención de utilizar un sistema

específico, lo que permite la identificación de la clave influencias en la aceptación en cualquier contexto dado.



**Figura 3. UTAUT. Fuente: Venkatesh, Morris y Davis et al., 2003.**

### 3.12 Contexto de la investigación

Una vez identificadas las contribuciones importantes a la literatura sobre la aceptación de la tecnología de la información por parte del usuario, es fundamental destacar que se realizó una comparación y síntesis de los modelos teóricos incluidos en esta revisión. Este análisis exhaustivo proporciona una base sólida para justificar la elección y adaptación de un instrumento de recolección de datos que sea fiable y válido, asegurando que esté debidamente fundamentado en la teoría existente.

Para el contexto de esta investigación, el instrumento que hemos identificado y que será la base para adaptarlo a nuestros objetivos, es el Modelo de Aceptación de Evaluación Basada en Computadora (Computer Based Assessment Acceptance Model, CBAAM) propuesto por Terzis y Economides (2011). Este instrumento tuvo como objetivo ampliar y dar un paso más en la investigación de la experiencia del estudiante con prácticas formativas de evaluación electrónica y a continuación se describen los antecedentes teóricos sobre los cuales está basado.

### **3.13 Aceptación de la evaluación basada en computadora**

UTUAT explica que los determinantes centrales de la aceptación de la TI son cuatro variables y las otras cuatro variables son moderadoras de las relaciones principales. Sin embargo, los sistemas de aprendizaje basados en tecnología son utilizados por un grupo de usuarios específico. Por lo tanto, los modelos anteriores no pueden reflejar completamente la experiencia de los estudiantes, lo que requiere una búsqueda de factores de motivación intrínsecos adicionales (Ong, Lai & Wang, 2004). En los modelos anteriores se han utilizado muchas variables para explicar la aceptación y la intención de utilizar un sistema de aprendizaje basado en tecnología (Ong et al., 2004; Teo, 2009). La Utilidad Percibida y la Facilidad de Uso Percibida de TAM se han utilizado en muchos estudios relacionados con la aceptación del aprendizaje basado en tecnología (Ong et al., 2004; Yi & Hwang, 2003).

UTUAT también se ha utilizado para explicar la adopción de un sistema de aprendizaje basado en tecnología (Wang, Wu & Wang, 2009). Otros investigadores utilizaron solo algunas variables como condiciones facilitadoras o influencia social en sus modelos propuestos (Teo, 2009). Además, algunos estudios agregaron variables más relevantes con el procedimiento de aprendizaje. Wang et al. (2009) usó Alegría Percibida de Moon y Kim (2001) y Autogestión del aprendizaje, que se define como la autodisciplina y la capacidad de aprendizaje autónomo (Smith, Murphy, & Mahoney, 2003), para explicar la intención de uso. Yi y Hwang, 2003 usaron el constructo Disfrute, lo que explica que usar un sistema informático se percibe personalmente agradable por derecho propio, aparte del valor instrumental de la tecnología (Davis et al., 1992).



Además, introdujeron la Orientación a Metas de Aprendizaje, que se definió como el enfoque del individuo a una tarea para comprender algo nuevo o para mejorar su nivel de competencia, y Autoeficacia Específica de la Aplicación (Yi & Hwang, 2003). Van Raaij y Schepers (2008) agregó Innovación Personal en el dominio de TI, que se define como la voluntad de un individuo de probar cualquier tecnología de información nueva (Agarwal & Prasad, 1999). Shih (2008) contribuye a la literatura de aceptación de TI con dos variables: Expectativas de Resultados Personales, que es la expectativa de resultado estimada por un individuo con respecto a si un comportamiento en particular producirá los resultados requeridos (Bandura, 1977); y Control de Comportamiento Percibido, que son las percepciones individuales de su control sobre el sistema para el aprendizaje basado en la Web (Shih, 2008). Además, muchos investigadores desarrollaron modelos causales para explicar la satisfacción de un alumno (por ejemplo, Sun, Tsai, Dedo, Chen & Yeh, 2008; Wang, 2003).

La Tabla 3 resume los constructos y los vínculos causales relacionados que han sido desarrollados en estudios previos.

**Tabla 3.** *Constructos y vínculos causales desarrollados en estudios previos*

<b>Constructo</b>	<b>Vínculos causales relacionados</b>	<b>Evidencias de apoyo</b>
Utilidad percibida (UP)	UP → Intención de Uso	Landry, Griffeth, y Hartman, 2006; Lee, 2008; Liao y Lu, 2008; Ong et al., 2004; Ong y Lai, 2006; Padilla-Melendez et al., 2008; Teo, 2009; Van Raaij y Schepers, 2008; Yi y Hwang, 2003
Facilidad de uso percibida (FDUP)	UP → Actitud FDUP → Intención de Uso FDUP → Actitud	Ngai, Poon, y Chan, 2007; Landry et al., 2006; Lee, 2008; Ong et al., 2004; Ong y Lai, 2006; Padilla-Melendez et al., 2008; Teo,

		2009; Van Raaij y Schepers, 2008; Yi y Hwang, 2003; Ngai et al., 2007
influencia social (IS)	IS → Intención de Uso	Van Raaij y Schepers, 2008; Wang et al., 2009
Percepción lúdica (PL)	PL → Intención de Uso	Wang et al., 2009
Autogestión del aprendizaje (ADA)	ADA → Intención de Uso	Wang et al., 2009
Autoeficacia Informática (AI)	AI → Intención de Uso AI → UP, FDUP	Padilla-Melendez et al., 2008; Ong et al., 2004; Ong y Lai, 2006; Teo, 2009
Condiciones facilitadoras (CF)	CF → Actitud	Teo, 2009; Teo, Lee, y Chai, 2008
Orientación a Metas de Aprendizaje (OMA)	OMA → Autoeficacia Específica de la Aplicación	Yi y Hwang, 2003
Expectativas de resultados personales (ERP)	ERP → Intención de Uso, Actitud	Shih, 2008
Contenido (C)	C → Satisfacción	Shee y Wang, 2008; Wang, 2003

Fuente: *Terzis y Economides (2011)*.

### 3.14 Modelo de investigación CBAAM e hipótesis

El modelo CBAAM fue construido sobre modelos de aceptación previos como: Modelo de Aceptación de Tecnología (TAM), Teoría del Comportamiento Planificado (TPB), y Teoría Unificada de la Aceptación y Uso de la Tecnología (UTAUT). Se agregaron dos variables adicionales (Contenido y Expectativa de Meta) sobre las variables de medición actuales. Este modelo combinó los siguientes constructos para estudiar la aceptación de un sistema de EBC:

#### ***Percepción lúdica***

La Percepción Lúdica (PL) se introdujo por primera vez en un modelo de aceptación como una creencia sobresaliente intrínseca que se forma a partir de la experiencia subjetiva del individuo con el sistema (Moon & Kim, 2001). Moon y Kim basado en los trabajos de Csikszentmihalyi (1975) y Deci y Ryan (1985) ampliaron TAM añadiendo la Percepción Lúdica. La alegría percibida se define por tres dimensiones:

(a) Concentración: Si el usuario está concentrado en la actividad.

(b) Curiosidad: Si se despierta la curiosidad cognitiva del usuario (Malone, 1981a; 1981b).

(c) Disfrute: Si el usuario disfruta de la interacción con el sistema.

Estas tres dimensiones están vinculadas y son interdependientes, pero no siempre se observan juntas en la práctica. Por lo tanto, cada dimensión no refleja la interacción total. Estas tres dimensiones de Percepción lúdica son factores muy importantes para la implementación exitosa de una Evaluación Basada en Computadora. Una EBC debe mantener la concentración, la curiosidad y el disfrute del alumno en niveles altos. Por lo tanto, CBAAM asumió que la Intención de Uso se ve afectada positivamente por la Percepción Lúdica como en la siguiente hipótesis:

H1: La Percepción Lúdica tendrá un efecto positivo en la Intención de Utilizar la EBC.

### ***Utilidad Percibida***

La Utilidad Percibida (UP) se define como el grado en que una persona cree que usar un sistema en particular mejorará su rendimiento (Davis, 1989). Numerosos investigadores aportan evidencias del efecto de la Utilidad Percibida sobre la Intención Conductual al usar un sistema de aprendizaje (Lee, 2008; Ong & Lai, 2006; Van Raaij & Schepers, 2008). Asimismo, los alumnos pueden creer que un sistema de EBC mejorar su conocimiento, comprensión y rendimiento para el curso. Además, si la EBC es útil para el alumno, ayudará a aumentar la concentración, la curiosidad y probablemente el disfrute del alumno. CBAAM también supone que la concentración, la curiosidad y el disfrute del alumno aumentarán como resultado del uso de un sistema útil que conduce a las siguientes hipótesis:

H2: La Utilidad Percibida tendrá un efecto positivo en la Intención de Utilizar la EBC.

H3: La Utilidad Percibida tendrá un efecto positivo en la Percepción Lúdica.

### ***Facilidad de uso percibida***

La Facilidad de Uso Percibida (FDUP) se define como el grado en que una persona cree que usar el sistema estaría libre de esfuerzo (Davis, 1989). Investigaciones anteriores han demostrado que se espera que la Facilidad de Uso Percibida influya directamente en la Utilidad Percibida y en la Intención Conductual de Uso (Agarwal & Prasad, 1999; Hu, Chau, Sheng y Tam, 1999; Venkatesh, 1999; Venkatesh y Davis, 1996). La Facilidad de Uso Percibida mejorará la Percibida Lúdica, ya que la facilidad de uso proporciona un uso fluido del sistema sin perturbaciones molestas. Para los efectos anteriores de la Facilidad de Uso Percibida, se hicieron las siguientes hipótesis:

H4: La Facilidad de Uso Percibida tendrá un efecto positivo en la Intención de Utilizar la EBC.

H5: La Facilidad de Uso Percibida tendrá un efecto positivo en la Utilidad Percibida.

H6: La Facilidad de Uso Percibida tendrá un efecto positivo en la Percepción Lúdica.

### ***Autoeficacia Informática***

La Autoeficacia Informática (AI) se define como las percepciones del individuo sobre su capacidad para utilizar computadoras (Compeau & Higgins, 1995). Los resultados anteriores muestran que existe un vínculo causal entre la Autoeficacia Informática y la Facilidad de Uso Percibida (Agarwal, Sambamurthy & Satair, 2000; Padilla-Melendez, Garrido-Moreno & Del Aguila-Obra, 2008; Venkatesh & Davis, 1996). Así, la EBC tiene un importante efecto directo sobre FDUO y una indirecta sobre la Intención Conductual para usar el sistema. En una EBC, la Autoeficacia Informática afecta a los estudiantes de varias maneras. Por ejemplo, los estudiantes con mejores habilidades de TI ganan tiempo significativo al hacer clic o escribir más rápido. El tiempo es un factor muy importante en una EBC. Entonces, la variable AI debe incluirse para una mejor explicación de la aceptación de EBC. Con base en estos hallazgos, se planteó la siguiente hipótesis:

H7: La Autoeficacia Informática tendrá un efecto positivo en la Facilidad de Uso Percibida.

### ***Influencia Social***

Taylor y Todd (1995) sugieren la Influencia Social (IS) como el efecto de la opinión de otras personas, la influencia superior y la influencia de los compañeros. La Influencia Social se puede definir por tres elementos: Norma Subjetiva, Imagen y Voluntariedad (Karahanna & Straub, 1999). Modelos anteriores utilizaron los siguientes constructos para medir la influencia social: Factores sociales (MPCU), Imagen (IDT) y Norma subjetiva (TRA, TPB, C-TAM-TPB y TAM2) (Venkatesh, Morris, Davis & Davis, 2003). TAM2 propone que la Norma Subjetiva y la Imagen influirán en las percepciones de los usuarios sobre la utilidad del sistema.

TAM2 propone que la norma subjetiva y la imagen influirán en las percepciones de los usuarios sobre la utilidad del sistema. TAM2 también sugiere que la Norma Subjetiva no tiene efecto directo sobre la Intención de Comportamiento si el uso del sistema es voluntario. Finalmente, en el modelo UTUAT que resume los ocho modelos anteriores, la Influencia Social es uno de los cuatro principales determinantes de la Intención Conductual. Un gran número de estudios de aceptación de tecnología utilizaron la Influencia Social en sus modelos propuestos y encontraron resultados útiles (Agarwal & Karahanna, 2000; Karahanna & Straub, 1999; Lu, Yu, Liu, & Yao, 2003; Taylor & Todd, 1995; Venkatesh & Davis, 2000; Venkatesh et al., 2003). Del mismo modo, la Influencia Social se ha utilizado en los modelos propuestos para LMS (Wang et al., 2009). La única hipótesis sobre la influencia social es:

H8: La Influencia Social tendrá un efecto positivo en la Utilidad Percibida.

### ***Condiciones Facilitadoras***

Las Condiciones Facilitadoras (CF) son factores que influyen en la creencia de un individuo para realizar un procedimiento. Las CF tienen muchos aspectos diferentes. La definición de CF depende del sistema y del proceso o de las personas que les apoyaran. Por ejemplo, un aspecto de las CF podría ser soporte técnico o relacionado con los servicios de apoyo en línea. Otra explicación dada es que las CF tienen que ver con

factores de recursos como el tiempo y dinero (Lu, Liu, Yu, & Wang, 2008). CF también pueden ser definidas por las políticas, regulaciones y entorno legal de un sistema. Las actividades de comunicación y la participación activa del personal organizativo también podrían definirse como CF (Bueno & Salmeron, 2008).

En nuestro estudio definimos como CF principalmente el apoyo durante la EBC. La EBC debe tener herramientas para ayudar a los estudiantes cuando se encuentran con dificultades con el sistema. Además, si la EBC tiene lugar en la universidad, un experto debe asistir durante la EBC para resolver las dudas de los estudiantes. sobre el uso de EBC o incluso el contenido de las preguntas. Por las razones anteriores, se planteó la siguiente hipótesis:

H9: Las Condiciones Facilitadoras tendrán un efecto positivo en la Facilidad de Uso Percibida.

### ***Expectativa de Meta***

Estudios previos describieron la necesidad de autodirección y orientación a objetivos en el aprendizaje a distancia (Smith et al., 2003; Yi & Hwang, 2003). Smith et al. (2003) propusieron la autogestión del aprendizaje como el grado en que un individuo siente que es autodisciplinado y puede participar en el aprendizaje autónomo. Además, Yi y Hwang (2003) introdujeron la Orientación a Metas de Aprendizaje como un determinante indirecto de aceptación del aprendizaje electrónico basándose en la investigación de Nicholls (1984). Nicholls (1984) propuso que existen dos tipos de metas.

La primera es la Orientación a Metas de Aprendizaje y la segunda es la Orientación a Metas de Desempeño. Las personas con Orientación a Metas de Aprendizaje quieren entender algo nuevo o mejorar su nivel de competencia. En contraste, los individuos con Orientación a Metas de Desempeño ven la habilidad como una entidad fija que revela su capacidad intelectual. Además, Shih (2008) introdujo las Expectativas de Resultados Personales como antecesor de la intención de uso basada en sobre las obras de Vroom (1964) y Bandura (1986). Vroom propuso que una mayor expectativa de resultados también aumenta la motivación individual para realizar una acción. Los resultados de Bandura (1986) reforzaron esta teoría, porque propuso que las expectativas con respecto

a las consecuencias de un comportamiento influyen fuertemente en las acciones individuales.

Motivados por tales estudios previos, Terzis y Economides (2011) proponen el constructo Expectativa de Meta (EM). La Expectativa de Meta (EM) es una variable que influye en la creencia de un individuo que está preparado adecuadamente para usar EBC. Se definen dos dimensiones de esta variable. En primer lugar, en una evaluación sumativa, los estudiantes tienen que estudiar y estar preparados para responder correctamente a las preguntas. Entonces, la primera dimensión es la preparación del estudiante para realizar la EBC. Esta dimensión mide si un estudiante está satisfecho con su preparación, no mide la preparación de aspectos cualitativos o cuantitativos. La segunda dimensión implica el nivel deseable de éxito para cada estudiante. Antes de la evaluación, cada estudiante trata de predecir su rendimiento basado en su estudio y el nivel de dificultad hipotética de la evaluación. El estudiante establece una meta con respecto a un porcentaje de respuestas correctas que le proporcionan un rendimiento satisfactorio. Se asumen las siguientes hipótesis:

H10: La Expectativa de Meta tendrá un efecto positivo en la Utilidad Percibida.

H11: La Expectativa de Meta tendrá un efecto positivo en la Percepción Lúdica.

### **Contenido**

Los sistemas de aprendizaje electrónico utilizan las TIC para automatizar la transmisión de contenidos. Basándose en Doll y Torkzadeh (1988), Wang (2003) propuso el Contenido (C) como uno de los determinantes de la satisfacción del alumno al usar sistemas de aprendizaje. Por otra parte, Shee y Wang (2008) propusieron que la variable Contenido del Sistema tiene un gran valor en la satisfacción de los estudiantes. También mencionaron la necesidad de expertos no técnicos, como maestros durante la construcción, el funcionamiento y el mantenimiento del sistema.

En CBAAM se estudian dos dimensiones del Contenido; el contenido del curso y el contenido de las preguntas. En cuanto al contenido del curso, se cree que afecta en gran medida la Utilidad Percibida y la Percepción Lúdica de la EBC. El contenido del curso puede determinar si es útil o no, interesante o no y finalmente difícil o no. En este modelo

también, el contenido de las preguntas se examina para determinar si son claras, fáciles para entender y relacionado con el contenido del curso. Estas dimensiones del contenido se proponen solo en este modelo. Modelos anteriores examinaron el contenido para diferentes propósitos. Por lo tanto, el modelo asume que el Contenido afectará la Utilidad Percibida, Percepción Lúdica, Expectativa de Meta e Intención de Uso, como en las siguientes hipótesis:

H12: El Contenido tendrá un efecto positivo en la Utilidad Percibida.

H13: El Contenido tendrá un efecto positivo en la Percepción Lúdica.

H14: El Contenido tendrá un efecto positivo en la Expectativa de Meta.

H15: El Contenido tendrá un efecto positivo en la Intención de Uso de la EBC.

En resumen, la figura 4 muestra el marco conceptual del modelo CBAAM y las relaciones hipotéticas entre los constructos adoptados. En este trabajo se analizaron los constructos que afectan la intención de utilizar la Evaluación Basada en Computadora mediante el CBAAM.

Relaciones hipotéticas entre los constructos adoptados:

H1: La percepción lúdica tendrá un efecto positivo en la intención de utilizar la Evaluación Basada en Computadora (EBC)

H2: La utilidad percibida tendrá un efecto positivo en la intención de utilizar la EBC

H3: La utilidad percibida tendrá un efecto positivo en la percepción lúdica

H4: La facilidad de uso percibida tendrá un efecto positivo en la intención de utilizar la EBC

H5: La facilidad de uso percibida tendrá un efecto positivo en la utilidad percibida

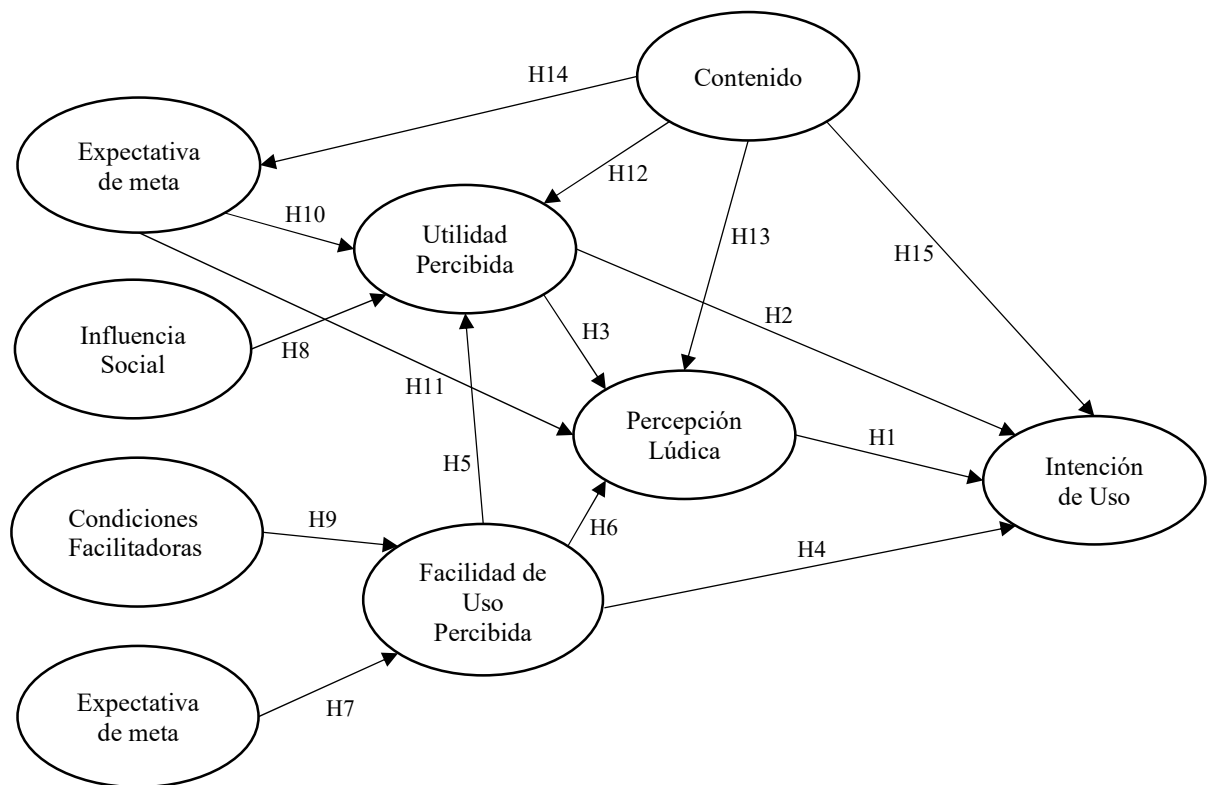
H6: La facilidad de uso percibida tendrá un efecto positivo en la percepción lúdica

H7: La autoeficacia informática tendrá un efecto positivo en la facilidad de uso percibida

H8: La influencia social tendrá un efecto positivo en la utilidad percibida

H9: Las condiciones facilitadoras tendrán un efecto positivo en la facilidad de uso percibida





**Figura 4.** Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM) de Terzis y Economides (2011).

H10: La expectativa de meta tendrá un efecto positivo en la utilidad percibida

H11: La expectativa de meta tendrá un efecto positivo en la percepción lúdica

H12: El contenido tendrá un efecto positivo en la utilidad percibida

H13: El contenido tendrá un efecto positivo en la percepción lúdica

H14: El contenido tendrá un efecto positivo en la expectativa de meta

H15: El contenido tendrá un efecto positivo en la intención de uso de la EBC

En este trabajo se ha optado por adoptar el Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM) de Terzis y Economides (2011) como pilar fundamental para investigar la aceptación de las evaluaciones electrónicas en un contexto educativo. Esta elección se justifica por varias razones esenciales que se alinean con los objetivos y el contexto de nuestra investigación.

Primero, el CBAAM se diseñó específicamente para abordar las particularidades de la aceptación tecnológica en el ámbito de las evaluaciones educativas, lo que lo hace especialmente pertinente para nuestro estudio. A diferencia de modelos más generales de aceptación tecnológica, como el TAM (Modelo de Aceptación Tecnológica) o el UTAUT (Teoría Unificada de Aceptación y Uso de la Tecnología), el CBAAM incorpora constructos y variables que reflejan de manera más directa los desafíos y consideraciones únicas de las evaluaciones basadas en computadora.

Segundo, Terzis y Economides (2011) enfatizan en su modelo la importancia de factores específicos de contexto, como la autoeficacia, la ansiedad ante la computadora y la actitud hacia la computadora, que se han demostrado críticos en la adopción de tecnologías de evaluación electrónica. Esta atención hacia los aspectos psicológicos y afectivos de la aceptación tecnológica es crucial para comprender las barreras y facilitadores en la adopción de evaluaciones electrónicas por parte de los estudiantes y docentes.

Tercero, el enfoque metodológico que subyace al CBAAM, basado en la teoría y apoyado por una base empírica, ofrece un marco robusto para el análisis y la interpretación de los datos recogidos en nuestra investigación. La capacidad de este modelo para integrar diversas perspectivas teóricas en un marco cohesivo nos permite abordar nuestra pregunta de investigación desde múltiples ángulos, enriqueciendo así la profundidad y la relevancia de nuestros hallazgos.

Finalmente, la elección del CBAAM se fundamenta en su relevancia específica para el contexto de evaluación electrónica, su enfoque integral en factores psicológicos, técnicos y su sólida base teórica. Por estas razones, el CBAAM constituye la base teórica más adecuada para guiar nuestra investigación sobre la aceptación de la evaluación automática de conceptos básicos de física basados en preguntas de respuesta corta.

### **3.15 Conclusiones**

Este capítulo sobre el marco teórico ofrece una perspectiva integral sobre la relevancia del formato de pregunta de respuesta corta para evaluar la comprensión conceptual en la educación, destacando la utilidad de las evaluaciones basadas en computadora. Mediante el análisis de los modelos de aceptación de tecnología, se identifican factores clave para la adopción efectiva de estas herramientas, señalando la importancia de superar barreras en profesores y estudiantes. La interdisciplinariedad de la investigación se evidencia en la conexión con campos como los inventarios de conceptos, e-assessment, procesamiento del lenguaje natural y modelos de aceptación tecnológica, proporcionando una base teórica robusta para abordar los desafíos y aprovechar las oportunidades de la evaluación basada en computadora en el contexto educativo actual. Este enfoque teórico fundamenta la investigación y abre camino para una implementación y exploración de innovaciones educativas.

## **CAPÍTULO 4. DISEÑO METODOLÓGICO**

### **4.1 Introducción**

Este capítulo aborda integralmente el diseño de investigación del estudio. Inicia con una panorámica de la filosofía de investigación adoptada, seguida por una descripción detallada del enfoque metodológico implementado. A continuación, se justifica el diseño de investigación elegido y su pertinencia específica para este estudio, incluyendo una discusión sobre la selección de los participantes. Se explican los fundamentos que motivaron la elección del enfoque y la metodología de investigación. En la sección dedicada a técnicas e instrumentos, se profundiza en los métodos cuantitativos empleados. Se esclarece el proceso de recolección de datos, junto con las estrategias adoptadas para el análisis de dichos datos. Asimismo, se establece la conexión entre la pregunta de investigación y los métodos utilizados. Para concluir, se examinan las consideraciones éticas vinculadas a la recolección de datos en este estudio.

### **4.2 Filosofía de investigación**

Según Saunders, Lewis y Thornhill (2009), la filosofía de investigación es una creencia sobre cómo deben recolectarse, investigarse y utilizarse los datos acerca de un fenómeno. Además, los autores afirmaron que "La filosofía de investigación es un término global relacionado con el desarrollo del conocimiento y la naturaleza de ese conocimiento" (p. 107). Más aún, sostuvieron que la filosofía de investigación adoptada abarca suposiciones esenciales sobre cómo el investigador ve el mundo. Estas suposiciones forman la base de la estrategia de investigación y los métodos elegidos.

La epistemología y la ontología son dos de los elementos principales de la filosofía del conocimiento que ayudan a los investigadores a determinar la elección adecuada de la filosofía de investigación. Según Doolin (1996), la epistemología es una filosofía de investigación relacionada con las suposiciones sobre lo que constituye conocimiento válido, mientras que la otra filosofía de investigación, la ontología, se refiere a las suposiciones sobre la naturaleza de la realidad física y social. En otras palabras, la epistemología es sobre 'la forma en que conocemos las cosas', mientras que la ontología es sobre 'lo que son las cosas'. Los supuestos epistemológicos y ontológicos influyen en

la metodología elegida por los investigadores. La metodología elegida aquí se refiere a lo que consideran métodos apropiados para obtener el conocimiento requerido. De manera congruente, existen diferencias filosóficas importantes entre un estudio cuantitativo que se centra en cifras y hechos, y un estudio cualitativo que se preocupa por entender el comportamiento humano desde la perspectiva del informante.

#### **4.3 Perspectiva de Investigación del Presente Estudio**

La filosofía de este estudio refleja los principios del positivismo, ya que se alinea con la postura filosófica del científico natural e implica trabajar con una realidad social observable para producir generalizaciones de tipo ley. Promete un conocimiento inequívoco y preciso y se origina en las obras de Francis Bacon (Bacon, 1620), Auguste Comte (Comte, 1830) y el grupo de filósofos y científicos de principios del siglo XX conocidos como el Círculo de Viena (Stadler, 2001). La etiqueta de positivismo se refiere a la importancia de lo que se 'postula' (en inglés "*posited*"), se refiere a los hechos, datos y fenómenos observables que son considerados como la base de conocimiento seguro y objetivo. Esto enfatiza el enfoque positivista en un método científico empirista estricto diseñado para producir datos y hechos puros no influenciados por la interpretación o el sesgo humano. La tabla 4 resumen alguna de las características del positivismo. Hoy en día existe una 'desconcertante variedad de positivismos', algunos cuentan hasta 12 variedades (Crotty 1998).

Los enfoques deductivo e inductivo son los principales tipos de enfoque de investigación. Usar un enfoque deductivo significa que un investigador desarrolla una teoría y una o más hipótesis y luego usa una estrategia de investigación para probarlas. Por el contrario, el uso de un enfoque inductivo implica la recopilación de datos y luego el desarrollo de la teoría como resultado del análisis de datos. La deducción se relaciona más con el positivismo y la inducción al interpretativismo (Saunders et al., 2009).

En este contexto, en la figura 5 se muestran los fundamentos metodológicos de este trabajo. El estudio realizado corresponde a la filosofía positivista, con enfoque basado en la deducción. Un método cuantitativo y una estrategia de tipo cuasiexperimental. El horizonte de tiempo es transversal, es decir, recolectamos los datos en un solo momento.

**Tabla 4. Características del positivismo**

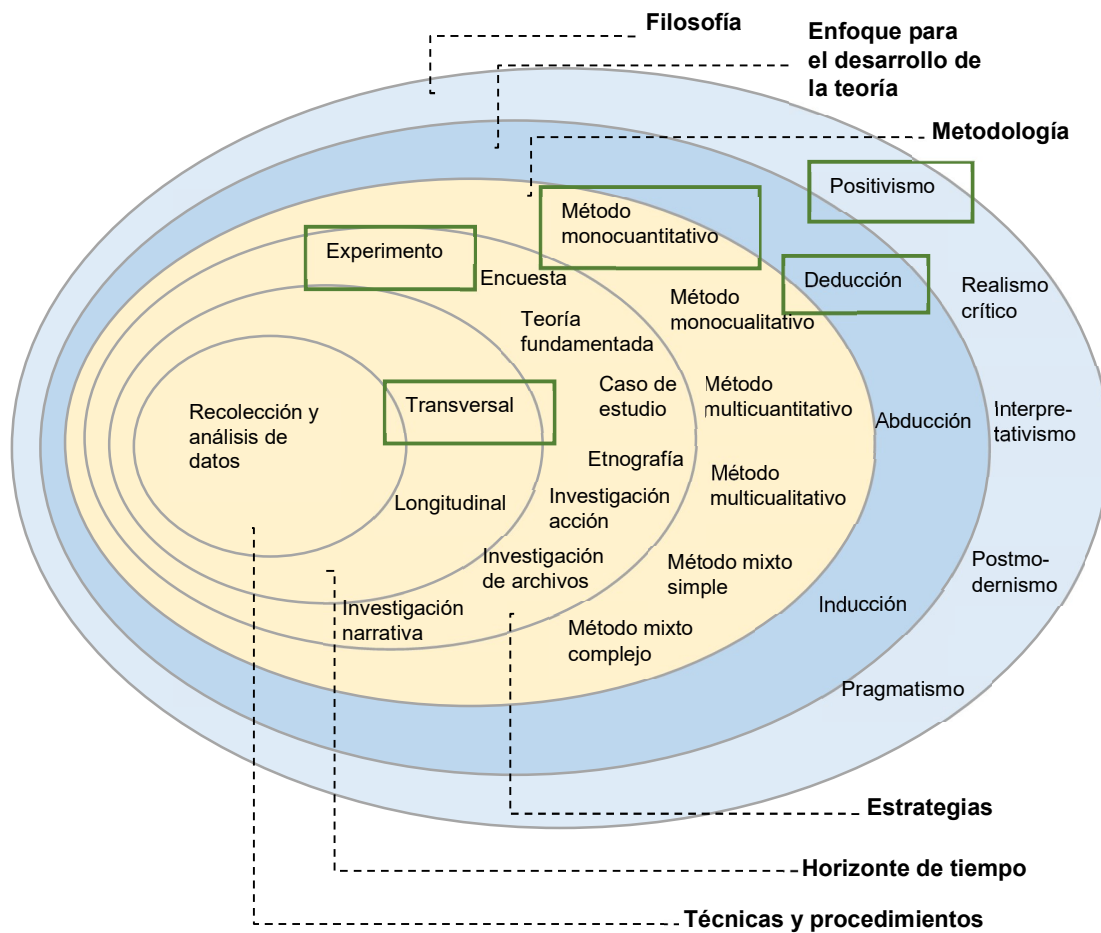
<b>Ontología</b> (naturaleza de la realidad o ser)	<b>Epistemología</b> (Qué constituye conocimiento aceptable)	<b>Axiología</b> (El rol de los valores)	<b>Método típico</b>
<b>POSITIVISMO</b>			
Reales, externos, independiente Una realidad verdadera (universalismo) Granular (cosas) Ordenado	Método científico Observables y medibles hechos Generalizaciones similares a leyes Números Explicación causal y predicción como contribución	Investigación libre de valores El investigador está separado, neutral e independiente de lo que es investigado Investigador mantiene postura objetiva	Típicamente deductivo, muy estructurado, muestras grandes, medición, métodos de análisis típicamente cuantitativos, pero sólo un rango de datos puede ser analizado

Fuente: *Saunders et al. (2019)*.

#### **4.4 Enfoque y alcance de la investigación**

La presente investigación se enmarca en un paradigma cuantitativo, de tipo cuasiexperimental y de alcance explicativo. De acuerdo con Creswell (2013), el enfoque cuantitativo es un tipo de investigación social que se basa en la recolección y análisis de datos numéricos y estadísticos para describir y explicar las relaciones causales entre variables. Este estudio se caracteriza por su naturaleza transversal, dado que la recolección de datos se realizó en un único momento, las variables se midieron tal y como se encuentran en su estado natural, sin ninguna manipulación (Sousa et al., 2007). Debido a que se pretende determinar el tipo de relación entre las variables de estudio y precisar cómo es dicha conexión, el alcance es explicativo (Hernández et al., 2010).

Dentro de este marco, el estudio persigue una comprensión detallada y la explicación de los fenómenos observados. En el ámbito cuantitativo, es posible realizar investigaciones predictivas que establezcan relaciones causales entre variables. Un ejemplo son los estudios de modelos explicativos basados en ecuaciones estructurales que proponen teorías para una mejor comprensión de un fenómeno específico.



**Figura 5. El 'diagrama cebolla'.**

Fuente: *Saunders, Lewis y Thornhill (2018).*

Este enfoque cuantitativo prioriza la medición objetiva de los fenómenos, empleando técnicas estadísticas para discernir patrones y establecer relaciones entre variables, enfatizando la precisión y la objetividad en la investigación.

#### 4.5 Diseño de investigación

El estudio de investigación que se presenta utiliza un diseño cuasi-experimental (et al., 2010). La distinción principal de este enfoque es que la asignación de los participantes a los grupos de intervención no se realiza de manera aleatoria (Keppel, 1991). En lugar de

ello, se emplean grupos preestablecidos, tal como ocurre dentro de un contexto educativo. Debido a las limitaciones para realizar asignaciones aleatorias en estos entornos, la investigación se llevó a cabo con grupos preexistentes. Específicamente, los grupos empleados corresponden a aquellos formados por estudiantes del primer semestre de las carreras de ingeniería. El diseño incluye exclusivamente la aplicación de una posprueba y la utilización de un grupo de control (Hernández et al., 2010). La tabla 5 muestra un resumen con la descripción de las variables de investigación. La figura 6 y 7 muestra un esquema de la metodología y el experimento respectivamente.

**Tabla 5.** Descripción de las variables de investigación

<b>Variable</b>	<b>Descripción de la variable</b>	<b>Instrumento de recolección de datos</b>
Variable independiente	Estrategia de evaluación automática implementada en plataforma MOODLE	Instrumento de evaluación que contiene 27 preguntas en formato de respuesta corta
Variable dependiente	Impacto que tendrá el uso de esta tecnología en los estudiantes	Modelo de Aceptación de Evaluación Basado en Computadora (CBAAM). Encuesta de 30 reactivos

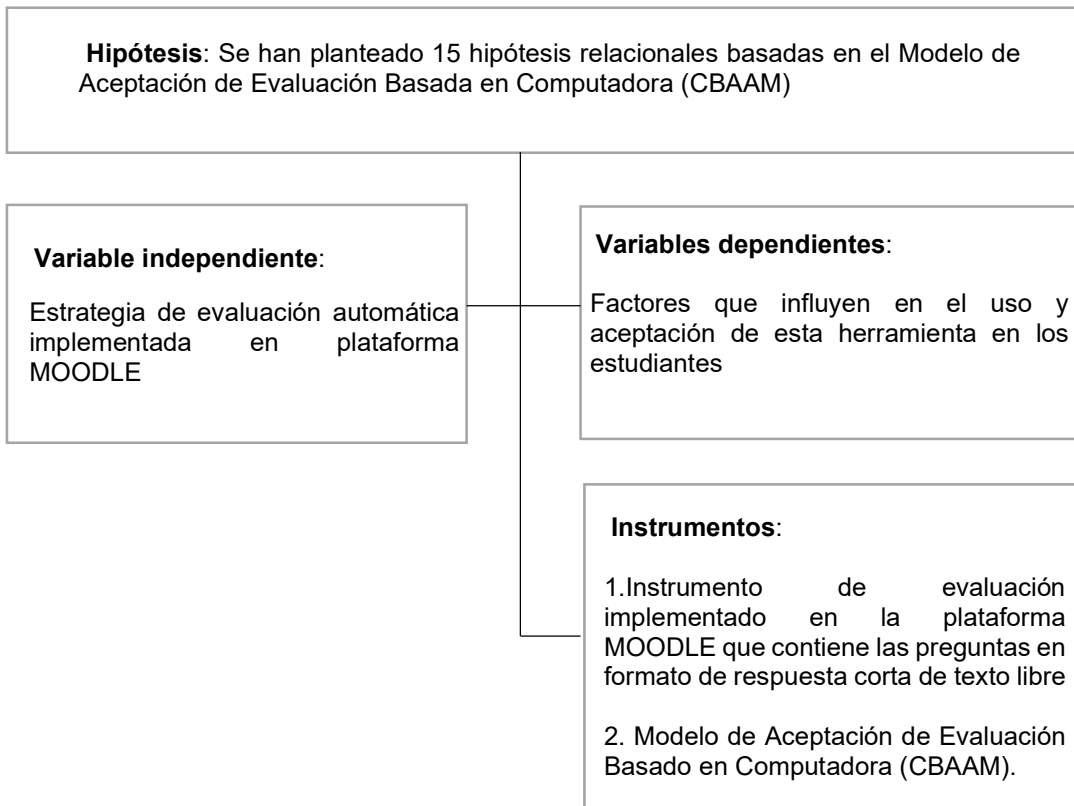
Fuente: elaboración propia.

#### **4.6 Población, muestra y muestreo**

##### **La población objetivo de la investigación**

En estadística, la población objetivo es el conjunto de individuos o el grupo al que se aplica la encuesta. Esto se refiere a los individuos que están siendo estudiados o que deben responder las preguntas de la encuesta (Kitchenham & Pfleeger, 2002). La universidad en cuestión es una institución educativa privada mexicana con 80 años de existencia, que cuenta con 26 campus en varios estados del país y una población total de aproximadamente 94,424 estudiantes.

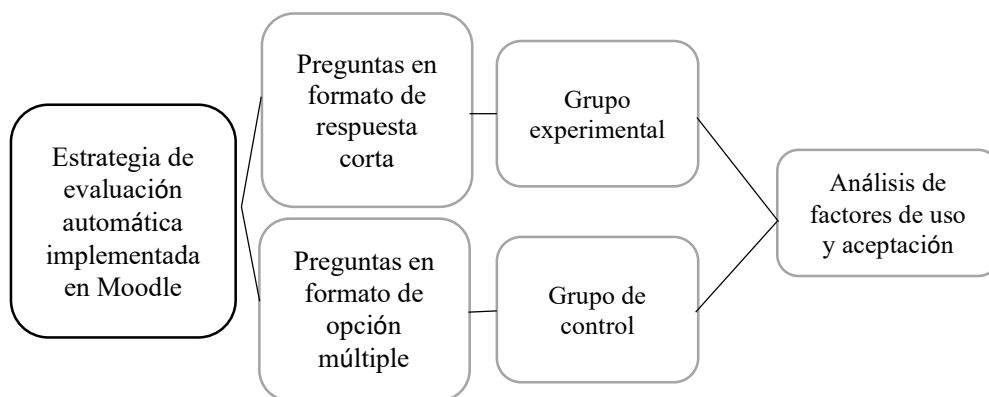




**Figura 6. Resumen metodología.**

Fuente: *elaboración propia.*

La presente investigación se llevó a cabo en un campus de esta universidad situado en la Ciudad de Puebla, por lo que se define a la población objetivo como los alumnos de primer semestre en dicho campus que cursen alguna carrera del área de ingeniería y que estén inscritos en las unidades de formación que impliquen física introductoria. Estos estudiantes son mujeres y hombres con un promedio de edad de 18 años, de un nivel socioeconómico medio a alto, y que cuentan con recursos para acceder a herramientas tecnológicas como laptops, teléfonos inteligentes e internet como parte de su cotidianeidad. La población objetivo contó con un total de 285 alumnos. Por tanto, el universo considerado para el experimento son todos los estudiantes universitarios de primer semestre de esta universidad, con sedes en varias ciudades.



**Figura 7. Esquema del experimento.**

Fuente: *elaboración propia.*

### ***El marco muestral de la investigación***

Después de establecer el alcance de la población objetivo, con el apoyo del departamento de dirección escolar de la universidad en cuestión, se generó una lista de todos los individuos de la población para la investigación. La diferencia entre una población y un marco muestral es que la población es general y el marco muestral es específico. Para asegurarse de que el marco muestral desarrollado sea adecuado, se procedió de la siguiente manera: (1) se incluyó a todos los participantes en la población objetivo; (2) se excluyó a todos los participantes que no estaban en la población objetivo; (3) se recopilaron los datos correspondientes a los horarios y ubicación del aula de clase. Con esta información y con la autorización previa de los profesores titulares se visitaron un total de 14 grupos para explicar los objetivos de la investigación e invitar a los alumnos a participar de forma voluntaria, solicitándoles que completaran los formularios de consentimiento informado.

### **Muestreo**

Kitchenham y Pfleeger (2002) afirmaron que una muestra utilizable es un subgrupo representativo de la población objetivo. Mencionaron que la parte más importante de su definición de muestra es la palabra “representativa”. Creen que los resultados de la investigación no se pueden generalizar a una población objetivo a menos que la muestra sea representativa, por lo que los investigadores deben asegurarse de que la muestra seleccionada sea representativa. Además de eso, incluso afirmaron que, de ser posible, “una población objetivo debería representarse como una lista finita de todos sus miembros”. Reconocieron explícitamente que, si un investigador trabaja con una población pequeña, posiblemente debería intentar obtener respuestas de todos los individuos de la población objetivo.

La muestra fue no probabilística dirigida (Arias, 2006) ya que se requería que los participantes cumplieran con los siguientes criterios de inclusión: 1) Estar inscritos en un periodo académico, 2) que estén cursando unidades de formación que involucren física introductoria. El criterio de exclusión es que no sean repetidores de la materia.

La población accesible son todos los alumnos de primer semestre de alguna carrera del área de ingeniería y que cumplan con los criterios de inclusión establecidos, y cuya participación está sujeta a la autorización de los profesores titulares de los cursos. El tamaño de la población es de 285 estudiantes. Las unidades de muestreo corresponden a la propia organización de los 14 grupos escolares que se visitaron. La muestra se formó con base en la aprobación de los docentes responsables y la participación voluntaria de los estudiantes, resultando en un total de 84 alumnos. La edad media fue de 18.39 años (DE = 0.94) con un rango de 17 a 22 años; por sexo respondieron 26 (30%) mujeres, 54 (65%) hombres y 4 (5%) no lo indicaron. Promedio escolar se ubicó en 93.043 (DE = 5.33).

En cuanto al tamaño de la muestra, el valor mínimo recomendado para este tipo de estudios está definido por las dos pautas siguientes: a) diez veces mayor que el número de elementos para el constructo más complejo, y b) diez veces el mayor número de

variables independientes que impactan una variable dependiente (Chin, 1998). Debido a que el modelo en cuestión tiene cuatro variables independientes que afectan a una dependiente (utilidad percibida), la muestra de 84 participantes excede el mínimo recomendado de 40.

#### **4.7 Recolección de los datos cuantitativos**

El diseño experimental manipula sistemáticamente una o más variables para evaluar cómo esta manipulación impacta en un resultado (o resultados) de interés. Es importante destacar que un experimento aísla los efectos de esta manipulación manteniendo constantes todas las demás variables. Cuando un grupo recibe un tratamiento y el otro grupo no (lo cual es una variable manipulada de interés), el experimentador puede aislar si el tratamiento y no otros factores influyen en el resultado (Creswell, 2013). Tal como se ha señalado previamente, esta investigación se distingue por la implementación de un diseño que incorpora exclusivamente una posprueba y un grupo de control, según Hernández et al. (2010). Además, para explorar las relaciones entre variables y abordar las preguntas e hipótesis planteadas, se utilizará la encuesta como técnica de recolección de datos.

Las encuestas ofrecen una descripción cuantitativa de las tendencias, actitudes y opiniones dentro de una población, además de examinar las relaciones entre variables al estudiar una muestra representativa. Los diseños de encuesta son herramientas valiosas para los investigadores, ya que facilitan la respuesta a tres categorías principales de preguntas: (a) preguntas descriptivas, que buscan cuantificar y describir características o comportamientos específicos; (b) preguntas sobre relaciones entre variables, dirigidas a explorar conexiones o asociaciones entre distintos factores; y (c) preguntas sobre relaciones predictivas, que intentan determinar cómo una variable puede influir en otra a lo largo del tiempo (Creswell, 2013).

#### **4.8 Instrumentos aplicados en la investigación**

Los instrumentos de recolección que se desarrollaron en este estudio fueron los siguientes:

1. Instrumento educativo de evaluación que contiene las 27 preguntas sobre conceptos de física en formato de respuesta corta de texto libre y que serán calificadas de forma automática en la plataforma MOODLE. Este cuestionario estará basado en el inventario de conceptos de física propuesto por Hestenes et al. (1992) y es el que se aplicará al grupo experimental. (Apéndice B)

2. Instrumento educativo de evaluación que contiene las 27 preguntas sobre conceptos de física en formato de opción múltiple y que serán calificadas de forma automática en la plataforma MOODLE. Este cuestionario estará basado en el inventario de conceptos de física propuesto por Hestenes et al. (1992) y es el que se aplicará al grupo control. (Apéndice C)

3. Modelo de Aceptación de Evaluación Basada en Computadora (Computer Based Assessment Acceptance Model, CBAAM). Este instrumento contiene 30 ítems con las siete opciones de respuesta (1 - Absolutamente en desacuerdo, 2 - Fuertemente en desacuerdo, 3 - En desacuerdo, 4 - Ni en acuerdo ni desacuerdo, 5 - De acuerdo, 6 - Fuertemente en acuerdo, 7 - Absolutamente de acuerdo). Esta escala fue desarrollada y validada por Terzis y Economides (2011) para investigar la intención de los estudiantes de utilizar la evaluación basada en computadora. (Apéndice A)

## **4.9 Confiabilidad de los instrumentos**

### **4.9.1 Instrumento de evaluación automática de conceptos básicos de física**

En relación con el instrumento educativo de evaluación, es importante destacar que, como resultado de la revisión de literatura, para el desarrollo de esta investigación se eligió utilizar una tecnología desarrollada en la comunidad de software libre y que se integra al sistema gestor de aprendizaje MOODLE a través de un plugin llamado *Pattern Match*. Con esta integración, ya se pueden construir preguntas en formato de respuesta libre y con el potencial de calificarse de forma automática.

Este sistema utiliza un algoritmo que identifica palabras clave, considerando tanto los sinónimos como los errores de escritura, aunque no incorpora reglas gramaticales ni

estructuras sintácticas en su funcionamiento. Una vez que se comprendieron los aspectos técnicos para la creación de preguntas en formato de respuesta libre y su correspondiente calificación automática, se procedió al desarrollo de nuestro instrumento de investigación dentro de la plataforma MOODLE que contiene 27 preguntas enfocadas a conceptos básicos de física, lo que implicó la creación de las reglas correspondencia para cada pregunta y que controlan la calificación automática.

### ***Una descripción general de los inventarios de conceptos***

Como ya se mencionó en el capítulo 2, hay muchos inventarios de conceptos que se utilizan en física y astronomía. Un inventario de conceptos es un instrumento de opción múltiple de nivel investigación, que está diseñado para evaluar la comprensión conceptual de un tema en particular (Lindell et al., 2007), con un mínimo contenido matemático.

### ***El inventario de conceptos sobre fuerza***

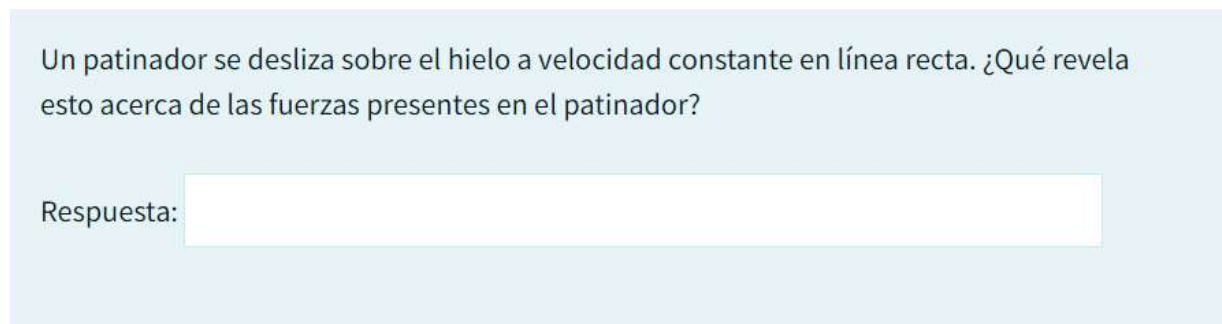
En el capítulo 3 se discutió el desarrollo y la evolución de herramientas diagnósticas esenciales en la enseñanza de la física, como la Prueba de Diagnóstico en Mecánica (MDT) y el Inventario de Conceptos de Fuerza (FCI). Estas herramientas no solo permiten identificar y evaluar la comprensión conceptual de los estudiantes sobre fenómenos físicos clave, sino que también resaltan la importancia de abordar conceptos erróneos comunes en la mecánica newtoniana.

La versión final del FCI, desarrollada por Halloun et al. en 1995, consta de 30 preguntas de opción múltiple, cada una cuidadosamente diseñada para minimizar el contenido matemático. Este instrumento cuenta con más de 50 publicaciones que avalan su validez, incluida su traducción al español.

Las preguntas que utilizaremos en la presente investigación estarán basadas en el FCI, con la variante que serán preguntas en el formato de respuesta corta de texto libre y con evaluación automática. Esto es en principio posible gracias a tecnología como la utilizada en el tipo de pregunta *Pattern Match* (PMatch) en OpenMark (Butcher & Jordan, 2010), que formó la base del tipo de pregunta *Pattern Match* (*patrón de coincidencia*) en el motor para la creación de cuestionarios en la plataforma Moodle.

### ***Pregunta de respuesta corta en la plataforma moodle***

Una vez instalado el plugin Pattern Match, el formato de preguntas de respuesta corta está disponible en la plataforma Moodle. Las preguntas de respuesta corta se escriben en Pattern Match especificando la redacción de la pregunta y las reglas de calificación correspondientes. Un ejemplo de una pregunta de respuesta corta sobre el tema de mecánica newtoniana se muestra en la Figura 8 a continuación.



Un patinador se desliza sobre el hielo a velocidad constante en línea recta. ¿Qué revela esto acerca de las fuerzas presentes en el patinador?

Respuesta:

**Figura 8.** *Ejemplo de una pregunta de respuesta corta.*

Fuente: *elaboración propia.*

Pattern Match califica las respuestas a las preguntas de respuesta libre de la siguiente manera. El autor de la pregunta especifica un conjunto de reglas de calificación, y cada regla está diseñada para hacer coincidir las respuestas con cadenas de caracteres (por ejemplo, en palabras), con capacidad de tener en cuenta también otros factores como letras omitidas y ordenadas incorrectamente, truncamiento, orden de las palabras y espaciado. Algunas de estas reglas coinciden con condiciones positivas que corresponden a respuestas correctas, mientras que otras reglas coinciden con condiciones negativas que corresponden a respuestas incorrectas. En el apéndice B se muestran las reglas calificación de cada pregunta contenida en el instrumento de evaluación automática de conceptos básicos de física.

Cuando el sistema califica una respuesta, se compara con cada regla en secuencia hasta que encuentra una coincidencia con una de las reglas o no coincide con ninguna. Cuando se encuentra una coincidencia con una regla, la respuesta se califica en función de si la regla corresponde a una condición de calificación correcta o incorrecta. Se otorga una puntuación de 1 a una respuesta que coincida con una regla de calificación correcta y se

otorga una puntuación de 0 a una respuesta que coincide con una regla de calificación incorrecta o que no coincide con ninguna regla de calificación. Las reglas de calificación de este trabajo están basadas en el trabajo previo de Butcher y Jordan (2010) y de Parker, M. A. J. (2020). Cabe mencionar que en la revisión de la literatura no se encontraron reportes del uso de este plugin en el idioma español.

A continuación, se resume el procedimiento general y las actividades realizadas para implementar el Instrumento de Evaluación Automática de Conceptos Básicos de Física, basado en preguntas de respuesta corta:

1. Se realizó una revisión exhaustiva de la literatura sobre herramientas para la evaluación automática de preguntas de respuesta corta, identificando las más efectivas.
2. A partir de la información recopilada, se seleccionó específicamente aquella herramienta óptima para diseñar evaluaciones automáticas destinadas a respuestas cortas.
3. Se optó por un inventario de conceptos en física, previamente basado en preguntas de opción múltiple y respaldado por estudios de validez y confiabilidad, para fundamentar el diseño del instrumento de evaluación automática.
4. Se procedió al desarrollo de una adaptación de este instrumento hacia el formato de preguntas de respuesta corta, conservando la integridad de los conceptos evaluados.
5. Se establecieron las reglas de calificación automática adaptadas al nuevo formato de preguntas de respuesta corta.
6. El instrumento de evaluación automática de conceptos básicos de física se implementó en la plataforma MOODLE de la universidad anfitriona en dos formatos distintos: el de preguntas de opción múltiple y el de preguntas de respuesta corta.
7. El instrumento fue piloteado en una muestra de 55 alumnos y se realizaron ajustes relacionados con las reglas de calificación para optimizar su eficacia.



#### **4.9.2 Modelo de Aceptación de Evaluación Basado en Computadora**

Como se mencionó anteriormente, tanto la prueba de validez como la de confiabilidad de la escala del CBAAM se había ya realizado por sus autores Terzis y Economides (2011), sin embargo, el proceso de traducción, adaptación y validación intercultural del instrumento para su uso requiere de un proceso metodológico riguroso con el objetivo de lograr la equivalencia entre el instrumento original y su traducción (Sousa & Rojjanasrirat, 2011, Muñiz et al., 2013).

El proceso consistió en los siguientes pasos:

1. Traducción del instrumento original al idioma de destino (traducción hacia adelante o de un solo sentido, la cual fue realizada por el autor, apoyada en la revisión por 2 colegas expertos en el área de adopción tecnológica.
2. Cegar retrotraducción (cegar traducción hacia atrás o traducción doble ciego) de la versión preliminar inicial traducida del instrumento. Ambas realizadas por 2 investigadores con amplia experiencia en el campo de adopción tecnológica.
3. Comparación de las versiones retrotraducida del instrumento. Con el objetivo de encontrar la equivalencia conceptual, semántica y de contenido. Durante este proceso también se consideraron las traducciones publicadas en el idioma destino de instrumentos parecidos y que son aplicados en el campo de la adopción tecnológica.
4. Pruebas de confiabilidad de la versión final del instrumento traducido en una muestra de la población objetivo que se llevó a cabo en el programa minitab.

Antes de implementar el instrumento CBAAM se realizó un estudio piloto con 55 estudiantes de ingeniería que cursaban el primer año, con el propósito de conocer su opinión acerca de la claridad de los reactivos y corregir los aspectos que, de forma o fondo, dificultaban la comprensión. A estos participantes se les informó el motivo de la prueba y se les solicitó su consentimiento informado a través de un formulario electrónico.

En seguida, se digitalizó el instrumento final y se agregó en la plataforma moodle para que estuviera disponible inmediatamente después de que los alumnos terminaran de

usar el sistema de evaluación basado en preguntas de respuesta corta con evaluación automática.

Los 30 ítems fueron evaluados por sus medias, desviaciones estándar, asimetría y curtosis. Para evaluar la confiabilidad de instrumento (CBAAM), se obtuvo un coeficiente de consistencia interna con la prueba de alpha de Cronbach con los 30 reactivos de 0.9190, que resultó alto. el coeficiente puede oscilar entre 0 y 1, en donde 0 significa nula confiabilidad y 1, confiabilidad total.

Se utilizó el programa minitab 21.3.1 para efectuar el cálculo; los coeficientes de consistencia interna de las subescalas se encontraron en un rango de 0.6838 a 0.9325, por lo que se concluye que la escala es confiable para su utilización en este contexto Únicamente, la subescala de expectativa de meta está por debajo de 0.70, mínimo que se considera adecuado. Los resultados de este análisis se muestran en el apéndice D. Es importante destacar que, para el estudio principal, se llevó a cabo nuevamente un análisis de fiabilidad y validez de los constructos del modelo, empleando otra muestra de la población objetivo. Los hallazgos se presentan en el capítulo siguiente.

Se presenta a continuación un resumen de los procedimientos y actividades llevados a cabo para la implementación del Instrumento CBAAM.

1. Se llevó a cabo una revisión profunda de la literatura relacionada con los modelos de aceptación tecnológica, con el fin de fundamentar teóricamente el estudio
2. Esta revisión se extendió al concepto de evaluaciones electrónicas (e-assessment) y términos relacionados, ampliando el marco teórico.
3. Con base en el análisis de la información obtenida, se hizo una selección cuidadosa de los modelos de aceptación de evaluaciones electrónicas que mejor se alineaban con los objetivos del estudio.
4. Se decidió emplear el Modelo de Aceptación de Evaluación Basado en Computadora propuesto por Terzis y Economides (2011), el cual cuenta con sólido respaldo en estudios previos de validez y confiabilidad.

5. Posteriormente, se desarrolló una adaptación de este modelo, que incluyó la traducción del instrumento al contexto lingüístico y cultural de la población objetivo.
6. El instrumento adaptado fue sometido a un piloto con una muestra de 55 estudiantes, lo que permitió identificar y realizar ajustes necesarios en términos de redacción y formato
7. A partir de esta fase piloto, se efectuaron pruebas de confiabilidad para asegurar la precisión y consistencia de la versión final traducida del instrumento.
8. Finalmente, en el marco del estudio principal, se efectuó una nueva evaluación de la fiabilidad y validez de los constructos del modelo, utilizando otra muestra de la población objetivo.

#### **4.10 Procedimientos para la recolección de datos**

El trabajo de campo se realizó en una universidad privada en Puebla capital y se dividió en dos etapas principales: inicialmente, en marzo de 2022, se ejecutó una prueba piloto; luego, durante octubre y noviembre del mismo año, se llevó a cabo el estudio principal. La finalidad de la prueba piloto fue realizar una evaluación preliminar de los instrumentos de investigación empleados y los aspectos operativos implicados, asegurando su adecuación y eficacia antes de proceder con la fase principal del estudio.

##### **4.10.1 Estudio piloto**

Una vez que se obtuvo la autorización de los profesores titulares, se visitaron dos grupos en la materia de Experimentación Física. En estas sesiones, se presentaron los objetivos del estudio a los estudiantes, quienes fueron invitados a participar voluntariamente y se les proporcionaron los formularios de consentimiento informado para completar.

Un total de 55 estudiantes manifestaron su interés por participar en el estudio. Para asegurar la equivalencia de grupos (Hernández et al., 2010) se llevó a cabo un proceso de asignación aleatoria a través de la plataforma Moodle, dividiendo a los estudiantes en un grupo experimental (n=28), que respondería al instrumento educativo que contiene las 27 preguntas en formato de respuesta corta y un grupo de control (n=27) que

respondería el mismo instrumento, pero con preguntas en formato de opción múltiple. La sesión se desarrolló en formato completamente virtual.

Durante esta sesión se aplicaron los siguientes instrumentos:

1. Instrumento educativo de evaluación basado en el inventario de conceptos de física propuesto por Hestenes et al., (1992), que contiene las 27 preguntas en formato de respuesta corta y que son evaluadas de forma automática en la plataforma MOODLE.
2. Instrumento educativo de evaluación basado en el inventario de conceptos de física propuesto por Hestenes et al., (1992), que contiene las 27 preguntas en formato de opción múltiple y que son evaluadas de forma automática en la plataforma MOODLE.
3. Modelo de Aceptación de Evaluación Basado en Computadora (CBAAM) propuesto por Terzis y Economides (2011) para medir las percepciones y actitudes de los estudiantes al utilizar el sistema de evaluación propuesto. Encuesta de 30 reactivos.

Durante esta fase de la prueba piloto, el objetivo principal era verificar los aspectos operativos implicados en el estudio y realizar una evaluación inicial de los instrumentos utilizados.

#### **4.10.2 Resultados preliminares de la prueba piloto**

El objetivo principal de la prueba piloto, como se ha indicado, consistió en realizar una evaluación general del procedimiento y los instrumentos utilizados. Los hallazgos obtenidos se detallan a continuación:

El instrumento desarrollado que contiene las 27 preguntas en formato de respuesta corta de texto libre y que son evaluadas de forma automática en la plataforma MOODLE, produjo resultados consistentes con su contraparte que contiene las preguntas en formato de opción múltiple. Esto es una muestra de que el funcionamiento de las reglas

correspondencia que se crearon para cada pregunta y que controlan la calificación automática es altamente confiable, sin embargo, se detectaron 2 preguntas que requerían ajustes.

Para el estudio principal, se consideraron los siguientes ajustes significativos:

- En las sesiones donde se explica el proyecto de investigación a los alumnos para solicitar su consentimiento informado y que proporcionen su correo para poder ser enrolados en la plataforma MOODLE, es importante darles unos minutos para que contesten el formato en ese momento, ya que si se deja la opción a que lo hagan después, la confirmación es muy baja, aunque hayan previamente manifestado su interés por participar.
- El tiempo recomendado en la literatura para que contesten el instrumento que contiene las 27 preguntas en formato de opción múltiple es de 30 minutos. En la prueba piloto también se les dio 30 minutos y no hubo comentarios por parte de los alumnos con relación a que no fue suficiente este tiempo.
- Otro aspecto importante, es solicitar a los alumnos que contesten el Instrumento para medir las percepciones y actitudes al utilizar el sistema de evaluación propuesto, inmediatamente después, ya que si se deja la opción a que lo hagan en otro momento, disminuirá la participación.

#### **4.10.3 Estudio principal**

Para el estudio principal se tuvieron en cuenta los aprendizajes clave identificados durante la prueba piloto. Después de obtener la autorización de los profesores titulares, se visitaron 14 grupos de primer semestre de ingeniería distribuidos en los siguientes cursos: 8 grupos de “Aplicación de las leyes de conservación en sistemas ingenieriles” y 6 grupos de “modelación del movimiento en ingeniería”. De esta manera, se logró recabar 202 consentimientos informados, aunque solo 84 estudiantes (41.5% del total) participaron voluntariamente en las sesiones para realizar el examen basado en computadora y, a continuación, completaron el instrumento diseñado para evaluar el uso y aceptación de este tipo de evaluación.

Al igual que en la prueba piloto, primero se registraron a los estudiantes en la plataforma Moodle, luego se procuró la equivalencia de grupos y los estudiantes fueron asignados al azar en uno de los dos grupos: un grupo experimental (n=41), para responder el instrumento educativo que contiene las 27 preguntas en formato de respuesta corta (apéndice B) y un grupo de control (n=43) para responder el mismo instrumento educativo, pero con preguntas en formato de opción múltiple (apéndice C). La sesión se desarrolló en formato completamente virtual.

#### **4.11 Procedimiento Estudio Principal**

El estudio principal se realizó mediante la programación de cuatro sesiones virtuales en octubre y noviembre de 2022. Los estudiantes participaron voluntariamente para completar el instrumento de evaluación sobre conceptos de física. La intervención consistió en la manipulación de la variable independiente, representada por dos estrategias de evaluación automática en diferentes formatos de preguntas ya establecidas en la plataforma MOODLE, con el objetivo de evaluar su impacto en las percepciones y actitudes de los estudiantes (variable dependiente).

El número total de estudiantes que participaron en el estudio principal durante las 4 sesiones virtuales que se programaron fue de 84. Al grupo experimental (n=41) se le aplicó el instrumento de evaluación de conceptos de física con las 27 preguntas en formato de respuesta corta y que serán evaluadas de forma automática en la plataforma MOODLE. Al grupo de control (n=43) se le aplicó exactamente el mismo instrumento, pero con el formato de preguntas de opción múltiple.

El investigador estuvo presente en las 4 sesiones virtuales y el tiempo que se les dio para responder el instrumento de evaluación de conceptos de física implementado en la plataforma Moodle fue de 30 minutos todas las sesiones. Posteriormente, se les pidió que contestaran las 30 preguntas correspondiente al instrumento de Modelo de Aceptación de Evaluación Basado en Computadora, que ya estaba disponible en también en la plataforma Moodle y que puede consultarse en el apéndice A.

#### **4.12 Análisis de los datos**

De acuerdo con Hernández-Sampieri y Mendoza Torres (2018), el enfoque cuantitativo es un conjunto de procesos secuencial y probatorio que, a partir de una idea, va acotándose hasta derivar en objetivos, preguntas de investigación e hipótesis. Las tres características principales del enfoque cuantitativo son:

- 1) La recolección de datos es en forma de puntuaciones. Es decir, los atributos de los objetos de investigación se miden con números.
- 2) El análisis de los datos numéricos se realiza en términos de su variación
- 3) El propósito del análisis es comparar grupos y relacionar los atributos mediante técnicas estadísticas. La variación entre los resultados de ambos grupos permitirá el análisis e interpretación de estos.

#### **4.13 Técnicas estadísticas**

Como se ha mencionado anteriormente, la presente investigación es de enfoque cuantitativo con diseño cuasi-experimental y de alcance explicativo con hipótesis relacionales. Por lo tanto, es factible someter dichas hipótesis a pruebas empíricas y análisis estadístico para determinar si son apoyadas o refutadas (Hernández-Sampieri y Mendoza Torres, 2018). Una prueba de hipótesis es una metodología que nos permite decidir si se puede aceptar o rechazar una afirmación sobre alguna población en función de la evidencia empírica obtenida de una muestra de datos.

La prueba de hipótesis examina dos hipótesis opuestas sobre la población: la hipótesis nula y la hipótesis alternativa (también llamada hipótesis de investigación). La hipótesis nula es la afirmación que se probará con la prueba. La hipótesis nula afirma que “no hay efecto” o “no hay diferencia”. Con base en los datos empíricos, la prueba de hipótesis establece si se puede rechazar la hipótesis nula.

La hipótesis alternativa es lo que se espera probar que es cierto. Para tomar esa decisión se emplea el valor p. Si el valor p es menor que el nivel de significancia (denotado como  $\alpha$  o alfa), entonces se puede rechazar la hipótesis nula. En investigación de ciencias

sociales el valor de alfa suele ser de 5% ( $\alpha=0.05$ ). Al diseñar una prueba de hipótesis, se establece la hipótesis nula como lo que queremos desaprobar. Entonces si  $p<0.05$ , rechazamos la hipótesis nula y tendremos prueba estadística de que la alternativa es verdadera. Si  $p>0.05$  se concluye que no hay evidencia estadística para rechazar la hipótesis nula.

La elección de una técnica estadística adecuada es un pilar fundamental en la investigación científica, especialmente en la exploración de modelos de aceptación de tecnología. En este estudio, se ha adoptado el método de modelo de ecuaciones estructurales basado en mínimos cuadrados parciales (Partial Least Squares-Structural Equation Modeling, PLS-SEM) debido a su robustez y flexibilidad, por lo que se considera que es la técnica estadística más adecuada para nuestro análisis y las pruebas de hipótesis derivadas del modelo teórico CBAAM (Terzis y Economides, 2011) que aplicamos en este trabajo. El software que se empleará para analizar los datos obtenidos será SmartPLS 4.0. En el apéndice E se detallan las variables del modelo, proporcionando sus definiciones tanto conceptual como operacional, así como la metodología específica empleada para su medición.

PLS-SEM es particularmente apropiado para modelos complejos que incluyen múltiples constructos y relaciones causales. Esta técnica es ventajosa en contextos donde el objetivo es la predicción y la explicación de las variables dependientes, lo cual es esencial en la evaluación de modelos de aceptación de tecnología. Además, PLS-SEM permite la evaluación simultánea de las relaciones entre constructos del modelo de medida y del modelo estructural. Esto es crucial cuando se manejan construcciones complejas y se busca explorar la fuerza y la significancia de las relaciones propuestas por las hipótesis. Por estas razones, PLS-SEM es idóneo para probar la validez y la fiabilidad del modelo de aceptación de tecnología en cuestión, así como para examinar la dirección y la fuerza de las relaciones hipotetizadas.

A continuación, se describen brevemente los aspectos generales de esta técnica de análisis, que se aplicará en niveles tanto descriptivos como inferenciales. Además, se



explicará la razón de su selección en función de los tipos de variables y los objetivos específicos de este estudio.

#### **4.14 Modelado de ecuaciones estructurales**

Uno de los métodos más importantes utilizados para el análisis de datos multivariados es la técnica de modelado de ecuaciones estructurales (SEM). Esto se debe a su capacidad para probar modelos causales lineales y aditivos sustentados teóricamente (Haenlein & Kaplan, 2004). SEM puede verse como una combinación de análisis factorial y regresión o análisis de trayectoria. Es una técnica de análisis multivariado muy poderosa que comprende otros métodos de análisis, como los casos especiales (Hox & Bechger, 1998).

Se cree que esta técnica es adecuada para este estudio ya que generalmente se utiliza cuando es necesario construir modelos teóricos donde los componentes básicos de estos modelos son los factores latentes. Las relaciones o vínculos entre estos constructos teóricos se representan mediante regresión o coeficientes de trayectoria entre los factores.

Según Wong (2013) y Sarstedt, Ringle, Smith, Reams y Hair (2014), las dos técnicas principales para estimar modelos de ecuaciones estructurales son el SEM basado en covarianza (CB-SEM) (Jöreskog, 1978, 1982) y el Mínimos cuadrados parciales SEM (PLS-SEM) (Wold, 1985). La técnica elegida para la estimación del modelo de este estudio es PLS SEM.

#### **4.15 Elección de PLS-SEM para la estimación del modelo de este estudio**

Aunque la aplicación inicial de SEM por parte de los investigadores incorporó un enfoque basado en covarianza (CB-SEM) (Haenlein & Kaplan, 2004), muchos estudios implementaron el método PLS basado en varianza (PLS-SEM). PLS-SEM fue construido originalmente por Wold (1980), quien afirmó que "el enfoque PLS para modelos de trayectoria con variables latentes está destinado principalmente a aplicaciones multidisciplinares y de otro tipo donde los problemas explorados son complejos y el conocimiento teórico es escaso". También afirmó que PLS se utiliza cuando están

involucradas las siguientes tres características: “(a) análisis predictivo causal, (b) complejidad de los problemas explorados, y (c) escasez de conocimientos teóricos previos”. Según Fornell y Bookstein (1982), PLS es una técnica de modelado de ecuaciones estructurales que se centra en un proceso iterativo que aumenta la varianza explicada de constructos endógenos.

PLS ejecuta un conjunto iterativo de reglas en las que los valores estadísticos necesarios se calculan con una serie de regresiones de mínimos cuadrados después de crear resultados estadísticos de constructos ponderando las sumas de elementos relacionados con cada constructo (Chin, 1998). Algunas de las razones del mayor uso de PLS-SEM son: (1) la capacidad de utilizar esta técnica en proyectos de investigación aplicada cuando el tamaño de la muestra es pequeño. En otras palabras, cuando hay participantes limitados en la investigación (Wong, 2013). Además, Vinzi, Trinchera y Amato (2010) declararon que PLS es un enfoque para el modelado de ecuaciones estructurales que no tiene suposiciones sobre las distribuciones de datos; (2) la capacidad del método PLS-SEM para manejar problemas de modelado difíciles que normalmente ocurren en las ciencias sociales (Hair, Ringle & Sarstedt, 2013).

En resumen, ambos enfoques de SEM (CB-SEM y PLS-SEM) tienen características individuales que los hacen adecuados para diferentes tipos de estudios de investigación. PLS en particular se considera un enfoque muy útil para el modelado de ecuaciones estructurales debido a su capacidad para explicar la varianza en las variables dependientes al examinar el modelo (Hair, Ringle & Sarstedt, 2012; Sarstedt et al., 2014).

Con base en los detalles enumerados anteriormente, la técnica PLS-SEM se considera un buen enfoque que es capaz de manejar problemas de modelado y datos relacionados con el modelo de adopción de la evaluación electrónica. Por lo tanto, en este estudio se utilizó PLS para analizar las hipótesis propuestas en el Capítulo 1.

#### **4.16 Procedimiento para analizar los datos**

Para analizar el modelo de adopción de la evaluación basada en computadora, en este trabajo se utilizó el enfoque PLS-SEM. De manera similar a la forma en que se utilizan otros métodos, la implementación de PLS-SEM depende de un conjunto de reglas y

pautas que se emplean para evaluar los resultados de la estimación estadística del modelo (Garson, 2016; Hair, Hult, Ringle & Sarstedt, 2016). Estas reglas dependen de la forma en que se vinculan las variables latentes (como se define en el modelo estructural), y del tipo de escala de medición del modelo, ya sea reflexiva o formativa (como se define en el modelo de medición). Los pasos generales para analizar e interpretar los datos mediante PLS-SEM son los siguientes:

1. **Definición del Modelo Conceptual:** Antes de realizar cualquier análisis, es fundamental definir un modelo conceptual. Este modelo se basa en la teoría y la investigación previa y especifica las relaciones hipotéticas entre las variables latentes (constructos no observados) y observadas (indicadores). En este estudio se aplicó el CBAAM de Terzis y Economides, 2011.

2. **Especificación del Modelo de Medida:** Determinar si los constructos son formativos o reflexivos. En un modelo reflexivo, se espera que los indicadores sean causados por el constructo. En un modelo formativo, se considera que los indicadores causan el constructo. En este estudio, se tiene un modelo reflexivo.

3. **Evaluación de la Fiabilidad y Validez:** Evaluar la fiabilidad y validez de los constructos del modelo. Esto incluye la fiabilidad interna (por ejemplo, Alfa de Cronbach), la validez convergente (por ejemplo, la varianza extraída media, AVE) y la validez discriminante (por ejemplo, el criterio Fornell-Larcker, la correlación de las raíces cuadradas del AVE con las correlaciones interconstructos).

4. **Estimación del Modelo Estructural:** Utilizar PLS-SEM para estimar las relaciones entre constructos. El algoritmo de mínimos cuadrados parciales busca maximizar la varianza explicada ( $R^2$ ) de los constructos dependientes.

5. **Evaluación de Resultados:** Evaluar los coeficientes de camino (betas), los niveles de significancia (a través de, por ejemplo, bootstrapping para obtener los t-values), y el

coeficiente de determinación ( $R^2$ ) para las relaciones estructurales. También considerar el efecto tamaño ( $f^2$ ) para evaluar la relevancia práctica de las relaciones.

Finalmente, la interpretación de los resultados en el marco de este estudio deberá permitirnos responder a la pregunta de investigación planteada en el Capítulo 1: ¿Cuáles son los factores que influyen en el uso y aceptación de una herramienta de evaluación automática de conceptos básicos de física basados en preguntas de respuesta corta en estudiantes de una universidad privada de Puebla? El análisis realizado proporcionará datos relevantes sobre cuestiones fundamentales, tales como la consistencia de las relaciones estimadas con las hipótesis propuestas, los constructos que ejercen una mayor influencia sobre otros y la calidad global del modelo en cuanto a su validez y fiabilidad.

Después de presentar un panorama general sobre la metodología PLS-SEM, el siguiente paso consistirá en examinar con mayor detalle el modelo estructural y de medición.

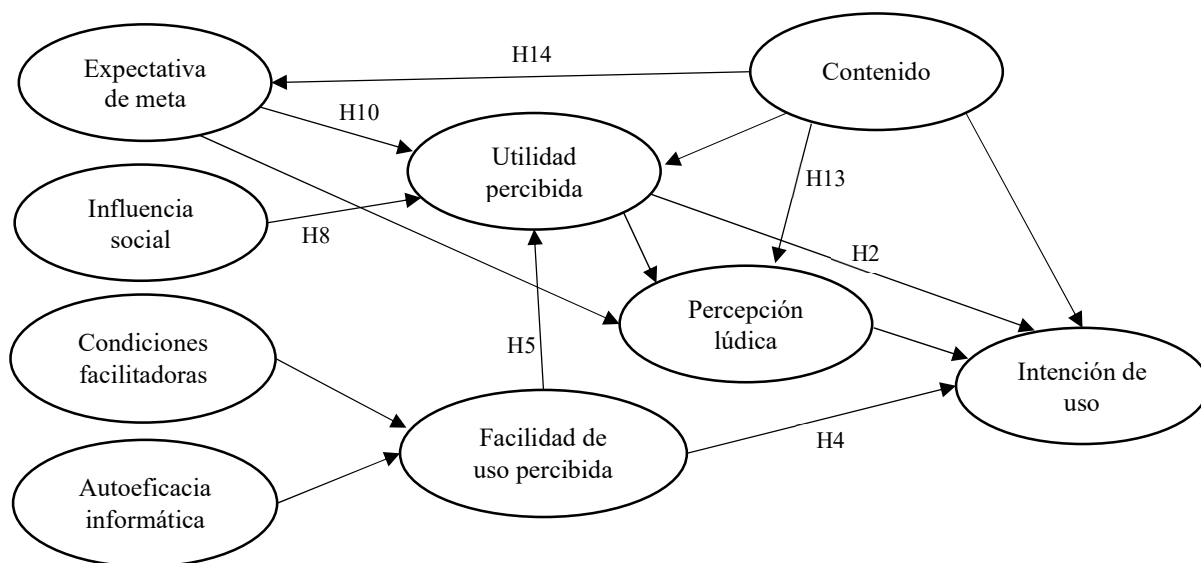
#### **4.17 Modelo estructural**

En el modelo estructural (también llamado modelo interno), las Variables Latentes (VL) (que se muestran en círculos en la Figura 9) están vinculadas entre sí de acuerdo con el modelo propuesto por Terzis y Economides (2011). Las variables latentes a veces se denominan constructos o factores (Garson, 2016). Estas VL se dividen en dos categorías, endógenas y exógenas. Las VL exógenas no tienen ningún predecesor en el modelo estructural, mientras que las variables endógenas tienen al menos otra variable latente que desempeña el papel de predecesor.

A continuación, se proporcionan detalles sobre las VL exógenas y endógenas del modelo estructural de este estudio:

##### **4.17.1 Las variables exógenas**

Una variable latente es exógena si no es un efecto de ninguna otra variable latente en el modelo (no hay flechas entrantes de otras variables latentes) (Garson, 2016). Hay nueve



**Figura 9.** Modelo de investigación (CBAAM) de Terzis y Economides (2011).

VL en el modelo mostrado en la Figura 9. Cuatro de ellas son exógenas y el resto son endógenas. Las cuatro variables exógenas son:

- Autoeficacia informática (AI)
- Influencia Social (IS)
- Condiciones Facilitadoras (CF)
- Contenido (C)

#### 4.17.2 Las variables endógenas

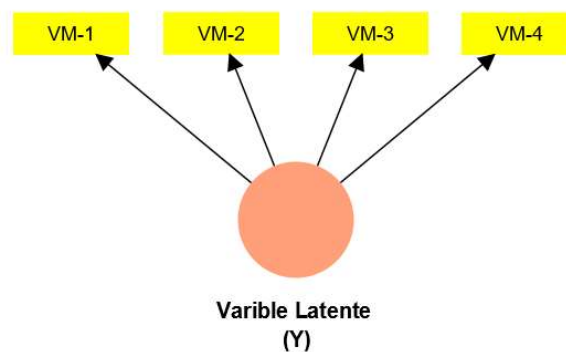
Una variable latente es endógena si es un efecto de al menos otra variable latente (hay al menos una flecha entrante de otra variable latente) (Garson, 2016). Las cinco variables endógenas de este modelo son:

- Facilidad de uso percibida (FDUP)
- Percepción Lúdica (PL)
- Utilidad Percibida (PU)

- Expectativa de Meta (EM)
- Intención de Uso (IU)

#### 4.18 El modelo de medición

En el modelo de medición (también llamado modelo externo), las Variables de Medición (VM) (que se muestran en cuadros amarillos) están vinculadas con sus variables latentes. A las variables de medición a veces se las denomina: variables observadas, variables manifiestas o indicadores (Monecke & Leisch, 2012). En PLS, una variable de medición solo puede relacionarse con una variable latente (Monecke & Leisch, 2012). La forma en que las variables de medición se relacionan con las variables latentes determina el tipo de escala de medición del modelo. Esto puede ser reflexivo (ver Figura 10) o formativo (ver Figura 11).

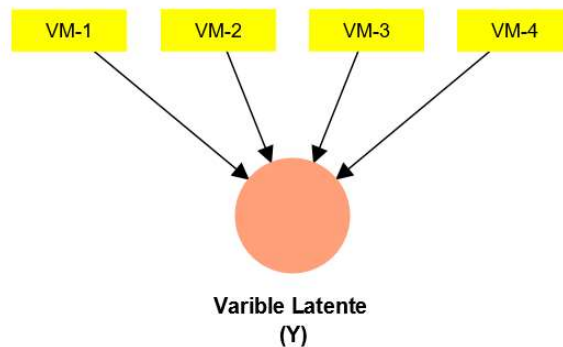


**Figura 10.** *Escala de Medición Reflexiva: La variable latente (Y) se mide mediante cuatro variables de medición VM-1, VM-2, VM-3 y VM-4 de forma reflexiva.*

Fuente: Hair Jr et al. (2016).

#### 4.19 Escala de medición reflexiva

Como se muestra en la Figura 10, en el caso de constructos medidos reflexivos, las variables de medición (indicadores) están altamente correlacionadas y son intercambiables. En otras palabras, se espera que las medidas tengan altas correlaciones. En este caso, se debe evaluar cuidadosamente la validez y confiabilidad de los indicadores (Haenlein & Kaplan, 2004; Hair et al., 2013).



**Figura 11.** *Escala de Medición Formativa: La variable latente (Y) se mide mediante cuatro variables de medición VM-1, VM-2, VM-3 y VM-4 de manera formativa.*

Fuente: Hair Jr et al. (2016).

En una escala de medición reflexiva, la dirección de la flecha va desde la variable latente hasta cada uno de los indicadores. Además, los cambios en la variable latente provocan directamente cambios en los indicadores asignados.

#### **4.20 Escala de medición formativa**

Por otro lado, como se ve en la Figura 11, no se espera que las medidas formativas se correlacionen. Se cree que los indicadores formativos han “formado” la variable latente. En el caso de constructos medidos formativamente, los indicadores pueden tener correlación negativa, positiva o incluso ninguna correlación entre ellos (Haenlein & Kaplan, 2004). En una escala de medición formativa, la dirección de la flecha va desde los indicadores formativos hasta la variable latente. Además, los cambios en uno o más de los indicadores provocan cambios en la variable latente.

Es importante mencionar que el modelo que se está utilizando en este estudio solo contiene constructos medidos reflexivamente (ver Figura 9).

#### **4.21 Evaluación del resultado PLS-SEM del modelo**

De acuerdo con Sarstedt et al. (2014), evaluar PLS-SEM implica realizar dos etapas. Como se ve en la Figura 12, la primera etapa (Etapa A) investiga el modelo de medición y examina la teoría de la medición. Esta investigación depende de si el modelo contiene

medidas reflexivas (ver Etapa A.1), medidas formativas (ver Etapa A.2) o ambas. Una vez que la investigación del modelo de medición concluye con resultados aceptables, comienza la Etapa B en la que implica investigar el modelo estructural y examinar la teoría estructural para decidir si las relaciones estructurales están teniendo un efecto importante y son relevantes de mencionarlas.

#### **4.22 Consideraciones éticas**

##### ***Descripción del instrumento, experimento o investigación***

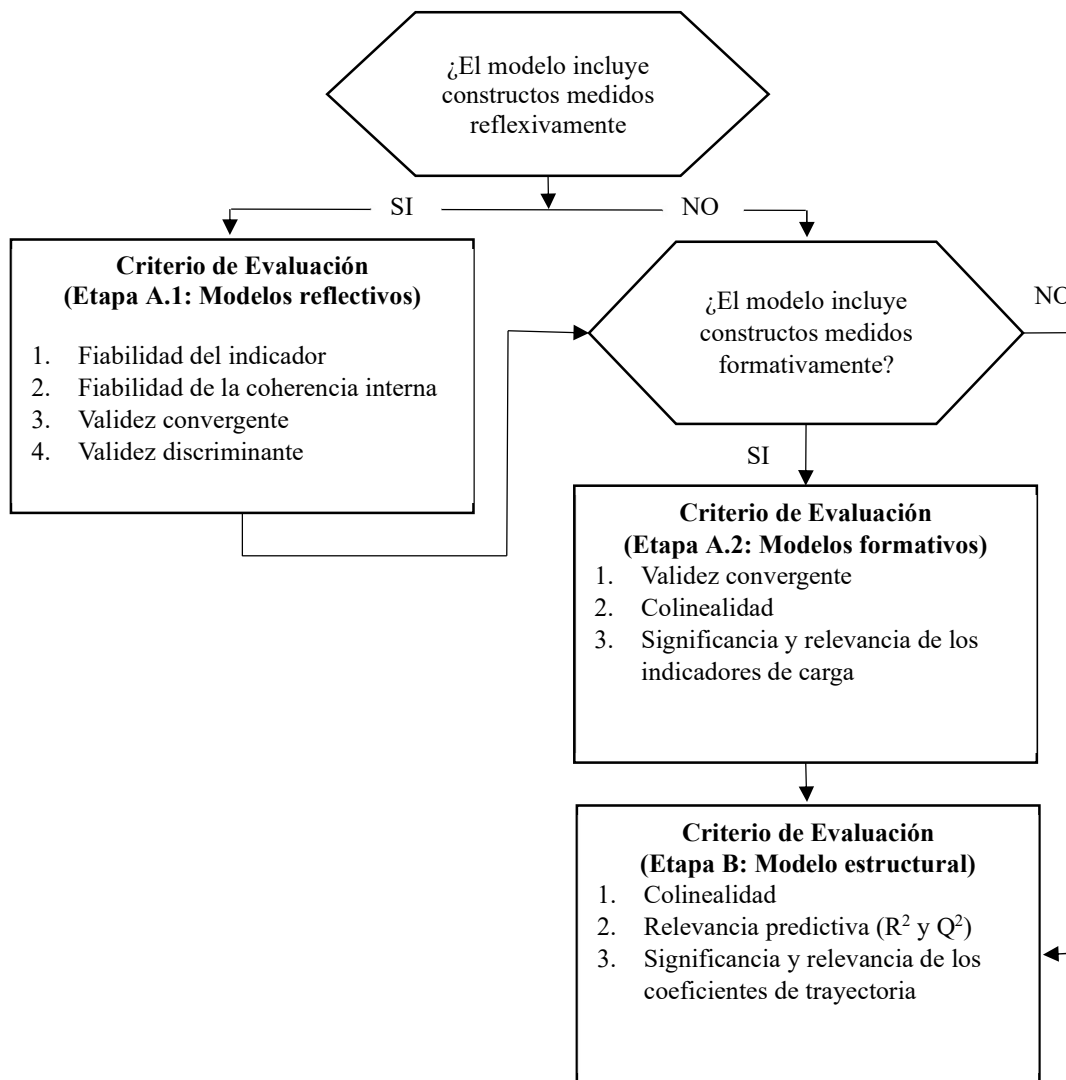
En este proyecto de investigación, la parte experimental que se tiene considerada es que los alumnos interactúen en la plataforma Moodle, para responder a un sistema de evaluación automática. Al finalizar el examen, se les pedirá que contesten una encuesta para medir la intención de utilizar dicho sistema.

##### ***Aspectos éticos que se toman en cuenta***

Realizar experimentos de la manera correcta es la clave para una investigación sobresaliente. La ciencia gira en torno a la experimentación, aprender a estructurar el diseño de un experimento es crucial para obtener datos útiles y resultados válidos a través de un experimento. Un investigador intenta aprender algo nuevo sobre el mundo y explica por qué sucede algo, sin embargo, al realizar experimentos involucrando a seres humanos es deber del investigador dirigir el estudio profesionalmente y lo más importante de una manera ética.

La investigación con seres humanos se han llenado de una historia de escándalos que a menudo dan forma a la manera en que la gente ve la ética en la investigación, la cuestión más apremiante en esta ética no es si debemos hacer investigación, pero podemos justificar la exposición de individuos al riesgo para el avance de la ciencia mediante la búsqueda de equilibrio entre el riesgo y la adquisición de conocimientos, a pesar de la mejores intenciones y cuidado la planificación y práctica de los experimentos. A veces las cosas van mal de vez en cuando, pueden surgir contratiempos técnicos, errores o falta de sensibilidad ética.





**Figura 12.** *Etapas de evaluación de PLS-SEM.*

Fuente: *Sarstedt et al. (2014).*

La ética que debe tenerse en cuenta al hacer investigación con seres humanos es presentada mediante la introducción del principio de consideraciones éticas encontradas en códigos internacionales de ética que guían la investigación humana. Se debe tener en cuenta que algunos países tienen normas en materia de experimentos con humanos,

pero en casos especiales como entrevistas o encuestas el investigador podría estar libre de regulación por las autoridades federales.

En este contexto, el protocolo de ética en la investigación de la institución educativa donde se llevó a cabo este estudio aborda los siguientes aspectos éticos:

1) Respeto. El respeto a las personas incorpora al menos dos convicciones éticas: en primer lugar, que las personas deben ser tratadas como agentes autónomos, y en segundo lugar, que las personas con una autonomía reducida tienen derecho a protección. Así pues, el principio del respeto de las personas se divide en dos exigencias morales distintas: la exigencia de reconocer la autonomía y la exigencia de proteger a los que tienen una autonomía reducida (Comisión Nacional, 1979).

2) Protección a las personas con capacidades disminuidas. En este aspecto, debemos estar muy atentos a las capacidades tanto física y cognitivas de los estudiantes que vamos a involucrar. Para garantizar su protección y respeto, lo primero que debemos hacer es informar sobre los objetivos de nuestro proyecto de investigación y su correspondiente experimentación para contar con su consentimiento de participación.

3) Beneficencia. Las personas son tratadas de manera ética no sólo respetando sus decisiones y protegiéndolas de todo daño, sino también esforzándose por asegurar su bienestar. Este trato se rige por el principio de beneficencia. El término 'beneficencia' se entiende a menudo para cubrir actos de bondad o caridad que van más allá de la obligación estricta. En este documento, beneficencia se entiende en un sentido más fuerte, como una obligación. Se han formulado dos reglas generales como expresión complementaria de acciones benéficas en este sentido: (1) no dañar y (2) maximizar los posibles beneficios y minimizar los posibles daños (Comisión Nacional 1979).

4) Justicia. ¿Quién debe recibir los beneficios de la investigación y soportar sus cargas? Esta es una cuestión de justicia, en el sentido de "equidad en la distribución" o "lo que se merece." Se comete una injusticia cuando se deniega algún beneficio al que tiene derecho una persona sin una buena razón o cuando se impone indebidamente alguna carga (Comisión Nacional 1979). Para esta consideración ética, lo más importante

sigue siendo informar de manera anticipada a los alumnos los objetivos del experimento y hacer una asignación aleatoria a los grupos diseñados para realizar el estudio.

5) Privacidad. En este aspecto, los elementos que debemos estar atentos cuidar son: no caer en una invasión de privacidad, cuidar la confidencialidad de los datos manejados y no generar un procedimiento del estudio que ponga en riesgo a los estudiantes. Debemos mantener el control sobre la extensión, el tiempo y las circunstancias de compartir los datos (físicos, conductuales, intelectuales) con otros.

En nuestro estudio, no se tienen contemplados menores de edad. No se harán preguntas intrusivas.

6) Confidencialidad. Por la naturaleza de nuestro estudio, los datos a recolectar son exclusivamente académicos. Todos los datos serán manejados en equipo de cómputo protegidos con contraseñas. El tratamiento de la información personal recopilada en una relación de confianza y con la expectativa de que no se haga pública de manera que contradiga el consentimiento o permiso otorgado por los participantes.

Los datos serán anonimizados desde el inicio del proceso de recogida, asignando códigos identificativos a cada participante, de modo que la información personal no pueda ser vinculada directamente a los datos recabados. Además, todos los datos serán almacenados en una base de datos segura, con acceso restringido solo al equipo de investigación. Estas medidas están diseñadas para asegurar que la información personal y los resultados de la investigación no sean divulgados de manera inapropiada (Sieber, 1992).

7) Procedimientos. La mayoría de la gente espera ser tratada con respeto y como seres autónomos. También esperan tomar decisiones por sí mismos y sobre qué información personal compartirán con otros. Por estas razones, la comunidad de investigación educativa está en consenso sobre la importancia del consentimiento informado durante la investigación. El proceso de consentimiento consiste en obtener el consentimiento informado del sujeto o del representante de los sujetos. El proceso de consentimiento implica lo siguiente:

- Proporcionar información específica a los sujetos sobre el estudio.
- Responder preguntas para asegurarse de que los sujetos entienden razonablemente la investigación y su papel en ella.
- Dar a los sujetos tiempo suficiente para considerar su respuesta.
- Obtención del acuerdo voluntario. Este documento es sólo para empezar el estudio, ya que el sujeto puede retirarse en un momento posterior, negarse a responder preguntas o completar tareas específicas durante la investigación.

8) Incentivos. En este momento no tenemos considerado algún tipo de incentivo en nuestro proyecto de investigación.

#### **4.23 Conclusiones**

En este capítulo se detalla el diseño metodológico del estudio, desde la filosofía de investigación hasta las técnicas de recolección y análisis de datos, justificando su pertinencia para los objetivos del estudio. Describe cómo los métodos cuantitativos aplicados se alinean con la pregunta de investigación y aborda las consideraciones éticas de la recolección de datos. En resumen, el capítulo establece un marco metodológico sólido que guía hacia resultados válidos y significativos para el campo de la innovación en evaluación educativa.

## **CAPÍTULO 5. RESULTADOS**

El propósito general de este estudio fue analizar el uso y la aceptación de una herramienta de evaluación automática de conceptos básicos de física basados en preguntas de respuesta corta, aplicando el Modelo de Aceptación de Evaluación Basado en Computadora (CBAAM, Terzis y Economides, 2011) en estudiantes de una universidad privada de Puebla. Este estudio sirvió como una comparación de dos instrumentos de evaluación automática sobre conceptos básicos de física en estudiantes universitarios: una evaluación basada en preguntas de opción múltiple (grupo control) en contraste con una evaluación basada en preguntas de respuesta corta (grupo experimental).

En este capítulo se presentan por separado los resultados del grupo experimental y de control. Como se mencionó en el capítulo anterior, evaluar PLS-SEM implica realizar dos etapas, la primera etapa (Etapa A) investiga el modelo de medición y examina la teoría de la medición. Esta investigación depende de si el modelo contiene medidas reflexivas (ver Etapa A.1), medidas formativas (ver Etapa A.2) o ambas. Una vez que la investigación del modelo de medición concluye con resultados aceptables, comienza la Etapa B en la que implica investigar el modelo estructural y examinar la teoría estructural para verificar si las relaciones estructurales están teniendo un efecto importante y son relevantes de mencionarlas.

El estudio se llevó a cabo en una universidad privada de la ciudad de Puebla, durante el semestre de otoño de 2022. Se analizaron los datos obtenidos por medio de los instrumentos de investigación que respondieron un total 84 estudiantes. Al grupo experimental (n=41) se le aplicó el instrumento de evaluación de conceptos de física con las 27 preguntas en formato de respuesta corta y que serán evaluadas de forma automática en la plataforma MOODLE. Al grupo de control (n=43) se le aplicó exactamente el mismo instrumento, pero con el formato de preguntas de opción múltiple. Posteriormente, se les pidió a los 2 grupos que contestaran las 30 preguntas correspondiente al instrumento de Modelo de Aceptación de Evaluación Basado en Computadora (CBAAM), que ya estaba disponible en también en la plataforma Moodle.

Este capítulo inicia con el análisis del modelo de medición (etapa A), donde se verifica la fiabilidad y validez de los constructos, posteriormente se presenta el análisis del modelo estructural (etapa B), evaluando la influencia y significancia de las relaciones entre constructos.

## **5.1 Modelo estructural**

En esta sección se detalla el análisis del modelo estructural y de medición aplicado al CBAAM (Terzis y Economides, 2011). Para profundizar en el análisis de los constructos que afectan la intención de utilizar la CBA dentro de un contexto educativo, se utilizó el software SmartPLS 4.0. Este programa permite una modelización de ecuaciones estructurales mediante el método de mínimos cuadrados parciales (PLS), lo cual es esencial para evaluar tanto la validez del modelo de medición como la fuerza y significancia de las relaciones hipotetizadas en el modelo estructural. La elección de SmartPLS 4.0 responde a su reconocida eficacia para el análisis de modelos complejos con variables latentes, facilitando así una interpretación clara y precisa de los datos.

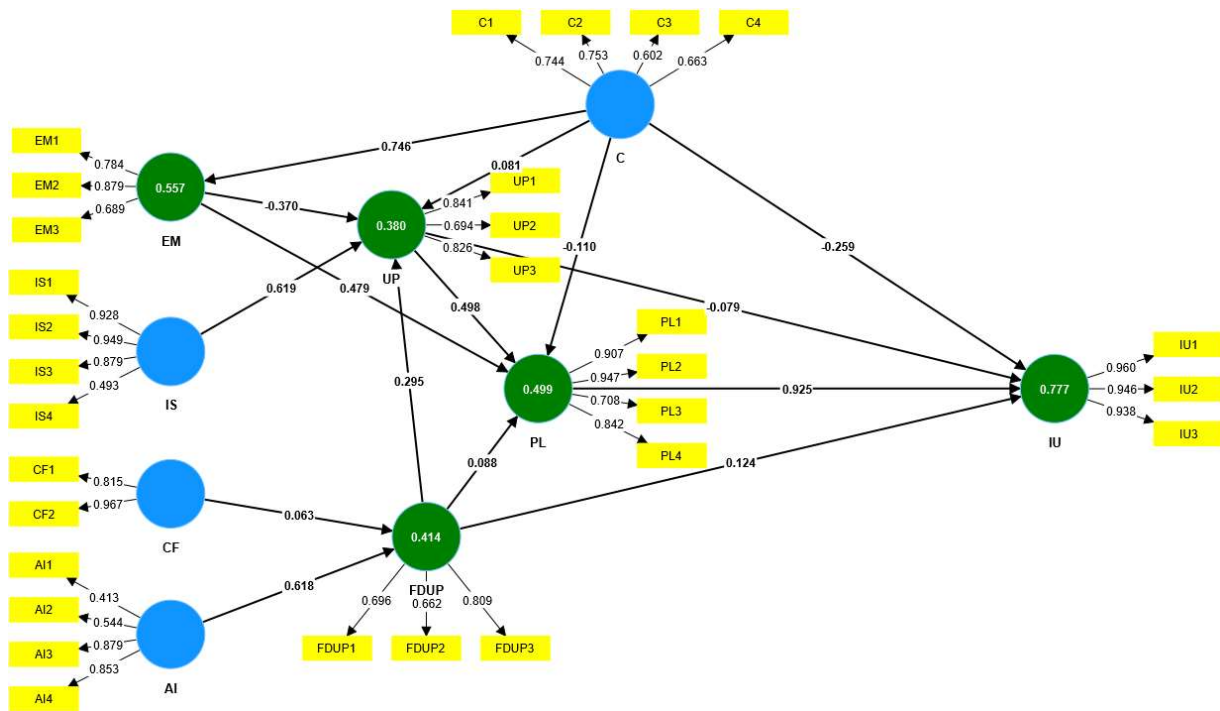
A continuación, se detalla cómo a través de este enfoque metodológico, se obtuvieron resultados relevantes sobre la aceptación del sistema de evaluación automática de conceptos básicos de física basados en preguntas de respuesta corta, en contraste con el formato de opción múltiple, contribuyendo significativamente al cuerpo de conocimiento existente en el campo de la evaluación electrónica.

El modelo estructural y de medición para los grupos experimental y de control se presentan en las Figuras 13 y 14, respectivamente.

En el capítulo anterior, se mencionó que las cuatro variables exógenas son:

- Autoeficacia informática (AI)
- Influencia Social (IS)
- Condiciones Facilitadoras (CF)

- Contenido (C)



**Figura 13.** Resultados PLS-SEM del modelo analizado para el grupo experimental. Las variables latentes se muestran en círculos y los moderadores (variables de medición) se muestran en cuadros amarillos.

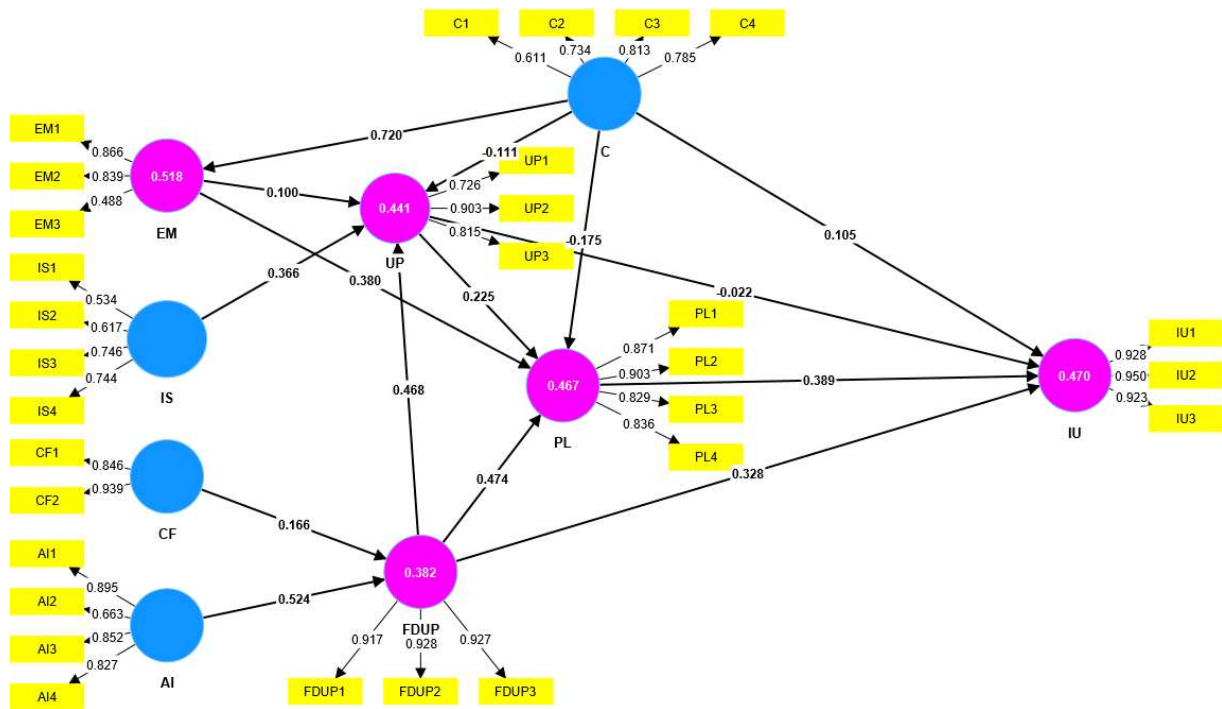
Fuente: elaboración propia.

Asimismo, se especificaron las cinco variables endógenas:

- Facilidad de uso percibida (FDUP)
- Percepción Lúdica (PL)
- Utilidad Percibida (PU)
- Expectativa de Meta (EM)
- Intención de Uso (IU)

Como se muestra en la Figura 13 y 14, cada una de estas variables endógenas tiene un valor en el medio del círculo. Este valor es el coeficiente de determinación,  $R^2$ . Este valor

muestra en qué medida (en porcentaje) las variables latentes exógenas explican esta variable endógena.



**Figura 14.** Resultados PLS-SEM del modelo analizado para el grupo control. Las variables latentes se muestran en círculos y los moderadores (variables de medición) se muestran en cuadros amarillos.

Fuente: elaboración propia.

## 5.2 Etapa A: Evaluación del modelo reflexivo

El modelo de medición que se muestra en la Figura 13 y 14 se basa en 30 elementos/indicadores reflectivos. Para analizar este modelo, el estudio utilizó el software SmartPLS 4.0 (Ringle, Wedne & Becker, 2015). El análisis del modelo reflexivo se realizó mediante la aplicación de lo siguiente:

1. Cargas del modelo exterior y fiabilidad del indicador.
2. Fiabilidad de la consistencia interna
3. Validez convergente
4. Validez discriminante



A continuación, se muestra cómo se ha realizado cada una de las pruebas enumeradas anteriormente:

### 5.2.1 Primero. Cargas del modelo exterior y fiabilidad del indicador

Las cargas del modelo externo o cargas de medición son los pesos de las trayectorias que conectan las variables latentes con los indicadores. Estos valores se utilizan para ver las correlaciones entre las VL y los indicadores. Las cargas del modelo exterior oscilan entre 0.0 y 1.0. Las cargas calculadas del modelo externo se utilizan para calcular la confiabilidad del indicador, que es igual al cuadrado de la carga de medición (Hair Jr et al., 2016). Como regla general, cuanto mayor sean las cargas, mejor y más fiable será el modelo exterior. Más concretamente, se prefiere tener valores iguales o superiores a 0.70 (Hulland, 1999). Las Tablas 6 y 7 muestran las cargas externas del modelo aplicado en este estudio:

**Tabla 6.** Cargas de modelo externo (grupo experimental)

	AI	C	CF	EM	FDUP	IS	IU	PL	UP
A11	0.413								
A12	0.544								
A13	0.879								
A14	0.853								
C1		0.744							
C2		0.753							
C3		0.602							
C4		0.663							
CF1			0.815						
CF2			0.967						
EM1				0.784					
EM2				0.879					
EM3				0.689					
FDUP1					0.696				
FDUP2					0.662				
FDUP3					0.809				
IS1						0.928			
IS2						0.949			
IS3						0.879			
IS4						0.493			
IU1							0.960		
IU2							0.946		
IU3							0.938		
PL1								0.907	
PL2								0.947	
PL3								0.708	
PL4								0.842	
UP1									0.841
UP2									0.694
UP3									0.826

Fuente: elaboración propia.

**Tabla 7. Cargas de modelo externo (grupo control)**

	AI	C	CF	EM	FDUP	IS	IU	PL	UP
AI1	0.895								
AI2	0.663								
AI3	0.852								
AI4	0.827								
C1		0.611							
C2		0.734							
C3		0.813							
C4		0.785							
CF1			0.846						
CF2			0.939						
EM1				0.866					
EM2				0.839					
EM3				0.488					
FDUP1					0.917				
FDUP2					0.928				
FDUP3					0.927				
IS1						0.534			
IS2						0.617			
IS3						0.746			
IS4						0.744			
IU1							0.928		
IU2							0.950		
IU3							0.923		
PL1								0.871	
PL2								0.903	
PL3								0.829	
PL4								0.836	
UP1									0.726
UP2									0.903
UP3									0.815

Fuente: *elaboración propia.*

Adicionalmente, las Tablas 8 y 9 muestran que la mayoría de las cargas están por encima de 0.7. Esto indica que cada una de las VL explica más del 50% de la varianza de su indicador (Henseler, Ringle & Sarstedt, 2012, p. 269). El valor (0.7) es el nivel en el que la varianza explicada es mayor que la varianza del error. Como regla general, en un modelo reflexivo, las cargas de trayectoria deben estar por encima de 0.70 (Henseler et al., 2012, p. 269).

### ***Interpretación de los resultados sobre las cargas del modelo exterior y fiabilidad del indicador***

Aunque el modelo CBAAM previamente validado se aplicó a un nuevo contexto, algunos

**Tabla 8. Confiabilidad de constructo y validez del modelo (grupo experimental)**

Variable Latente	Media	Desviación estándar	Indicador	Carga Factorial	Alpha de Cronbach	Confiabilidad compuesta	Varianza promedio extraída
Autoeficacia informática (AI)	6.679	0.486	AI1	0.413	0.667	0.781	0.492
			AI2	0.544			
			AI3	0.879			
			AI4	0.853			
Contenido (C)	6.115	0.767	C1	0.744	0.648	0.786	0.481
			C2	0.753			
			C3	0.602			
			C4	0.663			
Condiciones Facilitadoras (CF)	5.833	1.216	CF1	0.815	0.780	0.888	0.799
			CF2	0.967			
Expectativa de Meta (EM)	5.692	1.093	EM1	0.784	0.692	0.829	0.620
			EM2	0.879			
			EM3	0.689			
Facilidad de Uso Percibida (FDUP)	6.453	0.633	FDUP1	0.696	0.574	0.768	0.526
			FDUP2	0.662			
			FDUP3	0.809			
Influencia Social (IS)	5.256	1.481	IS1	0.928	0.865	0.896	0.695
			IS2	0.949			
			IS3	0.879			
			IS4	0.493			
Intención de Uso (IU)	5.966	1.438	IU1	0.960	0.943	0.964	0.898
			IU2	0.946			
			IU3	0.938			
Percepción Lúdica (PL)	5.788	1.151	PL1	0.907	0.875	0.915	0.732
			PL2	0.947			
			PL3	0.708			
			PL4	0.842			
Utilidad Percibida (UP)	5.983	0.888	UP1	0.841	0.721	0.831	0.623
			UP2	0.694			
			UP3	0.826			

Fuente: elaboración propia.

**Tabla 9. Confiabilidad de constructo y validez del modelo (grupo control)**

Variable Latente	Media	Desviación estándar	Indicador	Carga Factorial	Alpha de Cronbach	Confiabilidad compuesta	Varianza promedio extraída
Autoeficacia informática (AI)	6.349	1.018	AI1	0.895	0.843	0.886	0.663
			AI2	0.663			
			AI3	0.852			
			AI4	0.827			
Contenido (C)	5.337	1.046	C1	0.611	0.724	0.827	0.547
			C2	0.734			
			C3	0.813			
			C4	0.785			
Condiciones Facilitadoras (CF)	5.640	1.172	CF1	0.846	0.759	0.888	0.799
			CF2	0.939			
Expectativa de Meta (EM)	4.698	1.193	EM1	0.866	0.630	0.786	0.564
			EM2	0.839			
			EM3	0.488			
Facilidad de Uso Percibida (FDUP)	6.054	1.188	FDUP1	0.917	0.915	0.946	0.854
			FDUP2	0.928			
			FDUP3	0.927			
Influencia Social (IS)	5.343	0.926	IS1	0.534	0.659	0.758	0.444
			IS2	0.617			
			IS3	0.746			
			IS4	0.744			
Intención de Uso (IU)	6.039	1.003	IU1	0.928	0.927	0.953	0.872
			IU2	0.950			
			IU3	0.923			
Percepción Lúdica (PL)	5.506	1.002	PL1	0.871	0.885	0.919	0.740
			PL2	0.903			
			PL3	0.829			
			PL4	0.836			
Utilidad Percibida (UP)	5.829	0.736	UP1	0.726	0.749	0.857	0.669
			UP2	0.903			
			UP3	0.815			

Fuente: *elaboración propia.*

ítems presentan cargas bajas en ambos grupos, lo que sugiere la necesidad de revisar estos ítems para mejorar la adaptabilidad del modelo. Los constructos que mostraron altos valores de carga en ambos grupos (como "Condiciones Facilitadoras", "Facilidad de Uso Percibida", "Intención de Uso" y "Percepción Lúdica") confirman su validez en el nuevo contexto de investigación.

La mayoría de los constructos mantienen una fiabilidad y validez aceptables en ambos grupos, con algunos constructos destacando por su excelente desempeño ("Condiciones Facilitadoras", "Facilidad de Uso Percibida", "Intención de Uso"). Algunos constructos, como "Facilidad de Uso Percibida" en el grupo experimental y "Influencia Social" en ambos grupos, requieren una revisión más detallada debido a sus métricas más bajas.

### **5.2.2 Segundo. Fiabilidad de consistencia interna**

Después de verificar las cargas de los indicadores y asegurarse de que cada una de las variables latentes en el modelo explicara más de la mitad de la varianza de sus indicadores, se realizó una evaluación de la confiabilidad de la consistencia interna de las variables latentes. Por lo general, la confiabilidad de la consistencia interna se evalúa utilizando el "alfa de Cronbach". Sin embargo, esta medida tiende a proporcionar una medición conservadora en PLS-SEM (Garson, 2016). La literatura también recomienda el uso de la "confiabilidad compuesta" para verificar la confiabilidad de la consistencia interna y es considerada una mejor medida alternativa (Bagozzi & Yi, 1988; Hair et al., 2012).

Como regla general, cuantos mayores sean los valores de la confiabilidad compuesta, mejores y mayores serán los niveles de confiabilidad. Según Bagozzi y Yi (1988), valores mayores o iguales a 0.6 y menores de 0.7 se consideran "aceptables", mientras que valores mayores o iguales a 0.7 y menores de 0.95 se consideran "satisfactorios a buenos". Además, los valores superiores a 0.95 se consideran problemáticos. Estos valores (los superiores a 0.95) pueden indicar nuestra intención de que los indicadores

múltiples sean redundantes en lugar de ser medidas verdaderamente demostrativas de la variable latente (Hair Jr et al., 2016). La séptima columna de las Tablas 8 y 9 muestran los valores de confiabilidad compuesta.

### ***Interpretación de los resultados sobre las cargas del modelo exterior***

Se han obtenido niveles aceptables de confiabilidad de consistencia interna entre las nueve variables latentes reflexivas. Por tanto, se tienen evidencias de la fiabilidad de la consistencia interna en ambos grupos, lo que indica que el modelo CBAAM previamente validado se adapta bien a estos constructos en el nuevo contexto de esta investigación.

### **5.2.3 Tercero. Validez convergente**

El siguiente paso después de evaluar la confiabilidad de la consistencia interna implicó la evaluación de la validez convergente de las variables latentes medidas reflexivamente. Como afirman Bagozzi y Yi (1988) y Garson (2016), la validez convergente es el grado en que los indicadores utilizados para medir una misma variable latente concuerdan. Bagozzi y Yi (1988) sugieren que se puede utilizar la varianza promedio extraída (AVE) para evaluar la validez de la convergencia. La varianza promedio extraída (AVE) se calcula como la media de las cargas al cuadrado para todos los indicadores relacionados con la variable latente. El umbral AVE aceptable es 0.5 (Bagozzi & Yi, 1988).

Siguiendo esta recomendación, se calculó y evaluó el AVE de cada variable latente. Después de analizar las cifras calculadas que se muestran en las Tablas 8 y 9, se encontró que todos los valores de AVE son mayores que 0.50, (a excepción de tres constructo) que es el valor mínimo aceptable para la validez convergente (Bagozzi & Yi, 1988; Chin, 1998). Por tanto, se tienen evidencias de la validez convergente. Esto significa que, en promedio, cada una de las variables latentes explica más de la mitad de la varianza de sus indicadores.

### ***Interpretación de los resultados sobre la validez convergente***

Los constructos Condiciones Facilitadoras (CF), Facilidad de Uso Percibida (FDUP), Intención de Uso (IU) y Percepción Lúdica (PL) mantienen una alta validez convergente en ambos grupos, indicando que el modelo CBAAM previamente validado se adapta bien a estos constructos en el nuevo contexto de esta investigación. En ambos grupos, los constructos Autoeficacia Informática (AI) y Contenido (C) en el grupo experimental y Influencia Social (IS) en el grupo control muestran una validez convergente apenas aceptable, lo que sugiere que estos constructos pueden requerir ajustes adicionales para mejorar su validez en contextos similares.

#### **5.2.4 Cuatro. Validez discriminante**

La valoración de la validez discriminante tiene como objetivo asegurar que una variable latente reflexiva tenga la asociación más fuerte con sus propios indicadores. Según Sarstedt et al. (2014), la validez discriminante examina el grado en que una variable latente es empíricamente diferente de otras variables latentes en el modelo. Esto se determina mediante: (1) el cálculo de cuánto se correlaciona la variable latente con otras variables latentes; y (2) determinar en qué medida los indicadores representan sólo esta única variable latente.

El criterio de Fornell y Larcker (1981) es una de las medidas más importantes para probar la validez discriminante. En este método, se evalúa la raíz cuadrada de AVE para cada variable latente. La pauta recomendada es que la raíz cuadrada del AVE para cada variable latente debe ser mayor que su correlación con cualquier otra variable latente (Garson, 2016). El texto en negrita en la diagonal de las Tablas 10 y 11 representa la raíz cuadrada del AVE de cada una de las variables latentes. Por ejemplo, se encuentra que el AVE de la variable latente CF es 0.799 (de la tabla 8). Por tanto, su raíz cuadrada pasa a ser 0.894. Este valor es mayor que los valores de correlación en la columna de CF (0.427, 0.285, 0.523, 0.768, 0.721, 0.257) y también mayor que los valores en la fila de CF (0.359, 0.163) (ver tabla 10). También se realiza un análisis similar para las otras variables latentes (AI, C, EM, FDUP, IS, IU, PL y UP).

**Tabla 10.** Validez discriminante utilizando el criterio de Fornell-Lacker (grupo experimental)

	AI	C	CF	EM	FDUP	IS	IU	PL	UP
AI	<b>0.701</b>								
C	0.190	0.693							
CF	0.359	0.163	<b>0.894</b>						
EM	0.224	<b>0.746</b>	0.427	<b>0.788</b>					
FDUP	0.641	0.300	0.285	0.473	<b>0.725</b>				
IS	0.178	0.358	0.523	0.406	0.136	<b>0.833</b>			
IU	0.402	0.075	0.768	0.302	0.394	0.624	<b>0.948</b>		
PL	0.398	0.331	0.721	0.479	0.396	0.591	0.845	<b>0.856</b>	
UP	0.069	0.115	0.257	0.081	0.229	0.538	0.423	0.545	<b>0.789</b>

Fuente: elaboración propia.

**Tabla 11.** Validez discriminante utilizando el criterio de Fornell-Lacker (grupo control)

	AI	C	CF	EM	FDUP	IS	IU	PL	UP
AI	<b>0.814</b>								
C	0.174	<b>0.740</b>							
CF	0.465	0.427	<b>0.894</b>						
EM	-0.035	0.720	0.402	<b>0.751</b>					
FDUP	0.601	0.536	0.409	0.171	<b>0.924</b>				
IS	0.336	0.467	0.336	0.565	0.342	<b>0.666</b>			
IU	0.339	0.444	0.275	0.239	0.594	0.383	<b>0.934</b>		
PL	0.234	0.439	0.202	0.404	0.569	0.478	0.610	<b>0.860</b>	
UP	0.386	0.383	0.430	0.307	0.551	0.531	0.408	0.536	<b>0.818</b>

Fuente: elaboración propia.

El resultado indica que debido a que la raíz cuadrada de AVE (resaltada en negrita) en todas las columnas de variables latentes es mayor que las correlaciones (los números debajo), la validez discriminante está bien establecida.

Otro enfoque que implementaron los investigadores para evaluar la validez discriminante es evaluar las cargas cruzadas. Según Garson (2016) y Hair Jr et al. (2016), las cargas cruzadas son una alternativa al AVE como técnica de valoración de la validez discriminante para modelos reflexivos. La pauta recomendada es que, como mínimo, ninguna variable indicadora debe tener una correlación mayor con otra variable latente que con su propia variable latente (Garson, 2016; Hair Jr et al., 2016). De lo contrario no se establece la validez discriminante. Las Tablas 12 y 13 muestran que esta condición



se cumple. Cada una de las variables indicadoras exhibe una mayor carga sobre su variable latente que sobre cualquier otra variable latente incluida en el modelo.

**Tabla 12.** Validez discriminante usando cargas cruzadas (grupo experimental)

	AI	C	CF	EM	FDUP	IS	IU	PL	UP
AI1	0.413	0.009	0.413	0.076	0.201	0.386	0.525	0.565	0.454
AI2	0.544	-0.039	0.182	0.053	0.250	0.094	0.063	0.115	0.120
AI3	0.879	0.221	0.301	0.240	0.649	0.165	0.371	0.339	0.015
AI4	0.853	0.190	0.227	0.176	0.512	0.019	0.257	0.247	-0.089
C1	0.281	0.744	0.206	0.505	0.337	0.130	0.180	0.322	0.083
C2	0.165	0.753	0.043	0.645	0.272	0.286	0.006	0.212	0.159
C3	0.033	0.602	-0.079	0.371	0.120	0.439	-0.062	0.033	0.177
C4	0.000	0.663	0.230	0.500	0.057	0.221	0.046	0.292	-0.081
CF1	0.259	0.241	0.815	0.418	0.139	0.339	0.447	0.481	0.107
CF2	0.362	0.110	0.967	0.381	0.316	0.544	0.821	0.744	0.294
EM1	0.320	0.533	0.480	0.784	0.540	0.330	0.303	0.383	0.104
EM2	0.190	0.788	0.211	0.879	0.353	0.387	0.171	0.321	0.077
EM3	-0.003	0.381	0.366	0.689	0.222	0.222	0.271	0.468	0.000
FDUP1	0.261	0.075	0.144	0.339	0.696	-0.037	0.210	0.232	0.157
FDUP2	0.397	0.104	0.129	0.158	0.662	-0.057	0.195	0.219	0.308
FDUP3	0.633	0.375	0.298	0.481	0.809	0.276	0.396	0.372	0.082
IS1	0.218	0.262	0.301	0.252	0.166	0.928	0.564	0.512	0.581
IS2	0.124	0.289	0.567	0.390	0.093	0.949	0.587	0.593	0.487
IS3	0.129	0.483	0.640	0.534	0.100	0.879	0.572	0.533	0.385
IS4	0.273	0.279	0.566	0.388	0.277	0.493	0.561	0.394	0.009
IU1	0.361	0.082	0.663	0.247	0.353	0.571	0.960	0.825	0.449
IU2	0.439	0.024	0.737	0.287	0.432	0.604	0.946	0.826	0.434
IU3	0.338	0.113	0.789	0.330	0.328	0.601	0.938	0.744	0.310
PL1	0.398	0.403	0.650	0.526	0.370	0.629	0.830	0.907	0.403
PL2	0.435	0.224	0.723	0.375	0.373	0.582	0.874	0.947	0.485
PL3	0.314	0.402	0.396	0.414	0.335	0.261	0.416	0.708	0.393
PL4	0.207	0.149	0.647	0.337	0.284	0.482	0.686	0.842	0.589
UP1	0.050	-0.052	0.261	-0.059	0.149	0.362	0.346	0.512	0.841
UP2	0.074	0.297	0.047	0.122	0.315	0.107	0.060	0.281	0.694
UP3	0.054	0.140	0.224	0.151	0.166	0.624	0.448	0.440	0.826

Fuente: elaboración propia.

### **Interpretación de los resultados sobre la validez discriminante**

En general, tanto el grupo experimental como el grupo control muestran una validez

**Tabla 13.** Validez discriminante usando cargas cruzadas (grupo control)

	AI	C	CF	EM	FDUP	IS	IU	PL	UP
AI1	0.895	0.235	0.338	0.013	0.661	0.329	0.323	0.299	0.437
AI2	0.663	0.045	0.513	-0.020	0.153	0.228	0.093	-0.032	0.190
AI3	0.852	-0.007	0.343	-0.079	0.502	0.317	0.322	0.294	0.329
AI4	0.827	0.235	0.503	-0.041	0.398	0.188	0.257	-0.003	0.194
C1	0.446	0.611	0.271	0.176	0.741	0.221	0.447	0.453	0.443
C2	0.053	0.734	0.443	0.481	0.404	0.296	0.194	0.191	0.247
C3	-0.138	0.813	0.152	0.706	0.192	0.396	0.285	0.251	0.159
C4	0.175	0.785	0.411	0.687	0.319	0.432	0.369	0.384	0.300
CF1	0.352	0.232	0.846	0.251	0.276	0.251	0.153	0.104	0.277
CF2	0.464	0.484	0.939	0.435	0.430	0.337	0.309	0.233	0.460
EM1	0.184	0.734	0.349	0.866	0.253	0.527	0.309	0.373	0.238
EM2	-0.293	0.546	0.296	0.839	0.046	0.435	0.087	0.297	0.306
EM3	-0.018	0.146	0.297	0.488	0.000	0.253	0.086	0.232	0.117
FDUP1	0.511	0.538	0.345	0.229	0.917	0.360	0.503	0.551	0.538
FDUP2	0.532	0.542	0.419	0.182	0.928	0.368	0.696	0.546	0.496
FDUP3	0.627	0.399	0.367	0.058	0.927	0.213	0.429	0.476	0.495
IS1	0.218	0.205	0.300	0.161	0.161	0.534	0.107	0.167	0.115
IS2	0.090	0.311	0.278	0.354	0.005	0.617	0.080	0.080	0.231
IS3	0.095	0.292	0.067	0.544	0.146	0.746	0.169	0.442	0.374
IS4	0.412	0.391	0.328	0.357	0.438	0.744	0.469	0.413	0.496
IU1	0.279	0.418	0.210	0.248	0.538	0.388	0.928	0.623	0.374
IU2	0.352	0.382	0.290	0.164	0.586	0.297	0.950	0.561	0.318
IU3	0.321	0.444	0.272	0.261	0.539	0.390	0.923	0.521	0.455
PL1	0.151	0.422	0.152	0.430	0.452	0.368	0.433	0.871	0.482
PL2	0.113	0.328	0.065	0.435	0.390	0.508	0.500	0.903	0.462
PL3	0.135	0.220	0.078	0.155	0.435	0.271	0.324	0.829	0.332
PL4	0.344	0.470	0.330	0.325	0.626	0.447	0.727	0.836	0.516
UP1	0.395	0.254	0.506	0.268	0.410	0.460	0.179	0.373	0.726
UP2	0.300	0.363	0.330	0.325	0.427	0.441	0.314	0.488	0.903
UP3	0.269	0.316	0.253	0.171	0.507	0.410	0.475	0.445	0.815

Fuente: *elaboración propia.*

discriminante adecuada según el criterio de Fornell y Larcker. Sin embargo, algunos ítems presentan problemas en las cargas cruzadas, lo que sugiere que pueden necesitar ser revisados para mejorar la validez discriminante del modelo. Estos ajustes asegurarán que el modelo siga siendo robusto y aplicable en diferentes contextos educativos, proporcionando una representación precisa y diferenciada de los constructos evaluados.

### **5.3 Etapa B: Evaluación del modelo estructural**

Después de verificar la confiabilidad del indicador, la confiabilidad de la consistencia interna, la validez convergente y la validez discriminante del modelo de medición, las evidencias obtenidas sugieren que la calidad del modelo de medición es satisfactoria, tanto para el grupo experimental como el de control. En función de ello se continuó con la Etapa B la cual consistió en evaluar modelo estructural. Esto implicó realizar los siguientes pasos:

1. Probar el modelo estructural para detectar posibles problemas de colinealidad.
2. Evaluar la capacidad del modelo para predecir las variables latentes endógenas. Para facilitar esta evaluación, se utilizó: (a) el Coeficiente de determinación ( $R^2$ ); (b) Redundancia con validación cruzada ( $Q^2$ ).
3. Comprobación de la importancia del modelo estructural en el bootstrapping.

A continuación, se explican cada uno de los resultados obtenidos de los tres puntos enumerados anteriormente:

#### **5.3.1 Primero. Prueba del modelo estructural para detectar posibles problemas de colinealidad**

Según Garson (2016), el problema de multicolinealidad aparece cuando dos o más variables independientes están altamente intercorrelacionadas. Garson declaró que tener el problema de la multicolinealidad dificulta que el investigador evalúe la importancia relativa de una variable independiente comparada con otra. Por esta razón, el primer paso que se realizó para evaluar el modelo estructural fue examinar la colinealidad del modelo estructural. La razón es que la estimación de los coeficientes de trayectoria se basa en regresiones de mínimos cuadrados ordinarios (MCO) (Mooi & Sarstedt, 2011), y el problema de multicolinealidad en la regresión MCO infla los errores

estándar. Esto haría que las pruebas de significancia de las variables independientes no fueran confiables (Garson, 2016; Hair, Ringle & Sarstedt, 2011; Hair Jr et al., 2016).

Como pauta general, se considera que el modelo estructural tiene problemas de multicolinealidad cuando el coeficiente del Factor de Inflación de Varianza (VIF) es superior a 4.0 (Garson, 2016, pp. 71). En este trabajo se utilizó esta pauta general para examinar la colinealidad entre las variables latentes exógenas. Los resultados de esta prueba se muestran en las Tablas 14 y 15.

**Tabla 14.** Valores VIF del modelo interno (grupo experimental)

	AI	C	CF	EM	FDUP	IS	IU	PL	UP
AI					1.148				
C				1.000			1.177	2.302	2.294
CF					1.148				
EM								2.695	2.816
FDUP							1.234	1.368	1.303
IS									1.212
IU									
PL							1.707		
UP							1.434	1.068	

Fuente: elaboración propia.

**Tabla 15.** Valores VIF del modelo interno (grupo control)

	AI	C	CF	EM	FDUP	IS	IU	PL	UP
AI					1.276				
C				1.000			1.464	3.310	3.302
CF					1.276				
EM								2.563	3.014
FDUP							1.924	2.153	1.783
IS									1.633
IU									
PL							1.696		
UP							1.612	1.559	

Fuente: elaboración propia.

### **Interpretación de los resultados para problemas de colinealidad**

Como se muestra en las Tablas 14 y 15, los valores VIF resultantes de las variables exógenas que están directamente relacionadas con IU (C, FDUP, PL y UP) oscilaron entre 1.177 (C) y 1.707 (PL) (ver tabla 14), lo que significa que todos los coeficientes VIF son inferiores a 4.0. De manera similar, los valores VIF de las variables exógenas que están conectadas a las siguientes variables endógenas (C, IS, CF y AI) también son inferiores a 4.0. Esto confirma que no hay indicación de colinealidad entre cada conjunto de variables latentes exógenas. Por tanto, los resultados del modelo estructural no se ven afectados negativamente por la colinealidad tanto el grupo experimental como el grupo control.

### **5.3.2 Segundo. Evaluación de la capacidad del modelo para predecir las variables latentes endógenas**

Para evaluar la capacidad del modelo estructural con el objetivo de predecir constructos endógenos, se utilizaron los valores tanto del coeficiente de determinación ( $R^2$ ) como de la redundancia con validación cruzada ( $Q^2$ ). A continuación, se explican los resultados de cada una de estas dos evaluaciones:

#### ***Coefficiente de determinación ( $R^2$ )***

El valor  $R^2$  se utiliza para evaluar la precisión predictiva del modelo.  $R^2$  es una medida del tamaño del efecto general para el modelo estructural. Chin (1998, p. 323) explica que los resultados superiores al valor 0.67 son “sustanciales”, los valores superiores a 0,33 son “moderados” y los valores superiores a 0.19 son “débiles”. Los valores  $R^2$  calculados de las variables latentes endógenas del modelo utilizado en este estudio se muestran en las Tablas 16 y 17.

Como se observa en las Tablas 16 y 17 y basándose en la explicación de Chin (1998) de los valores  $R^2$ , se examinó el poder predictivo de las variables latentes endógenas. La intención de uso (IU), que es la principal medida de resultado del modelo, tiene un poder de predicción sustancial para el caso del grupo experimental ( $R^2 = 0.777$ ) y

moderado ( $R^2 = 0.470$ ) para el grupo control, mientras que el poder de predicción de EM, FDUP, PL y UP obtuvieron un poder de moderado.

**Tabla 16.** Valores  $R^2$  de las variables latentes endógenas (grupo experimental)

	R-square	Poder de predicción
EM	0.557	Moderado
FDUP	0.414	Moderado
IU	0.777	Sustancial
PL	0.499	Moderado
UP	0.380	Moderado

Fuente: *elaboración propia.*

**Tabla 17.** Valores  $R^2$  de las variables latentes endógenas (grupo control)

	R-cuadrada	Poder de predicción
EM	0.518	Moderado
FDUP	0.382	Moderado
IU	0.470	Moderado
PL	0.467	Moderado
UP	0.441	Moderado

Fuente: *elaboración propia.*

### ***Interpretación de los resultados sobre la capacidad del modelo para predecir las variables latentes endógenas***

El grupo experimental muestra un poder de predicción sustancial para IU y moderado para las demás variables, lo que indica que el modelo en este grupo es particularmente fuerte para predecir IU. En contraste, el grupo control presenta un poder de predicción moderado para todas las variables, sugiriendo una capacidad de predicción uniforme pero no tan fuerte como en el caso de IU en el grupo experimental. En resumen, el grupo experimental demuestra un mejor desempeño predictivo para IU en comparación con el grupo control, mientras que ambos grupos exhiben un poder de predicción moderado para las demás variables.

### **Redundancia con validación cruzada ( $Q^2$ )**

Se emplearon medidas de redundancia con validación cruzada ( $Q^2$ ), a este procedimiento también se le llama *blindfolding* (vendar los ojos), para evaluar la relevancia predictiva de todas las variables endógenas latentes en el modelo de este estudio. Como afirma Garson (2016), el *blindfolding* es un procedimiento de reutilización de muestras que se utiliza para calcular el valor  $Q^2$  (Geisser, 1974, p. 33). Este procedimiento sólo se aplica a las variables latentes medidas reflexivamente. Garson (2016) proporcionó una descripción más detallada de cómo se calculan los valores de  $Q^2$ :

“El *blindfolding* es una técnica de reutilización de muestras que comienza con el primer punto de datos y omite cada décimo punto de datos en los indicadores de la construcción endógena. Luego, el procedimiento estima los parámetros del modelo de ruta PLS utilizando los puntos de datos restantes. Los puntos de datos omitidos se consideran valores faltantes y se tratan en consecuencia cuando se ejecuta el algoritmo PLS-SEM (por ejemplo, mediante el reemplazo del valor medio). Las estimaciones resultantes se utilizan luego para predecir los puntos de datos omitidos. La diferencia entre los puntos de datos verdaderos (es decir, omitidos) y los pronosticados se utiliza luego como entrada para la medida  $Q^2$ ” (p. 116)

Es importante mencionar que SmartPLS 4.0 proporciona dos enfoques para calcular los valores de  $Q^2$ : el primero es la redundancia con validación cruzada; y el segundo es la comunalidad con validación cruzada. Se utilizó el enfoque de redundancia con validación cruzada recomendado por Hair Jr et al. (2016). Según Garson (2016) y Hair Jr et al. (2016), tener valores de  $Q^2$  mayores que cero para las variables latentes endógenas confirma la relevancia predictiva del modelo estructural. Como se presenta en la Tabla 18 y 19, realizar el procedimiento de *blindfolding* con una Distancia de Omisión (OD) de seis arrojó valores de  $Q^2$  que están por encima de cero. El valor recomendado de la Distancia de Omisión (OD) para la mayoría de los tipos de investigación es de 5 a 10 (Hair Jr, Sarstedt, Ringle, & Mena, 2012). Esto se observó en todas las variables latentes endógenas del grupo experimental (IU: 0.302, EM: 0.521, FDUP: 0.259, PL: 0.303 y

UP:0.090) (ver tabla 18). En el caso del grupo control, también se obtuvieron valores de  $Q^2$  mayores que cero (ver tabla 19). Por lo tanto, se confirma la relevancia predictiva del modelo para las variables latentes endógenas.

**Tabla 18.** Redundancia con validación cruzada de variables latentes ( $Q^2$ ) (grupo experimental)

	Q <sup>2</sup> -cuadrada
EM	0.521
FDUP	0.259
IU	0.302
PL	0.303
UP	0.090

Fuente: *elaboración propia*.

**Tabla 19.** Redundancia con validación cruzada de variables latentes ( $Q^2$ ) (grupo control)

	Q-cuadrada
EM	0.467
FDUP	0.275
IU	0.114
PL	0.049
UP	0.205

Fuente: *elaboración propia*.

### ***Interpretación de los resultados sobre la relevancia predictiva de modelo***

El grupo experimental demuestra un mejor desempeño predictivo en general, especialmente para las variables EM e IU, en comparación con el grupo control. El grupo control, aunque presenta un buen poder predictivo para EM y un poder predictivo moderado para FDUP y UP, muestra un desempeño inferior en IU y PL.

### **5.3.3 Tercero. Comprobar la significancia y relevancia de las rutas estructurales**

#### ***A. La significancia del modelo:***



El paso final del análisis del modelo estructural considera las fortalezas y la importancia de los coeficientes de trayectoria (las relaciones hipotéticas entre las variables latentes). SmartPLS 4.0 (Ringle et al., 2015) proporciona un procedimiento llamado bootstrapping que es una técnica utilizada para generar estadísticas t para probar la importancia de los modelos interno y externo. Este procedimiento utiliza métodos de remuestreo para calcular la importancia de los coeficientes PLS. En esta técnica, se toma una gran cantidad de submuestras de arranque (por ejemplo: 5000 submuestras) de la muestra inicial con reemplazo.

Esto se realiza para producir errores estándar de arranque, que luego generarán valores t aproximados que se utilizan para determinar la importancia de la ruta estructural. Siguiendo las recomendaciones de Hair Jr et al. (2016), el procedimiento de bootstrapping se realizó con un nivel de significancia del 5%. Los resultados de ejecutar este procedimiento se muestran en las Tablas 20 y 21. Como se muestra en estas tablas, los coeficientes de ruta del modelo estructural están resaltados en negrita y enumerados en la columna Estadísticas T.

Después de analizar los coeficientes de ruta presentados en la columna de estadísticos T, es evidente que algunos de estos coeficientes de ruta son significativos y otros no. Como regla general, todos los valores del estadístico t superiores a 1.65 son significativos con un nivel de significancia de 0.10, todos los valores del estadístico t superiores a 1.96 son significativos con un nivel de significancia de 0.05 y todos los valores del estadístico t superiores a 2.58 son significativos con un nivel de significancia de 0.01. nivel (Garson, 2016). Aplicando esta regla general para el caso del grupo experimental podemos concluir que:

El vínculo (AI -> FDUP) (4.905), el vínculo (C -> EM) (9.169), el vínculo (IS -> UP) (3.112), vínculo (PL -> IU) (6.154), el vínculo (UP -> PL) (2.846), son todos estadísticamente significativos al nivel de significancia de 0.01.

Además, el vínculo (C -> IU) (2.092) y el vínculo (EM -> PL) (2.021) son significativos al nivel de significancia de 0.05.

Finalmente, se comprueba estadísticamente que el vínculo (C -> PL) (0.504), el vínculo (C -> UP) (0.323), el vínculo (CF -> FDUP) (0.436), el vínculo (EM -> UP) (1.612), el vínculo (FDUP -> IU) (0.962), el vínculo (FDUP -> PL) (0.475), el vínculo (FDUP -> UP) (1.436), y el vínculo (UP -> IU) (0.588) no son estadísticamente significativos.

En resumen, el análisis del grupo experimental revela que las relaciones entre AI y FDUP, C y EM, IS y UP, PL y IU, y UP y PL son las más robustas y estadísticamente significativas, contribuyendo de manera importante al modelo estructural. Además, las relaciones entre C e IU, y EM y PL también son significativas, aunque con un nivel de certeza menor. Sin embargo, las demás relaciones no presentan una significancia estadística, indicando una menor influencia en la estructura general del modelo.

**Tabla 20.** Estadísticos T de coeficientes de trayectoria [modelo estructural (interno)] (grupo experimental)

	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics ( O/STDEV )	P values
AI -> FDUP	0.618	0.614	0.126	<b>4.905</b>	<b>0.000</b>
C -> EM	0.746	0.751	0.081	<b>9.169</b>	<b>0.000</b>
C -> IU	-0.259	-0.250	0.124	<b>2.092</b>	<b>0.037</b>
C -> PL	-0.110	-0.041	0.218	<b>0.504</b>	<b>0.614</b>
C -> UP	0.081	0.067	0.250	<b>0.323</b>	<b>0.747</b>
CF -> FDUP	0.063	0.110	0.144	<b>0.436</b>	<b>0.663</b>
EM -> PL	0.479	0.427	0.237	<b>2.021</b>	<b>0.043</b>
EM -> UP	-0.370	-0.316	0.229	<b>1.612</b>	<b>0.107</b>
FDUP -> IU	0.124	0.155	0.129	<b>0.962</b>	<b>0.336</b>
FDUP -> PL	0.088	0.104	0.186	<b>0.475</b>	<b>0.635</b>
FDUP -> UP	0.295	0.282	0.206	<b>1.436</b>	<b>0.151</b>
IS -> UP	0.619	0.580	0.199	<b>3.112</b>	<b>0.002</b>
PL -> IU	0.925	0.908	0.150	<b>6.154</b>	<b>0.000</b>
UP -> IU	-0.079	-0.082	0.135	<b>0.588</b>	<b>0.556</b>
UP -> PL	0.498	0.522	0.175	<b>2.846</b>	<b>0.004</b>

Fuente: *elaboración propia.*

Para el caso del grupo control se concluye que:

El vínculo (C -> EM) (10.537), el vínculo (FDUP -> PL) (2.953), el vínculo (IS -> UP) (2.920), vínculo (FDUP -> UP) (2.791), el vínculo (PL -> IU) (2.648), son todos estadísticamente significativos al nivel de significancia de 0.01.

Además, el vínculo (EM -> PL) (2.025) y el vínculo (FDUP -> IU) (2.261) son significativos al nivel de significancia de 0.05.

Asimismo, el vínculo (AI -> FDUP) (1.749) y el vínculo (UP -> PL) (1.764), son significativos con un nivel de significancia de 0.10.

Finalmente, se comprueba estadísticamente que el vínculo (C -> IU) (0.883), el vínculo (C -> PL) (0.716), el vínculo (C -> UP) (0.431), el vínculo (CF -> FDUP) (0.835), el vínculo (EM -> UP) (0.454) y el vínculo (UP -> IU) (0.134) no son estadísticamente significativos.

En resumen, el análisis del grupo control revela que las relaciones entre C y EM, FDUP y PL, IS y UP, FDUP y UP, y PL y IU son las más robustas y estadísticamente significativas, contribuyendo de manera importante al modelo estructural. Además, las relaciones entre EM y PL, y FDUP e IU también son significativas, aunque con un nivel de certeza menor. Sin embargo, las demás relaciones no presentan una significancia estadística, indicando una menor influencia en la estructura general del modelo.

### ***Interpretación de los resultados sobre la significancia y relevancia de las rutas estructurales***

En resumen, el grupo experimental demuestra un mejor desempeño predictivo en términos de trayectorias significativas comparado con el grupo control. Las trayectorias AI -> FDUP y UP -> PL son notablemente más fuertes en el grupo experimental, mientras que el grupo control muestra fortalezas en trayectorias relacionadas con FDUP.

**Tabla 21.** Estadísticos T de coeficientes de trayectoria [modelo estructural (interno)] (grupo control)

	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics ( O/STDEV )	P values
AI -> FDUP	0.524	0.495	0.299	<b>1.749</b>	<b>0.080</b>
C -> EM	0.720	0.737	0.068	<b>10.537</b>	<b>0.000</b>
C -> IU	0.105	0.103	0.119	<b>0.883</b>	<b>0.377</b>
C -> PL	-0.175	-0.144	0.245	<b>0.716</b>	<b>0.474</b>
C -> UP	-0.111	-0.093	0.258	<b>0.431</b>	<b>0.666</b>
CF -> FDUP	0.166	0.243	0.199	<b>0.835</b>	<b>0.404</b>
EM -> PL	0.380	0.354	0.188	<b>2.025</b>	<b>0.043</b>
EM -> UP	0.100	0.091	0.221	<b>0.454</b>	<b>0.650</b>
FDUP -> IU	0.328	0.316	0.145	<b>2.261</b>	<b>0.024</b>
FDUP -> PL	0.474	0.436	0.160	<b>2.953</b>	<b>0.003</b>
FDUP -> UP	0.468	0.431	0.168	<b>2.791</b>	<b>0.005</b>
IS -> UP	0.366	0.400	0.125	<b>2.920</b>	<b>0.004</b>
PL -> IU	0.389	0.416	0.147	<b>2.648</b>	<b>0.008</b>
UP -> IU	-0.022	-0.037	0.163	<b>0.134</b>	<b>0.893</b>
UP -> PL	0.225	0.260	0.128	<b>1.764</b>	<b>0.078</b>

Fuente: *elaboración propia.*

Después de analizar el coeficiente de trayectoria para el modelo interno (estructural), se evaluó el nivel de significancia del modelo externo. Esto se llevó a cabo verificando el estadístico t relacionado con las cargas del modelo externo. Como se demuestra en las tablas 22 y 23, los valores de los estadísticos t de todas las cargas son mayores que 1.96. Por lo tanto, todas las cargas del modelo externo son significativas al nivel de significancia de 0.01.

### ***Interpretación de los resultados sobre la significancia del modelo externo***

La mayoría de las cargas son altamente significativas en ambos grupos, lo que refuerza la validez de las relaciones entre las variables observadas y sus constructos latentes en el modelo. En resumen, ambos grupos muestran una fuerte consistencia en las cargas del modelo externo, con la mayoría de las cargas siendo significativas. Las diferencias menores en las cargas no significativas sugieren áreas donde el modelo puede necesitar ajustes adicionales dependiendo del grupo específico.

**Tabla 22.** Estadísticos *T* y valores *P* de las cargas del modelo externo (grupo experimental)

	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics ( O/STDEV )	P values
AI1 <- AI	0.413	0.419	0.258	1.602	0.109
AI2 <- AI	0.544	n/a	n/a	n/a	n/a
AI3 <- AI	0.879	0.879	0.080	10.987	0.000
AI4 <- AI	0.853	0.841	0.097	8.811	0.000
C1 <- C	0.744	0.732	0.130	5.725	0.000
C2 <- C	0.753	0.728	0.141	5.327	0.000
C3 <- C	0.602	0.595	0.171	3.525	0.000
C4 <- C	0.663	0.673	0.116	5.708	0.000
CF1 <- CF	0.815	0.739	0.280	2.907	0.004
CF2 <- CF	0.967	0.883	0.330	2.933	0.003
EM1 <- EM	0.784	0.777	0.113	6.959	0.000
EM2 <- EM	0.879	0.875	0.110	8.016	0.000
EM3 <- EM	0.689	0.655	0.190	3.626	0.000
FDUP1 <- FDUP	0.696	0.665	0.178	3.909	0.000
FDUP2 <- FDUP	0.662	0.633	0.217	3.053	0.002
FDUP3 <- FDUP	0.809	0.809	0.107	7.594	0.000
IS1 <- IS	0.928	0.905	0.204	4.556	0.000
IS2 <- IS	0.949	0.901	0.210	4.528	0.000
IS3 <- IS	0.879	0.824	0.211	4.164	0.000
IS4 <- IS	0.493	0.454	0.250	1.970	0.049
IU1 <- IU	0.960	0.954	0.025	37.907	0.000
IU2 <- IU	0.946	0.940	0.029	33.115	0.000
IU3 <- IU	0.938	0.938	0.029	32.077	0.000
PL1 <- PL	0.907	0.904	0.056	16.243	0.000
PL2 <- PL	0.947	0.945	0.021	44.271	0.000
PL3 <- PL	0.708	0.706	0.192	3.677	0.000
PL4 <- PL	0.842	0.844	0.056	15.124	0.000
UP1 <- UP	0.841	0.837	0.140	5.988	0.000
UP2 <- UP	0.694	0.664	0.249	2.787	0.005
UP3 <- UP	0.826	0.807	0.181	4.554	0.000

Fuente: *elaboración propia.*

**Tabla 23.** Estadísticos T y valores P de las cargas del modelo externo (grupo control)

	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics ( O/STDEV )	P values
AI1 <- AI	0.895	0.812	0.326	2.743	0.006
AI2 <- AI	0.663	0.431	0.477	1.390	0.165
AI3 <- AI	0.852	0.678	0.321	2.654	0.008
AI4 <- AI	0.827	0.700	0.298	2.774	0.006
C1 <- C	0.611	0.588	0.184	3.328	0.001
C2 <- C	0.734	0.690	0.156	4.713	0.000
C3 <- C	0.813	0.797	0.124	6.538	0.000
C4 <- C	0.785	0.796	0.074	10.566	0.000
CF1 <- CF	0.846	0.808	0.138	6.115	0.000
CF2 <- CF	0.939	0.938	0.060	15.689	0.000
EM1 <- EM	0.866	0.866	0.066	13.109	0.000
EM2 <- EM	0.839	0.798	0.166	5.054	0.000
EM3 <- EM	0.488	0.457	0.265	1.844	0.065
FDUP1 <- FDUP	0.917	0.905	0.053	17.447	0.000
FDUP2 <- FDUP	0.928	0.926	0.029	31.498	0.000
FDUP3 <- FDUP	0.927	0.910	0.055	16.874	0.000
IS1 <- IS	0.534	0.446	0.302	1.767	0.077
IS2 <- IS	0.617	0.532	0.282	2.186	0.029
IS3 <- IS	0.746	0.716	0.143	5.225	0.000
IS4 <- IS	0.744	0.754	0.111	6.687	0.000
IU1 <- IU	0.928	0.929	0.030	30.597	0.000
IU2 <- IU	0.950	0.947	0.020	48.360	0.000
IU3 <- IU	0.923	0.918	0.038	24.429	0.000
PL1 <- PL	0.871	0.861	0.049	17.662	0.000
PL2 <- PL	0.903	0.897	0.031	29.207	0.000
PL3 <- PL	0.829	0.823	0.061	13.682	0.000
PL4 <- PL	0.836	0.846	0.036	23.524	0.000
UP1 <- UP	0.726	0.710	0.161	4.514	0.000
UP2 <- UP	0.903	0.904	0.042	21.627	0.000
UP3 <- UP	0.815	0.804	0.110	7.378	0.000

Fuente: elaboración propia.

### **B. La relevancia del modelo estructural**

En cuanto a la relevancia, se utilizaron los valores de los coeficientes de trayectoria. Estos valores explican qué tan fuerte es el efecto de una variable latente sobre otra

variable latente. Los valores del coeficiente de trayectoria varían de -1 a +1. Los valores más cercanos a +1 significan relaciones positivas fuertes y los valores más cercanos a -1 significan relaciones negativas fuertes.

La valoración de los coeficientes de trayectoria del modelo estructural no debe limitarse únicamente a los efectos directos. Por lo tanto, también se consideraron los efectos totales. Los efectos totales se muestran en las tablas 24 y 25.

La consideración de los efectos totales ayudó a examinar la influencia de cada una de las variables latentes exógenas sobre una variable latente objetivo utilizando todas las variables latentes mediadoras. Como resultado, se espera que esto proporcione una idea más clara sobre las diferentes relaciones del modelo estructural.

El análisis de los valores mostrados en la tabla 24 que corresponde al grupo experimental muestra que la relación más fuerte es entre PL e IU (0.925), seguida de la relación entre C y EM (0.746), luego la relación entre IS y UP (0.619). Sin embargo, la variable latente que tiene el efecto más fuerte en IU es PL (0.925), seguida de UP (0.382), luego FDUP (0.318), luego EM (0.302), luego IS (0.236), luego AI (0.197), luego CF (0.020) y por último C (-0.105).

**Tabla 24.** *Efectos directos, indirectos y totales (grupo experimental)*

Variable dependiente	R <sup>2</sup>	Variable independiente	Efecto directo	Efecto indirecto	Efecto total
Intención de Uso	0.470	Percepción Lúdica	0.925	0.000	0.925***
		Utilidad Percibida	-0.079	0.461	0.382*
		Facilidad de uso Percibida	0.124	0.194	0.318**
		Autoeficacia Informática	0.000	0.197	0.197*
		Influencia Social	0.000	0.236	0.236
		Condiciones Facilitadoras	0.000	0.020	0.020
		Expectativa de Meta	0.000	0.302	0.302
		Contenido	-0.259	0.154	-0.105

Percepción Lúdica	0.467	Utilidad Percibida	0.498	0.000	0.498**
		Facilidad de uso Percibida	0.088	0.147	0.235
		Autoeficacia Informática	0.000	0.145	0.145
		Influencia Social	0.000	0.308	0.308*
		Condiciones Facilitadoras	0.000	0.014	0.014
		Expectativa de Meta	0.479	-0.185	0.294
		Contenido	-0.110	0.260	0.150
Utilidad Percibida	0.441	Facilidad de uso Percibida	0.295	0.000	0.295
		Autoeficacia Informática	0.000	0.182	0.182
		Influencia Social	0.619	0.000	0.619***
		Condiciones Facilitadoras	0.000	0.018	0.018
		Expectativa de Meta	-0.370	0.000	-0.370
		Contenido	0.081	-0.276	-0.195
Facilidad de uso Percibida	0.382	Autoeficacia Informática	0.618	0.000	0.618***
		Condiciones Facilitadoras	0.063	0.000	0.063
Expectativa de Meta	0.518	Contenido	0.746	0.000	0.746***

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Fuente: *elaboración propia*.

Para el caso del grupo control la tabla 25 muestra que la relación más fuerte es entre FDUP y PL (0.579), seguida de la relación entre FDUP e IU (0.543), luego la relación entre AI y FDUP (0.524). Sin embargo, la variable latente que tiene el efecto más fuerte en IU es FDUP (0.543), seguida de PL (0.389), luego AI (0.284), luego EM (0.155), luego C (0.141), luego CF (0.090), luego UP (0.066) y por último IS (0.024).

**Tabla 25.** *Efectos directos, indirectos y totales (grupo control)*

Variable dependiente	R <sup>2</sup>	Variable independiente	Efecto directo	Efecto indirecto	Efecto total
		Percepción Lúdica	0.389	0.000	0.389***

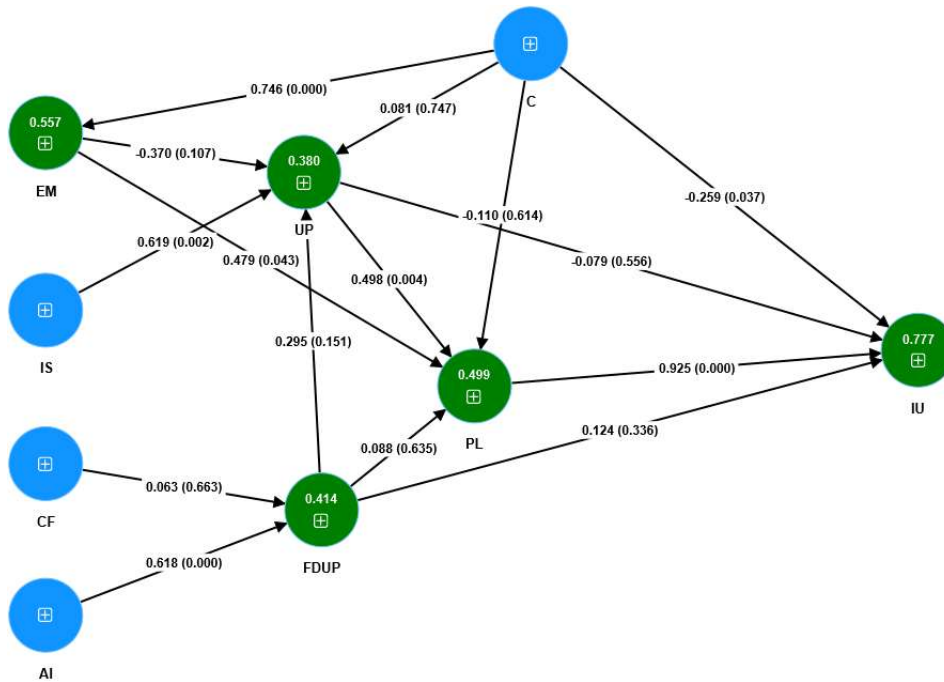


Intención de Uso	0.470	Utilidad Percibida	-0.022	0.088	0.066
		Facilidad de uso Percibida	0.328	0.215	0.543***
		Autoeficacia Informática	0.000	0.284	0.284
		Influencia Social	0.000	0.024	0.024
		Condiciones Facilitadoras	0.000	0.090	0.090
		Expectativa de Meta	0.000	0.155	0.155
		Contenido	0.105	0.036	0.141
Percepción Lúdica	0.467	Utilidad Percibida	0.225	0.000	0.225*
		Facilidad de uso Percibida	0.474	0.105	0.579***
		Autoeficacia Informática	0.000	0.303	0.303
		Influencia Social	0.000	0.082	0.082
		Condiciones Facilitadoras	0.000	0.096	0.096
		Expectativa de Meta	0.380	0.023	0.403**
		Contenido	-0.175	0.265	0.090
Utilidad Percibida	0.441	Facilidad de uso Percibida	0.468	0.000	0.468***
		Autoeficacia Informática	0.000	0.245	0.245
		Influencia Social	0.366	0.000	0.366***
		Condiciones Facilitadoras	0.000	0.078	0.078
		Expectativa de Meta	0.100	0.000	0.100
		Contenido	-0.111	0.072	-0.039
Facilidad de uso Percibida	0.382	Autoeficacia Informática	0.524	0.000	0.524*
		Condiciones Facilitadoras	0.166	0.000	0.166
Expectativa de Meta	0.518	Contenido	0.720	0.000	0.72***

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Fuente: *elaboración propia*.

Finalmente, en la figura 15 y 16 se muestran los resultados de los coeficientes de ruta del modelo CBAAM aplicado a este estudio.



**Figura 15.** Coeficientes de ruta del modelo CBAAM aplicado a este estudio (grupo experimental)

Fuente: elaboración propia.

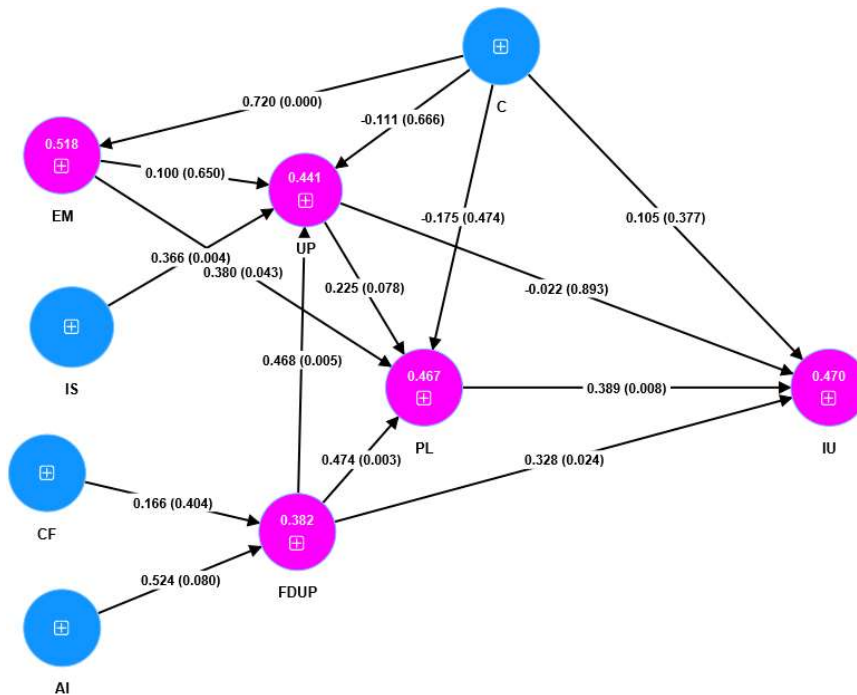
### **Interpretación de los resultados sobre la relevancia del modelo estructural**

Para el grupo experimental, se observa un modelo robusto con varias relaciones significativas, especialmente en la Percepción Lúdica e Intención de Uso, con fuertes efectos directos e indirectos. La Percepción Lúdica y la Utilidad Percibida juegan roles clave en la determinación de la Intención de Uso.

Para el grupo control, se observa un modelo sólido con relaciones significativas, aunque algunas relaciones son marginalmente significativas. La Facilidad de Uso Percibida y la Influencia Social son determinantes clave en la Percepción Lúdica y la Utilidad Percibida.

En resumen, ambos grupos muestran modelos estructurales con relaciones significativas que destacan diferentes aspectos de la percepción y uso de la tecnología. El grupo

experimental muestra un modelo ligeramente más robusto con más relaciones directas e indirectas significativas, mientras que el grupo control también muestra un modelo sólido, pero con algunas diferencias en la magnitud y significancia de las relaciones.



**Figura 16.** Coeficientes de ruta del modelo CBAAM aplicado a este estudio (grupo control)

Fuente: elaboración propia.

### C. Conclusión de importancia y relevancia

Como se resume en las Tablas 26 y 27, algunas de las relaciones del modelo estructural son significativas y otras no. Las relaciones que son significativas confirman las hipótesis correspondientes que fueron propuestas conforme al modelo CBAAM, mientras que las hipótesis relacionadas con el resto de estas relaciones se rechazan. Los valores t y los valores p de estas relaciones se obtuvieron utilizando el procedimiento de bootstrapping que está disponible en SmartPLS 4.0.

**Tabla 26.** Prueba de significancia de los coeficientes de trayectoria del modelo estructural (grupo experimental)

Hipótesis	Trayectoria		Coefficiente de trayectoria	t-values	p-values	Resultados
H1	PL->IU	La percepción lúdica tendrá un efecto positivo en la intención de utilizar la EBC	0.925***	6.154	0.000	Aceptada
H2	UP->IU	La utilidad percibida tendrá un efecto positivo en la intención de utilizar la EBC	-0.079	0.588	0.556	Rechazada
H3	UP->PL	La utilidad percibida tendrá un efecto positivo en la percepción lúdica	0.498***	2.846	0.004	Aceptada
H4	FDUP->IU	La facilidad de uso percibida tendrá un efecto positivo en la intención de utilizar la EBC	0.124	0.962	0.336	Rechazada
H5	FDUP->UP	La facilidad de uso percibida tendrá un efecto positivo en la utilidad percibida	0.295	1.436	0.151	Rechazada
H6	FDUP->PL	La facilidad de uso percibida tendrá un efecto positivo en la percepción lúdica	0.088	0.475	0.635	Rechazada
H7	AI->FDUP	La autoeficacia informática tendrá un efecto positivo en la facilidad de uso percibida	0.618***	4.905	0.000	Aceptada
H8	IS->UP	La influencia social tendrá un efecto positivo en la utilidad percibida	0.619***	3.112	0.002	Aceptada
H9	CF->FDUP	Las condiciones facilitadoras tendrán un efecto positivo en la facilidad de uso percibida	0.063	0.436	0.663	Rechazada

H10	EM->UP	La expectativa de meta tendrá un efecto positivo en la utilidad percibida	-0.370	1.612	0.107	Rechazada
H11	EM->PL	La expectativa de meta tendrá un efecto positivo en la percepción lúdica	0.479**	2.021	0.043	Aceptada
H12	C->UP	El contenido tendrá un efecto positivo en la utilidad percibida	0.081	0.323	0.747	Rechazada
H13	C->PL	El contenido tendrá un efecto positivo en la percepción lúdica	-0.110	0.504	0.614	Rechazada
H14	C->EM	El contenido tendrá un efecto positivo en la expectativa de meta	0.746***	9.169	0.000	Aceptada
H15	C->IU	El contenido tendrá un efecto positivo en la intención de uso de la EBC	-0.259**	2.092	0.037	Aceptada

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Fuente: *elaboración propia*.

**Tabla 27.** Prueba de significancia de los coeficientes de trayectoria del modelo estructural (grupo control)

Hipótesis	Trayectoria		Coefficiente de trayectoria	t-values	p-values	Resultados
H1	PL->IU	La percepción lúdica tendrá un efecto positivo en la intención de utilizar la EBC	0.389***	2.648	0.008	Aceptada

H2	UP->IU	La utilidad percibida tendrá un efecto positivo en la intención de utilizar la EBC	-0.022	0.134	0.893	Rechazada
H3	UP->PL	La utilidad percibida tendrá un efecto positivo en la percepción lúdica	0.225	1.764	0.078	Rechazada
H4	FDUP->IU	La facilidad de uso percibida tendrá un efecto positivo en la intención de utilizar la EBC	0.328**	2.261	0.024	Aceptada
H5	FDUP->UP	La facilidad de uso percibida tendrá un efecto positivo en la utilidad percibida	0.468***	2.791	0.005	Aceptada
H6	FDUP->PL	La facilidad de uso percibida tendrá un efecto positivo en la percepción lúdica	0.474***	2.953	0.003	Aceptada
H7	AI->FDUP	La autoeficacia informática tendrá un efecto positivo en la facilidad de uso percibida	0.524*	1.749	0.080	Rechazada
H8	IS->UP	La influencia social tendrá un efecto positivo en la utilidad percibida	0.366***	2.920	0.004	Aceptada
H9	CF->FDUP	Las condiciones facilitadoras tendrán un efecto positivo en la facilidad de uso percibida	0.166	0.835	0.404	Rechazada
H10	EM->UP	La expectativa de meta tendrá un efecto positivo en la utilidad percibida	0.100	0.454	0.650	Rechazada
H11	EM->PL	La expectativa de meta tendrá un efecto positivo en la percepción lúdica	0.380**	2.025	0.043	Aceptada

H12	C->UP	El contenido tendrá un efecto positivo en la utilidad percibida	-0.111	0.431	0.666	Rechazada
H13	C->PL	El contenido tendrá un efecto positivo en la percepción lúdica	-0.175	0.716	0.474	Rechazada
H14	C->EM	El contenido tendrá un efecto positivo en la expectativa de meta	0.720***	10.537	0.000	Aceptada
H15	C->IU	El contenido tendrá un efecto positivo en la intención de uso de la EBC	0.105	0.883	0.377	Rechazada

\*p < 0.10, \*\*p < 0.05, \*\*\*p < 0.01.

Fuente: *elaboración propia*.

### ***Interpretación sobre los resultados de las hipótesis confirmadas y rechazadas***

#### ***Hipótesis aceptadas en ambos grupos:***

H1 (PL -> IU): La percepción lúdica tiene un efecto positivo significativo en la intención de utilizar la EBC tanto en el grupo experimental como en el grupo control. Esto sugiere que, independientemente del grupo, la percepción de diversión o entretenimiento asociada con la herramienta es un factor clave que impulsa la intención de uso.

H11 (EM -> PL): La expectativa de meta tiene un efecto positivo significativo en la percepción lúdica en ambos grupos. Este hallazgo indica que las expectativas y objetivos que los usuarios tienen pueden influir en cómo perciben la diversión asociada con la herramienta, en ambos contextos.

H14 (C -> EM): El contenido tiene un efecto positivo significativo en la expectativa de meta tanto en el grupo experimental como en el grupo control. Esto demuestra que el

contenido de la herramienta es crucial para formar expectativas positivas, independientemente del grupo.

***Hipótesis aceptadas solo en el grupo experimental:***

H3 (UP -> PL): La utilidad percibida tiene un efecto positivo en la percepción lúdica en el grupo experimental, sugiriendo que los usuarios que perciben la herramienta como útil también tienden a encontrarla más divertida.

H7 (AI -> FDUP): La autoeficacia informática tiene un efecto positivo en la facilidad de uso percibida solo en el grupo experimental, indicando que la confianza en las propias habilidades informáticas es más relevante en este grupo.

H8 (IS -> UP): La influencia social tiene un efecto positivo en la utilidad percibida en el grupo experimental, lo que sugiere que la percepción de utilidad de la herramienta está más influenciada por la opinión de otros en este grupo.

H15 (C -> IU): El contenido tiene un efecto negativo en la intención de uso de la EBC en el grupo experimental, lo cual es un resultado inesperado que podría indicar problemas con el contenido específico percibido por los usuarios de este grupo.

***Hipótesis aceptadas solo en el grupo control:***

H4 (FDUP -> IU): La facilidad de uso percibida tiene un efecto positivo en la intención de utilizar la EBC solo en el grupo control, indicando que este grupo valora más la facilidad de uso como un factor para la intención de uso.

H5 (FDUP -> UP): La facilidad de uso percibida tiene un efecto positivo en la utilidad percibida solo en el grupo control, sugiriendo que este grupo ve la facilidad de uso como un componente de la utilidad.



H6 (FDUP -> PL): La facilidad de uso percibida tiene un efecto positivo en la percepción lúdica solo en el grupo control, indicando que este grupo asocia más la facilidad de uso con la diversión.

***Hipótesis rechazadas en ambos grupos:***

H2 (UP -> IU): La utilidad percibida no tiene un efecto significativo en la intención de utilizar la EBC en ninguno de los grupos.

H9 (CF -> FDUP): Las condiciones facilitadoras no tienen un efecto significativo en la facilidad de uso percibida en ambos grupos.

H10 (EM -> UP): La expectativa de meta no tiene un efecto significativo en la utilidad percibida en ambos grupos.

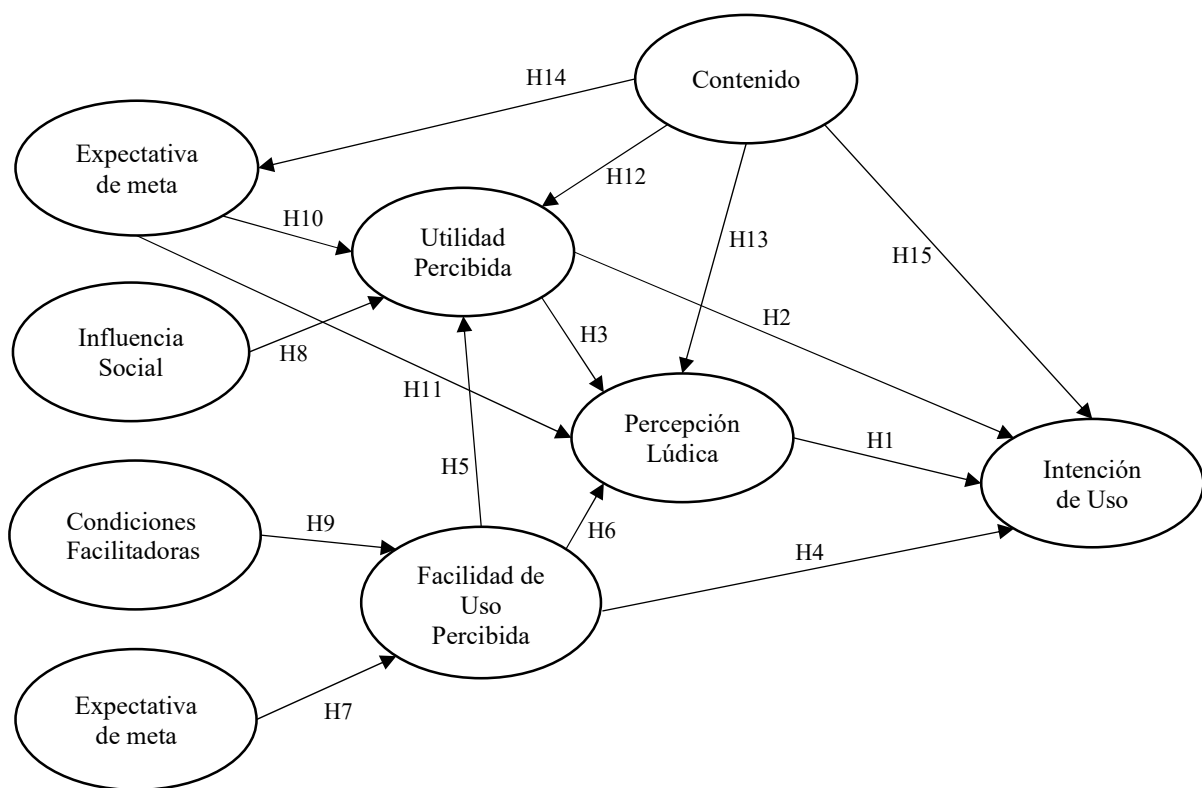
H12 (C -> UP): El contenido no tiene un efecto significativo en la utilidad percibida en ambos grupos.

H13 (C -> PL): El contenido no tiene un efecto significativo en la percepción lúdica en ambos grupos.

En resumen, la percepción lúdica, la expectativa de meta y el contenido son factores importantes que afectan tanto la intención de uso como la percepción lúdica en ambos grupos. Sin embargo, se observan diferencias en la relevancia de la utilidad percibida, la autoeficacia informática, la influencia social y la facilidad de uso percibida entre los grupos experimental y control. Estos hallazgos indican que diferentes factores pueden ser más o menos influyentes dependiendo del formato de preguntas utilizadas, lo cual es crucial para desarrollar estrategias efectivas de implementación y promoción de la herramienta de evaluación basada en computadora (EBC).

## 5.4 Discusión

El objetivo general de esta investigación fue analizar el uso y la aceptación de una herramienta de evaluación automática de conceptos básicos de física basados en preguntas de respuesta corta. Las relaciones entre los constructos que influyen en la adopción y uso de esta tecnología en los alumnos se exploraron a través de una serie de 15 hipótesis relacionales basadas en el Modelo de Aceptación de Evaluación Basado en Computadora (CBAAM, Terzis y Economides, 2011) mostrado en la figura 17.



**Figura 17.** Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM) de Terzis y Economides (2011).

La relevancia de este estudio se enmarca en el creciente interés por la integración de tecnologías de evaluación en la educación superior, especialmente en áreas como la física, donde la comprensión conceptual es fundamental. La investigación se desarrolló mediante una comparación de dos instrumentos de evaluación automática: por un lado,

un grupo control que utilizó una evaluación basada en preguntas de opción múltiple, y por otro, un grupo experimental que empleó preguntas de respuesta corta.

Este enfoque comparativo permitió examinar las actitudes, percepciones y nivel de aceptación de los estudiantes hacia estas metodologías de evaluación. La discusión que sigue se centra en interpretar los hallazgos de esta investigación, considerando tanto las implicaciones pedagógicas como las posibles direcciones futuras en el campo de la evaluación basada en computadora (EBC). Todos los aportes a esta área de investigación ayudan a las instituciones a tener una implementación exitosa de la EBC. En la literatura, la Utilidad Percibida, la Facilidad de Uso Percibida, la Percepción Lúdica y la Importancia Percibida se consideran elementos principales de la Intención de Utilizar EBC (Terzis et al., 2013).

El estudio mostró que la Percepción Lúdica tiene un impacto directo en la Intención de Uso para los 2 tipos de instrumentos de evaluación, mientras que los constructos que tienen un impacto indirecto en la Intención de Uso son la Utilidad Percibida, Facilidad de Uso Percibida, Autoeficacia Informática, Influencia social, Condiciones Facilitadoras, Expectativa de Meta y Contenido (véase tablas 24 y 25).

El constructo Contenido que se utilizó de esta manera por primera vez en el modelo CBAAM tuvo un impacto directo en la Intención de Uso como sugiere la hipótesis de este estudio para el caso del grupo experimental. Sin embargo, para el grupo control no tuvo un impacto directo, lo cual es el mismo resultado es que obtuvieron Terzis y Economides, 2011, en el estudio que realizaron para un instrumento de evaluación que contenía preguntas en formato de opción múltiple. Esto es uno de los resultados más importante de destacar, ya que el contenido está directamente relacionado con el formato de las preguntas. Como se mencionó en capítulo 3, las preguntas de respuesta corta permiten a los que los estudiantes utilizar sus propios conocimientos para construir sus respuestas en lugar de elegir entre una lista de opciones, como responder a preguntas de opción múltiple (Martinez, 1991). Este proceso les da la oportunidad de explorar su propio aprendizaje y tener una experiencia más lúdica.

En CBAAM se estudian dos dimensiones del Contenido; el contenido del curso y el contenido de las preguntas. En cuanto al contenido del curso, se cree que afecta en gran medida la Utilidad Percibida y la Percepción Lúdica de la EBC. El contenido del curso puede determinar si es útil o no, interesante o no y finalmente difícil o no. En este modelo también, el contenido de las preguntas se examina para determinar si son claras, fáciles para entender y relacionado con el contenido del curso. Basándose en estos resultados, cuando el contenido de una EBC se diseña cuidadosamente, el EBC sería más útil y lúdica para los estudiantes, por lo que sería más probable que se utilizará.

El contenido tiene un efecto directo sobre la Utilidad Percibida, la Percepción Lúdica y la Expectativa de Meta que indica una influencia indirecta en la Intención de Uso. En cuanto a la Expectativa de Meta que también se utilizó por primera vez en este modelo, se demostró que los estudiantes encuentran una EBC lúdica cuando tienen buenas expectativas del sistema. Además, el efecto positivo de la influencia social en la utilidad percibida proporcionada por TAM2 (Venkatesh & Davis, 2000) también fue apoyado por este modelo. En cuanto al efecto positivo de la Influencia Social en la Utilidad Percibida proporcionada por TAM2 también fue apoyado por este modelo.

Los resultados de este estudio se relacionan con el papel de la Autoeficacia Informática, el Contenido, las Condiciones de Facilitadoras, la Expectativa de Meta, la Facilidad de Uso Percibida, la Influencia Social, la Percepción Lúdica y la Utilidad Percibida en la intención de uso de la EBC por parte de los estudiantes.

La tabla 28 resume los resultados de este estudio para el grupo experimental y control. Se muestran las 15 hipótesis y si fueron respaldadas por el modelo o no. El estudio concluye que es más probable que un sistema sea utilizado por los estudiantes si es lúdico y esto también se confirma estudios previos. Además, una EBC es más probable que sea lúdica cuando es fácil de usar.

**Tabla 28. Resultados de las pruebas de hipótesis**

Hipótesis	Trayectoria	Resultados grupo experimental	Resultados grupo control
H1	PL->IU	Aceptada	Aceptada
H2	UP->IU	Rechazada	Rechazada
H3	UP->PL	Aceptada	Rechazada
H4	FDUP->IU	Rechazada	Aceptada
H5	FDUP->UP	Rechazada	Aceptada
H6	FDUP->PL	Rechazada	Aceptada
H7	AI->FDUP	Aceptada	Rechazada
H8	IS->UP	Aceptada	Aceptada
H9	CF->FDUP	Rechazada	Rechazada
H10	EM->UP	Rechazada	Rechazada
H11	EM->PL	Aceptada	Aceptada
H12	C->UP	Rechazada	Rechazada
H13	C->PL	Rechazada	Rechazada
H14	C->EM	Aceptada	Aceptada
H15	C->IU	Aceptada	Rechazada

Fuente: *elaboración propia.*

La facilidad de uso percibida es uno de los constructos principales en TAM (Davis, 1989) y el efecto directo sobre la Intención de Uso era lo que se esperaba, lo cual se comprobó en el grupo control ya que se aceptaron las tres hipótesis relacionadas con este constructo. Sin embargo, para el caso del grupo experimental estas mismas hipótesis se rechazaron. Esto posiblemente se explica por el hecho de que los estudiantes encuentran un tipo de formato de examen más fácil que el otro para demostrar su conocimiento y comprensión de los temas evaluados.

Además, el análisis muestra que la Utilidad Percibida no tiene un efecto directo sobre la Intención de Uso para los dos grupos. Este es un resultado controvertido, ya que una posible interpretación es que los estudiantes no encuentran un tipo de formato de examen más útil que el otro para demostrar su conocimiento y comprensión de los temas evaluados. Los estudios previos respaldan un efecto muy fuerte de la Utilidad Percibida en la intención de uso (por ejemplo, TAM). Por otro lado, la Utilidad Percibida tiene un

efecto indirecto positivo significativo en la Intención de Uso a través de la Percepción Lúdica. Por lo tanto, si una EBC es útil, es más probable que sea Lúdica.

#### **5.4.1 Contribuciones significativas**

Este estudio aborda con profundidad la intersección entre tecnología educativa y pedagogía, a través del análisis de la aceptación estudiantil hacia sistemas de evaluación basada en computadora (EBC), utilizando preguntas de respuesta corta en contraste con las tradicionales preguntas de opción múltiple. La diferenciación en la aceptación de estos dos formatos, respaldada por el Modelo de Aceptación de Evaluación Basado en Computadora (CBAAM), revela una preferencia notable hacia el formato de respuesta corta, sugiriendo que este formato no solo facilita una evaluación más profunda de la comprensión conceptual, sino que también mejora la percepción lúdica de la evaluación por parte de los estudiantes.

La introducción del constructo Contenido en el modelo CBAAM como un influenciador directo de la Intención de Uso en el grupo experimental es un hallazgo innovador. Este resultado es coherente con la teoría pedagógica que sostiene que las preguntas abiertas promueven un nivel más alto de reflexión crítica y autoexpresión, permitiendo a los estudiantes aplicar y expresar su conocimiento de manera más integral que en el formato de opción múltiple. Además, este hallazgo resalta la importancia de considerar el diseño de contenido en el desarrollo de herramientas de EBC, subrayando que un contenido bien diseñado puede mejorar significativamente la utilidad percibida y la percepción lúdica de la herramienta, fomentando una mayor aceptación y uso por parte de los estudiantes.

La discusión sobre la Percepción Lúdica y su impacto directo en la Intención de Uso para ambos tipos de instrumentos de evaluación enfatiza la relevancia de crear experiencias de evaluación que no solo sean educativas, sino también atractivas para los estudiantes. Este enfoque hacia una evaluación más lúdica y participativa puede ser crucial para el

éxito de la implementación de la EBC en contextos educativos contemporáneos, donde la motivación y la participación del estudiante son desafíos constantes.

Sin embargo, es fundamental reconocer las limitaciones inherentes al estudio, como la contextualización específica de la muestra y la concentración en la física como área temática. Estas limitaciones abren oportunidades para futuras investigaciones que podrían expandir la comprensión del impacto del formato de evaluación en diversas disciplinas y contextos educativos, así como explorar otros constructos que puedan influir en la aceptación de la EBC por parte de los estudiantes.

En suma, este estudio no solo aporta a la teoría existente con hallazgos empíricos significativos sobre la preferencia de formato de evaluación y la aceptación de la EBC, sino que también ofrece una contribución práctica al diseño pedagógico, recomendando una consideración cuidadosa del contenido y la metodología de evaluación para fomentar experiencias de aprendizaje más ricas y atractivas. La profundización en estos aspectos es vital para el avance de la pedagogía moderna y la integración efectiva de la tecnología en la educación.

#### **5.4.2 Similitudes y diferencias con estudios previos**

Al vincular los hallazgos de este estudio con la literatura existente, se evidencia una continuidad y, al mismo tiempo, se destacan novedosas diferencias en el entendimiento del impacto de diferentes formatos de evaluación en la aceptación y uso de la evaluación basada en computadora (EBC) por parte de los estudiantes. Estudios anteriores, como los realizados por Terzis y Economides (2011), han establecido la importancia de la Utilidad Percibida, la Facilidad de Uso Percibida y la Percepción Lúdica en la adopción de tecnologías educativas. Sin embargo, este trabajo amplía el marco de referencia al introducir y validar el impacto del Contenido como un factor directamente influyente en la Intención de Uso de la EBC, particularmente para preguntas de respuesta corta. Esta contribución resalta una dimensión que ha tenido poca relevancia en la literatura previa,

proponiendo que la naturaleza del contenido puede ser tan decisiva como las percepciones de utilidad y facilidad de uso.

Este estudio también se distingue por su enfoque comparativo, examinando la adopción de EBC entre grupos de control y experimental con diferentes formatos de evaluación. Mientras que la investigación previa ha tendido a concentrarse en un único formato de evaluación, este enfoque dual permite una comparación directa de cómo el formato de las preguntas puede influir en la percepción estudiantil y la aceptación de la EBC. Este método comparativo representa una innovación metodológica significativa, proporcionando evidencia empírica de que el formato de respuesta corta puede mejorar significativamente la percepción lúdica y la utilidad percibida de la EBC en comparación con las preguntas de opción múltiple.

Además, la incorporación del constructo Contenido en el análisis del modelo CBAAM y su validación como un influenciador directo de la Intención de Uso para preguntas de respuesta corta es una contribución específica de este estudio. A diferencia de investigaciones anteriores que han enfocado predominantemente en aspectos técnicos y de percepción de la EBC, este hallazgo subraya la relevancia del contenido de la evaluación en la aceptación de la tecnología. Este enfoque único ofrece una nueva perspectiva sobre cómo la presentación del material de evaluación afecta la experiencia de aprendizaje y evaluación del estudiante, invitando a futuras investigaciones a considerar el contenido de la evaluación como un componente esencial en el diseño de EBC.

En conjunto, estas contribuciones específicas enriquecen significativamente el campo de estudio de la evaluación basada en computadora, ofreciendo nuevas direcciones para la investigación futura y prácticas recomendadas para la implementación efectiva de EBC en entornos educativos. Este trabajo no solo se alinea con la literatura existente en términos de destacar la importancia de la percepción estudiantil en la aceptación de la EBC, sino que también avanza en el debate al identificar el contenido de la evaluación y



el formato de las preguntas como factores cruciales para mejorar la aceptación y efectividad de la EBC en la educación superior.

### **5.4.3 Limitaciones, Impacto y Recomendaciones**

A pesar de los hallazgos significativos de este estudio, es crucial reconocer sus limitaciones y proponer áreas para futuras investigaciones que puedan superarlas. Una de las principales limitaciones radica en la generalización de los resultados, ya que la muestra se centra en estudiantes de una institución de educación superior específica y en el contexto de la enseñanza de conceptos básicos de física. Además, se identificaron dos sesgos importantes: Por un lado, tenemos que de acuerdo con la caracterización que se hizo de la muestra de alumnos que participaron en el estudio, se observó que el 75% tiene un promedio académico alto situado en el rango de 100-90, la literatura reporta que los efectos de la evaluación basada en computadora con retroalimentación inmediata son más notorios en estudiantes de rendimiento académico medio-bajo (Jordan y Mitchell, 2009). De manera similar tenemos que de los estudiantes que participaron en el estudio 68% fueron hombres y 32% fueron mujeres, la literatura reporta que los estudios sobre la aceptación de la tecnología no son iguales para los hombres y mujeres (Venkatesh et al. (2003).

Futuras investigaciones podrían expandir este estudio a una variedad más amplia de contextos educativos, disciplinas y niveles académicos para examinar si los patrones de aceptación de la EBC y el impacto de los distintos formatos de evaluación se mantienen o varían.

Otra limitación se relaciona con el diseño del estudio, que, aunque innovador, compara solo dos tipos de formatos de evaluación. Estudios posteriores podrían incluir una gama más diversa de formatos de evaluación, como preguntas de desarrollo, para entender mejor cómo diferentes métodos de evaluación influyen la percepción y aceptación estudiantil de la EBC. Además, sería valioso explorar en profundidad cómo variables

demográficas y académicas específicas, como el género, la edad, el historial académico y la familiaridad previa con la tecnología, pueden afectar la aceptación de la EBC.

En términos de relevancia práctica, este estudio destaca la importancia de considerar cuidadosamente el diseño de contenido y el formato de las evaluaciones al implementar sistemas de EBC. Para maximizar la utilidad percibida y la percepción lúdica de la EBC, se recomienda a los profesores y diseñadores instruccionales integrar preguntas de respuesta corta que fomenten la reflexión y permitan a los estudiantes expresar su conocimiento de manera más completa. Además, este estudio subraya la necesidad de desarrollar contenidos de evaluación que sean relevantes y desafiantes, promoviendo un aprendizaje más profundo y significativo.

Basándose en los resultados obtenidos, se sugiere que las instituciones educativas proporcionen capacitación y recursos a los profesores para el desarrollo e implementación efectiva de EBC, centrando la atención en las mejores prácticas para la creación de contenido de evaluación y en el uso de tecnologías que faciliten una experiencia de evaluación lúdica y enriquecedora para los estudiantes.

Mientras este estudio ofrece información valiosa sobre la adopción de la EBC en la educación superior, las limitaciones destacadas y las sugerencias para futuras investigaciones proporcionan un camino claro para ampliar nuestra comprensión y mejorar la práctica educativa en torno a las evaluaciones basadas en computadora. La relevancia práctica de estos hallazgos resalta el potencial de la EBC para transformar las experiencias de evaluación en la educación superior, alentando a una exploración más profunda y aplicación cuidadosa de estas tecnologías en el futuro.

Una recomendación fundamental que surge de esta disertación, considerando la amplia gama de modelos de aceptación de tecnología y la adopción particular del Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM) de Terzis y Economides (2011), es el desarrollo de un modelo innovador que aborde específicamente el impacto del formato de las preguntas en la adopción y efectividad de la evaluación electrónica. A

través del análisis exhaustivo realizado, se ha identificado una oportunidad significativa para enriquecer nuestra comprensión de cómo el diseño de las preguntas puede influir en la percepción de los usuarios sobre la utilidad, la facilidad de uso y la percepción lúdica de las herramientas de evaluación electrónica.

Este nuevo modelo debería incorporar constructos que examinen detalladamente las características específicas de los formatos de preguntas, como preguntas de respuesta corta frente a preguntas de opción múltiple, y su relación con los procesos cognitivos y emocionales de los estudiantes. Al hacerlo, podremos ofrecer recomendaciones más precisas y fundamentadas sobre cómo diseñar evaluaciones electrónicas que no solo mejoren el rendimiento académico, sino que también aumenten el compromiso y la satisfacción del estudiante.

Además, este modelo enriquecido permitiría a los profesores y diseñadores instruccionales comprender mejor las dinámicas subyacentes en la selección de formatos de preguntas y adaptar sus estrategias de evaluación para maximizar los beneficios educativos mientras minimizan las barreras para la adopción de la EBC. Por lo tanto, el desarrollo de este modelo podría representar otro paso importante hacia el avance de la innovación educativa y la mejora de las prácticas de evaluación en entornos de aprendizaje digital.

#### **5.4.4 Aprendizajes y Reflexiones finales**

Reflexionar sobre este proceso de investigación revela tanto los desafíos enfrentados como las profundas lecciones aprendidas en el camino. Desde el inicio, el propósito de explorar la aceptación y uso de herramientas de evaluación basadas en computadora (EBC) por parte de estudiantes universitarios se presentó no solo como una oportunidad para contribuir al campo de la innovación educativa, sino también como un desafío personal para profundizar mi comprensión de cómo la tecnología puede transformar el aprendizaje y la evaluación.

Una de las lecciones más significativas ha sido la importancia de un diseño metodológico riguroso. La elección de emplear un enfoque comparativo entre preguntas de respuesta corta y de opción múltiple demostró ser fundamental para explorar las complejidades de cómo diferentes formatos de evaluación pueden influir en la percepción estudiantil y su aceptación de la EBC. Este enfoque no solo enriqueció mi entendimiento del tema, sino que también subrayó la relevancia de aplicar teorías y modelos existentes, como el CBAAM, para investigar nuevas áreas de interés.

Durante el tiempo que se llevó a cabo esta investigación, se ha hecho evidente la importancia de ser flexible y adaptable. Enfrentar desafíos operativos, tales como obtener la autorización de profesores para interactuar con sus alumnos y asegurar el apoyo de la institución educativa para implementar un módulo especial de evaluación en la plataforma Moodle oficial, requirió un esfuerzo considerable. Estas habilidades de adaptación resultan fundamentales no solo en el ámbito investigativo, sino también en el educativo, un campo donde las necesidades y los entornos se transforman de manera continua.

La interacción con los participantes y el análisis de los datos recopilados revelaron la complejidad de las percepciones humanas y la diversidad de factores que influyen en la adopción de tecnología. Este estudio reafirmó lo que en mi experiencia como docente he observado en el aula: la necesidad de diseñar experiencias educativas que no solo sean pedagógicamente sólidas sino también atractivas y accesibles para los estudiantes.

En términos de contribuciones a la práctica docente, este estudio destaca el valor de incorporar variedad y flexibilidad en los métodos de evaluación. Al revelar cómo las preferencias y percepciones estudiantiles varían significativamente entre los formatos de evaluación, la investigación subraya la importancia de diseñar evaluaciones que no solo midan el rendimiento académico, sino que también mejoren la experiencia de aprendizaje.

Finalmente, este proceso de investigación ha sido un recordatorio de que la innovación educativa es un proceso iterativo, donde cada descubrimiento lleva a nuevas preguntas y oportunidades de mejora. Las lecciones aprendidas y los hallazgos de este estudio no solo contribuyen a la literatura existente, sino que también abren nuevas direcciones para próximas investigaciones. En última instancia, esta experiencia ha reforzado mi compromiso con la mejora continua de la educación a través de la innovación y la investigación.

## **CAPÍTULO 6. CONCLUSIONES**

En el presente trabajo, se realizó un estudio sobre los factores que influyeron en el comportamiento de los estudiantes hacia la intención de utilizar una evaluación basada en computadora en la educación superior. El enfoque fue cuantitativo, de tipo cuasi-experimental y de alcance explicativo. La pregunta de investigación general *¿Cuáles son los factores que influyen en el uso y aceptación de una herramienta de evaluación automática de conceptos básicos de física basados en preguntas de respuesta corta en estudiantes de una universidad privada de Puebla?* Se respondió mediante el análisis de los constructos que afectan la intención de utilizar la EBC, empleando el Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM) de Terzis y Economides (2011).

El modelo probado y la medición se apoyaron en los datos recopilados. Los resultados sugieren que hay una variabilidad considerable en la aceptación de las hipótesis entre el grupo experimental, que empleó preguntas de respuesta corta, y el grupo control, que utilizó preguntas de opción múltiple. Esto indica que las diferencias en la metodología de evaluación pueden tener un impacto significativo en la percepción de los usuarios sobre aspectos como la Percepción Lúdica, la Utilidad Percibida y la Facilidad de Uso Percibida, que son factores clave en el modelo CBAAM.

Por lo anterior, se podría afirmar que el tipo de instrumento de evaluación automática influye en la aceptación de las hipótesis relacionadas con el modelo CBAAM. Las diferencias observadas podrían atribuirse a las características inherentes de los dos tipos de preguntas, lo que sugiere que la elección del instrumento de evaluación debe considerarse cuidadosamente en función de los objetivos pedagógicos y las preferencias de los estudiantes. Es importante considerar estas diferencias al implementar evaluaciones automáticas en entornos educativos y al diseñar futuras investigaciones para explorar más a fondo cómo las distintas modalidades de evaluación afectan la aceptación y el rendimiento de los estudiantes.

A partir de los resultados obtenidos, se concluye que se cumplió el objetivo general: *Analizar el uso y la aceptación de una herramienta de evaluación automática de conceptos básicos de física basados en preguntas de respuesta corta, en estudiantes de una universidad privada de Puebla.*

El proceso para lograr el objetivo general consistió en alcanzar cuatro objetivos específicos. Para la primera etapa de la investigación se plantearon dos objetivos específicos: *Identificar las técnicas y herramientas empleadas actualmente para la evaluación automática de respuestas cortas.* Para posteriormente, *Implementar en una herramienta tecno-pedagógica las estrategias de evaluación automática identificadas para respuestas cortas.* Al respecto, podemos mencionar que, como resultado de una revisión exhaustiva de la literatura, se eligió utilizar una tecnología desarrollada en la comunidad de software libre que se integra al sistema gestor de aprendizaje MOODLE a través de un plugin llamado *Pattern Match*. Con esta integración, ya se pueden construir preguntas en formato de respuesta libre con el potencial de calificarse automáticamente. Esta tecnología está basada en un algoritmo de coincidencia de palabras que busca palabras clave y considera sinónimos y errores ortográficos; sin embargo, no utiliza la gramática o la sintaxis en sus algoritmos. A pesar de estas limitaciones, la elección de esta herramienta ofrece una solución práctica y accesible para la evaluación automática de respuestas cortas.

Es importante destacar que durante nuestra revisión no se encontraron reportes previos del uso de este plugin en idioma español. Esto representa una contribución significativa a la práctica docente, ya que este documento se convierte en una referencia valiosa para profesores interesados en implementar evaluaciones más innovadoras sin necesidad de ser expertos en procesamiento del lenguaje natural. Esta investigación no solo enriquece la literatura existente en aspectos teóricos, sino que también ofrece un modelo práctico para otros educadores que deseen mejorar sus métodos de evaluación utilizando tecnologías accesibles y eficaces.

Una vez comprendidos los aspectos técnicos para la creación de preguntas en formato de respuesta libre y su correspondiente calificación automática, se procedió al desarrollo del instrumento de investigación en la plataforma MOODLE. Este instrumento incluye 27 preguntas diseñadas para evaluar conceptos básicos de física, basadas en el Inventario de Conceptos de Física (FCI) de Hestenes et al. (1992). Este proceso implicó la creación de reglas de correspondencia específicas para cada pregunta, las cuales controlan la calificación automática.

Podemos concluir que esta etapa del trabajo corresponde al logro de los dos primeros objetivos específicos planteados, marcando un avance significativo en el uso de herramientas tecno-pedagógicas para la evaluación automática en el contexto educativo.

En la segunda etapa de la investigación, se abordó el tercer objetivo específico: *Adaptar una escala de medición para evaluar la Aceptación de la Evaluación Basada en Computadora*. Para ello, se realizó un estudio piloto de la versión en español (Apéndice A) del instrumento asociado al Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM). Los resultados sugieren que tanto las subescalas como la versión completa del instrumento en español muestran datos de confiabilidad adecuados para su aplicación en estudiantes de una universidad privada en México. La versión en español reveló índices de consistencia interna adecuados y congruentes con los reportados para la versión original en inglés de Terzis y Economides (2011), indicando un nivel de homogeneidad y coherencia entre los ítems de la escala. El total de la versión en español del CBAAM alcanzó un coeficiente de consistencia alto y estadísticamente significativo, aunque la subescala de expectativa de meta mostró excepciones.

Para la versión final del instrumento y el estudio principal, se utilizó el enfoque de modelado de ecuaciones estructurales (SEM). El análisis del modelo de medición evaluó la fiabilidad y validez de los constructos, revelando que, aunque algunos ítems presentaron cargas bajas y requieren revisión, constructos clave como Condiciones Facilitadoras, Facilidad de Uso Percibida, Intención de Uso y Percepción Lúdica demostraron altos valores de carga, confirmando su validez en el nuevo contexto de esta



investigación. Sin embargo, Facilidad de Uso Percibida en el grupo experimental e Influencia Social en ambos grupos mostraron métricas más bajas, sugiriendo la necesidad de una revisión más detallada.

La validez convergente y discriminante fue de manera general aceptable, aunque algunos ítems presentaron problemas que requieren ajustes adicionales. En resumen, el análisis de la escala de medición para la EBC mediante SEM mostró resultados positivos en términos de fiabilidad y validez, proporcionando una base sólida para el análisis de las hipótesis planteadas en el modelo estructural.

El cumplimiento de este objetivo en específico, consideramos que tiene implicaciones prácticas muy importantes ya que aporta un instrumento con evidencias de confiabilidad y validez para la medición del Uso y Aceptación de la Evaluación Basada en Computadora en el contexto educativo mexicano. Adicionalmente, permite a los investigadores conocer las formas en que los estudiantes perciben el uso de las TIC específicamente aplicadas al proceso de evaluación.

Por otra parte, desde el punto de vista teórico, al tratarse de una adaptación de un instrumento ya publicado en otro idioma y contexto, permite la comparación de los constructos e indicadores observables, ya que una de las limitaciones más importantes de este tipo de estudios es que se utilizan muestras muy específicas de estudiantes para responder sobre sus creencias. Esta es una de las posibles razones a las que se encontró que ciertas hipótesis fueron validadas en el contexto de estudio de Terzis y Economides (2011), pero no en el contexto de esta investigación realizada en 2022. Adicionalmente, la época en la que se realizó cada estudio pudo influir notablemente en los resultados, debido a que la familiaridad y la predisposición hacia la tecnología han evolucionado a lo largo del tiempo.

Para la etapa final del trabajo se planteó el cuarto objetivo específico: *Analizar los factores que afectan la intención de los estudiantes al utilizar la herramienta tecnopedagógica para la evaluación automática de preguntas de respuesta corta*. Este análisis

se centró en el papel de la Autoeficacia Informática, el Contenido, las Condiciones Facilitadoras, la Expectativa de Meta, la Facilidad de Uso Percibida, la Influencia Social, la Percepción Lúdica y la Utilidad Percibida en la intención de uso de la EBC por parte de los estudiantes. El modelo CBAAM aplicado en este estudio planteaba 15 hipótesis. Los resultados del análisis del modelo estructural revelan que tanto el grupo experimental como el grupo control mostraron un número similar de hipótesis aceptadas, con 7 hipótesis confirmadas en cada grupo (ver tabla 28). Sin embargo, las áreas de influencia y las relaciones significativas presentan diferencias notables entre los dos grupos.

En ambos grupos, la percepción lúdica demostró tener un efecto positivo significativo en la intención de uso de la herramienta de evaluación basada en computadora (EBC). Esto destaca la importancia de diseñar herramientas que no solo sean funcionales, sino también atractivas y entretenidas para los usuarios. Además, tanto la expectativa de meta como el contenido de la herramienta resultaron ser factores cruciales. La expectativa de meta tiene un impacto positivo en la percepción lúdica, mientras que el contenido influye positivamente en la formación de expectativas. Estos hallazgos subrayan la necesidad de alinear los objetivos educativos con las funcionalidades de la EBC y de asegurar que el contenido sea relevante y de alta calidad.

Las diferencias entre los grupos experimental y control son notables. En el grupo experimental, se observó que la utilidad percibida tiene un efecto positivo en la percepción lúdica, sugiriendo que los usuarios que perciben la herramienta como útil también la encuentran más divertida. Además, la autoeficacia informática influye positivamente en la facilidad de uso percibida, lo que destaca la importancia de la confianza en las habilidades informáticas. También se encontró que la influencia social tiene un impacto positivo en la utilidad percibida, indicando que la percepción de utilidad de la herramienta está influenciada por la opinión de otros. Sin embargo, un hallazgo inesperado fue que el contenido tiene un efecto negativo en la intención de uso en el grupo experimental, lo cual podría indicar problemas específicos con el contenido percibido por estos estudiantes.

Por otro lado, en el grupo control, la facilidad de uso percibida tiene un impacto significativo en varios aspectos: intención de uso, utilidad percibida y percepción lúdica. Esto sugiere que los usuarios de este grupo valoran más la simplicidad y accesibilidad de la herramienta. La facilidad de uso se presenta como un componente esencial para la aceptación y uso de la EBC en contextos donde se utilizan formatos de preguntas más tradicionales.

La justificación de estos resultados radica en varios factores. En primer lugar, el modelo aplicado CBAAM fue desarrollado en 2011 y esta investigación se llevó a cabo en 2022, un periodo en el que las tecnologías y las expectativas de los usuarios han evolucionado significativamente. Esto puede explicar por qué algunas hipótesis originales no se confirmaron y por qué surgieron nuevas relaciones significativas. Además, el contexto específico de nuestra investigación, incluyendo las características de los estudiantes y el entorno educativo, podría haber influido en los resultados obtenidos.

Estos hallazgos son relevantes ya que proporcionan una visión actualizada sobre los factores que afectan la intención de uso de herramientas tecno-educativas para la evaluación automática. Ofrecen también una base para futuras investigaciones y desarrollos en el campo, asegurando que las herramientas se adapten mejor a las necesidades y expectativas de las y los estudiantes actuales. Al comprender mejor estas dinámicas, los educadores y desarrolladores pueden crear herramientas más eficaces y atractivas, promoviendo una adopción más amplia y efectiva en diversos contextos educativos.

Este estudio representa una contribución significativa al campo de la innovación educativa, particularmente en la evaluación basada en computadora dentro de la educación superior. Una de las aportaciones más destacadas es la validación empírica de la influencia que el tipo de instrumento de evaluación tiene sobre la aceptación y percepción de los estudiantes, según el modelo CBAAM. La diferenciación en la aceptación entre preguntas de respuesta corta y preguntas de opción múltiple subraya la importancia de la metodología de evaluación en la percepción estudiantil sobre

factores clave como la Percepción Lúdica, la Utilidad Percibida y la Facilidad de Uso Percibida. Estos hallazgos sugieren que la elección del formato de evaluación no solo afecta la comprensión conceptual, sino también la disposición de los estudiantes hacia la evaluación misma.

Desde un enfoque práctico, este trabajo ofrece perspectivas valiosas para profesores y diseñadores instruccionales en entornos de educación superior, al destacar la necesidad de seleccionar cuidadosamente los instrumentos de evaluación automatizada en función de los objetivos pedagógicos y las preferencias de los estudiantes. Además, la integración exitosa de la tecnología de *Pattern Match* dentro del sistema MOODLE y el desarrollo de un instrumento de evaluación que abarca conceptos básicos de física demuestran la viabilidad técnica y pedagógica de implementar evaluaciones de respuesta corta calificadas automáticamente.

El estudio también aporta a la comunidad académica una herramienta validada en español para evaluar la aceptación de la evaluación basada en computadora, enriqueciendo el cuerpo de recursos disponibles para la investigación futura y la práctica educativa. Este instrumento, que ha demostrado confiabilidad y validez, facilita la exploración de la percepción estudiantil sobre el uso de TIC en procesos de evaluación, abriendo camino para investigaciones comparativas entre diferentes contextos y culturas educativas.

Adicionalmente, los hallazgos resaltan que, para fomentar una mayor aceptación de las evaluaciones basadas en computadora, es crucial que estas sean percibidas como lúdicas, útiles y fáciles de usar por parte de los estudiantes. Esto indica que, más allá de la selección del tipo de pregunta, la experiencia general del usuario juega un papel fundamental en la disposición a utilizar estas tecnologías. Por lo tanto, este estudio no solo profundiza nuestra comprensión sobre la adopción de evaluaciones electrónicas, sino que también guía el diseño de futuras herramientas de evaluación para maximizar su efectividad y aceptación.

El estudio concluye que un sistema es más probable que sea utilizado por los estudiantes si es lúdico y la EBC es más probable que sea lúdica cuando es fácil de usar y útil. Finalmente, el modelo de aceptación estudiado para la evaluación basada en computadora explica aproximadamente el 47% de la varianza de la Intención para Usar EBC con preguntas de opción múltiple y del 77% para el caso de la EBC con preguntas de respuesta corta. Por lo tanto, se recomienda investigar otras variables que afectan la Intención de Uso y aplicar el modelo en otros contextos para mayor confirmación.

### ***Trabajo futuro***

Un trabajo a futuro podría expandir este estudio a una variedad más amplia de contextos educativos, disciplinas y niveles académicos para examinar si los patrones de aceptación de la evaluación basada en computadora (EBC) y el impacto de los distintos formatos de evaluación se mantienen o varían.

Además, sería valioso explorar otros aspectos, tales como variables demográficas y académicas específicas, como el género, la edad, el historial académico y la familiaridad previa con la tecnología, afectan la aceptación de la EBC. Esta línea de investigación ayudaría a identificar posibles diferencias en la aceptación y uso de la herramienta, ofreciendo así una base para desarrollar estrategias de implementación más inclusivas y efectivas.

Una recomendación fundamental que surge de este trabajo, considerando la adopción del Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM) de Terzis y Economides (2011), es desarrollar un modelo innovador que aborde específicamente el impacto del formato de las preguntas en la adopción y efectividad de la evaluación electrónica. El análisis exhaustivo realizado ha identificado una oportunidad significativa para comprender mejor cómo el diseño de las preguntas puede influir en la percepción de los usuarios sobre la utilidad, la facilidad de uso y la percepción lúdica de las herramientas de evaluación electrónica.

Este nuevo modelo debería incorporar constructos que examinen detalladamente las características específicas de los formatos de preguntas, como preguntas de respuesta corta frente a preguntas de opción múltiple, y su relación con los procesos cognitivos y

emocionales de los estudiantes. Al hacerlo, se podrán ofrecer recomendaciones más precisas y fundamentadas sobre cómo diseñar evaluaciones electrónicas que no solo mejoren el rendimiento académico, sino que también aumenten el compromiso y la satisfacción del estudiante.

Otro estudio futuro podría centrarse en crear una lista de directrices para evaluar las soluciones de software de evaluación electrónica y comparar las funciones que ofrecen diferentes proveedores. Esto ayudaría a las instituciones de educación superior a elegir el proveedor y el software más adecuado. La lista de pautas debería incluir secciones sobre tipos de preguntas y respuestas, seguridad de la evaluación, posibles problemas técnicos y requisitos de capacitación.

En resumen, análisis futuros deberían centrarse en la ampliación del estudio a diversos contextos y formatos de evaluación, soluciones de software, así como en el desarrollo de un modelo que integre las características específicas de los formatos de preguntas y su impacto en la percepción de los estudiantes. Estos esfuerzos contribuirán significativamente a la mejora de las prácticas de evaluación y a la promoción de herramientas tecnológicas en la educación superior.

## Referencias

- Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, 24, 665–694.
- Agarwal, R., & Prasad, J. (1999). Are individual differences germane to the acceptance of new information technologies. *Decision Sciences*, 30(2), 361-391.
- Agarwal, R., Sambamurthy, V., & Stair, R. M. (2000). Research report: the evolving relationship between general and specific computer self-efficacy and an empirical assessment. *Information Systems Research*, 11(4), 418–430.
- Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
- Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior 1. *Journal of applied social psychology*, 32(4), 665-683.
- Al-Qdah, M., & Ababneh, I (2017). Comparing online and paper exams: Performances and perceptions of Saudi students. *International Journal of Information and Education Technology*, 7(2), 106.
- Appiah, M., & Van Tonder, F. (2018). E-assessment in higher education: A review. *International Journal of Business Management & Economic Research*, 9(6), 1454-1460.
- Arias, F. G. (2006). *El proyecto de investigación* (5ta. Ed.). Caracas, Venezuela: Editorial Episteme.
- Bacon, D. R. (2003). Assessing learning outcomes: A comparison of multiple-choice and short answer questions in a marketing context. *Journal of Marketing Education*, vol. 25, no. 1, pp. 31-36.
- Bacon, F. (1620). *Novum Organum*.
- Bagozzi, R. P. (2007). The Legacy of the Technology Acceptance Model and a Proposal for a Paradigm Shift. *Journal of the Association for Information Systems*, 8(4), 243-254.
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the academy of marketing science*, 16(1), 74-94.

- Bailey, J. M., Johnson, B., Prather, E. E. and Slater, T. F. (2012). Development and Validation of the Star Properties Concept Inventory. *International Journal of Science Education*, vol. 34, no. 14, pp. 2257-2286.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2), 191.
- Bandura, A. (1986). *Social Foundations of Thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Beichner, R. J. (2014). History and evolution of active learning spaces. *New directions for teaching and learning*, 2014(137), 9-16.
- Bell, B. & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85, 536-553.
- Ben-Shakter, G. & Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of different guessing tendencies, *Journal of Educational Measurement*, vol. 28, no. 1., pp. 23-25.
- Betts, L. R. and Elder, T. J., Hartley, J. and Truman, M. (2009). Does correction for guessing reduce students' performance on multiple-choice examinations? Yes? No? Sometimes? . *Assessment and Evaluation in Higher Education*, vol. 34, no. 1, pp.1-15.
- Biggs, J. & Tang, C. (2011). *Teaching for quality learning at university: what the student does*. 4th ed. Berkshire: Open University Press.
- Birenbaum, M. & Tatsuoka, K. K. (1987). Open-ended versus multiple choice response formats - it does make a difference for diagnostic purposes. *Applied Psychological Measurement*, 11(4), 385-395.
- Bridgeman, B. (1991). A comparison of open-ended and multiple-choice questionformats for the quantitative section of the Graduate Record Examinations General Test, *ETS Research Report Series*, vol. 2, pp. 1-25.
- Bruyn, E. de, Mostert, E., & Schoor, A. (2011). Computer-based Testing - The Ideal Tool to Assess on the Different Levels of Bloom's Taxonomy. In 14th International Conference on Interactive Collaborative Learning (ICL2011) (p. 444 - 449). Issue: September, Publisher: IEEE. J. Sheard (Eds.), *Proceedings of the 2nd melbourne computing education conventicle* (pp. 1-8). Melbourne



- Bueno, S., & Salmeron, J. L. (2008). TAM-based success modeling in ERP. *Interacting with Computers*, 20(6), 515–523.
- Bukie, O.F. (2014). Understanding technologies for e-assessment: A systematic review approach. *Journal of Emerging Trends in Computing and Information Sciences*, 5(12), 936-947.
- Bull, J. and McKenna, C. (2003) *Blueprint for computer-aided assessment*, London, RoutledgeFalmer.
- Burrows, S., Gurevych, I., & Stein, B. (2015). The eras and trends of automatic short answer grading. *International Journal of Artificial Intelligence in Education*, 25(1), 60-117.
- Burton (2005). Multiple-choice and true/false tests: myths and misapprehensions. *Assessment and Evaluation in Higher Education*, vol. 30, no. 1, pp. 65-72.
- Butcher, P. (2008). Online assessment at the Open University us-ing open source software: moodle, openmark and more. IN: Khandia, F. In 12th CAA International Computer Assisted Assessment Conference: Proceed-ings of the Conference on 8th and 9th July (pp. 65-78)
- Butcher, P.G., & Jordan, S.E. (2010). A comparison of human and computer marking of short free-text student responses. *Computers & Education*, 55(2), 489-499.
- Carter, M. (2007). *Ways of Knowing, Doing, and Writing in the Disciplines*. *College Composition and Communication*, 385-418.
- Cheng, E. W. (2019). Choosing between the theory of planned behavior (TPB) and the technology acceptance model (TAM). *Educational Technology Research and Development*, 67(1), 21-37.
- Chi, M.T., Feltovich, P.J., & Glasner, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research*, 295(2), 295-336.
- Choudrie, J., & Dwivedi, Y. K. (2005). A survey of citizens' awareness and adoption of e-government initiatives, the 'government gateway': A United Kingdom perspective. *Iseing. Org*, 5, 1-13, vol. 104, no. 4, pp. 454-496.

- Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of physics*, 50(1), 66-71.
- Comer, D. K., Clark, C. R., & Canelas, D. A. (2014). Writing to Learn and Learning to Write Across the Disciplines: Peer-to-Peer Writing in Introductory-Level MOOCs. *International Review of Research in Open and Distance Learning*, 15(5), 26-82.
- Comisión Nacional para la Protección de los Sujetos Humanos de Investigación Biomédica y del Comportamiento. (1979). Informe Belmont: Principios éticos y pautas para la protección de los sujetos humanos de investigación. Departamento de Salud, Educación y Bienestar de los EE.UU. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html>
- Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: development of a measure and Initial test. *MIS Quarterly*, 19(2), 189-211.
- Comte, A. (1830). *Cours de Philosophie Positive*.
- Conner, M., & Armitage, C. J. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology*, 28(15), 1429-1464.
- Conole, G., & Warburton, B. (2005). A review of computer-assisted assessment. *ALT-J, Research in Learning Technology*, 13(1), 17-31. <https://doi.org/10.1080/0968776042000339772>
- Craven, P. (2009). History and challenges of e-assessment: the "Cambridge Approach" perspective - e-assessment research and development 1989 to 2009. *Cambridge Assessment*, pp. 1-11.
- Creswell, J. W. (2013). *Qualitative Inquiry & Research Design: Choosing among Five Approaches* (3rd ed.). Thousand Oaks, CA: SAGE.
- Crisp, G. (2007) *The e-assessment Handbook*, London, Continuum.
- Crisp, G. (2011). *Teacher's handbook on e-Assessment: A handbook to support teachers in using e-assessment to improve and evidence student learning and outcomes*. San Francisco, California: Creative Commons.
- Crotty, M. (1998) *The Foundations of Social Research*. London: Sage.
- Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. San Francisco: Jossey-Bass.

- Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340
- Davis, F. D. (2011). Foreword in technology acceptance in education: Research and issues. Sense Publishers.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology*, 22, 1111-1132.
- Davis, M. M., Spohrer, J. C., & Maglio, P. P. (2011). Guest editorial: How technology is changing the design and delivery of services. *Operations Management Research*, 4(1-2), 1-5. <https://doi.org/10.1007/s12063-011-0046-6>
- Deane Sorcinelli, M., & Elbow, P. (1997). *Writing to Learn: Strategies for Assigning and Responding to Writing Across the Disciplines*. San Francisco, California: Jossey-Bass.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-Determination in Human behavior*. New York: Plenum Press.
- Del Gobbo, E., Guarino, A., Cafarelli, B., Grilli, L., & Limone, P. (2023). Automatic evaluation of open-ended questions for online learning. A systematic mapping. *Studies in Educational Evaluation*, 77
- Dick-Perez, M., Luxford, C. J., Windus, T. L. and Holme, T. (2016). A Quantum Chemistry Concept Inventory for Physical Chemistry Classes. *Journal of Chemical Education*, vol. 93, no. 4, pp. 605-612.
- Dillon, A., & Morris, M. G. (1996). User acceptance of new information technology: theories and models. *Annual review of information science and technology*.
- Ding, L., & Caballero, M. D. (2014). Uncovering the hidden meaning of cross-curriculum comparison results on the Force Concept Inventory. *Physical Review Special Topics-Physics Education Research*, 10(2).
- Doll, W. J., & Torkzadeh, G. (1988). The measurement of end-user computing satisfaction. *MIS Quarterly*, 12(2), 259–274.
- Doolin, B. (1996). Alternative views of case research in information systems. *Australasian Journal of Information Systems*, 3(2).

- Downing, S. M. (2003). Guessing on selected-response examinations, *Medical Education*, vol. 37, no. 8, pp. 670-671.
- Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education*, 11(5), 481-490.
- Dwivedi, Y. K., & Williams, M. D. (2008). Demographic influence on UK citizens' e-government adoption. *Electronic Government: an International Journal*, 5(3), 261-274.
- Elbow, P. (1994). *Writing for Learning -- Not Just for Demonstrating Learning*. Amherst: University of Massachusetts, Amherst, National Teaching and Learning Forum.
- Emig, J. (1977). Writing as a Mode of Learning. *College Composition and Communication*, 28(2), 122-128.
- Farrell, T. & Rushby, N. (2016). Assessment and learning technologies: an overview. *British Journal of Educational Technology*, 47(1):106-120.
- Ferrao, M. (2010). E-assessment within the Bologna paradigm: evidence from Portugal. *Assessment and Evaluation in Higher Education*, vol. 35, no. 7, pp. 819-830.
- Fishbein, M., & Ajzen I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research, 181-202.
- Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing research*, 440-452.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing research*, 39-50.
- Forsman, S. (1985). Writing to Learn Means Learning to Think. *Roots in the Sawdust*, 162-174.
- Funk, S. C. and Dickson, K. L. (2011). Multiple-choice and short-answer exam performance in a college classroom, *Teaching of Psychology*, vol. 38, no. 4, pp. 273-277.
- Garson, G. D. (2016). *Partial least squares: Regression and structural equation models*: Statistical Associates Publishers.
- Geisser, S. (1974). A Predictive Approach to the Random Effect Model. *Biometrika*, 61(1), 101-107. doi:10.2307/2334290

- Gibbs, G. (2006). Why Assessment is changing. In *Innovative Assessment in Higher education*, pp.11-22. Routledge.
- Gipps, C. and Murphy, P. (1994). *A fair test? Assessment, achievement, and equality*, Open University Press, Buckingham.
- Glynn, S.M. & Muth, K.D. (1994). Reading and writing to learn science: Achieving scientific literacy. *Journal of Research in Science Teaching*, 31(9), 1057-1073.
- Granic, A., & Marangunic, N. (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology*, 50(5), 2572-2593. <https://doi.org/10.1111/bjet.12864>
- Gray, K., Rebello, N. S. and Zollman, D. (2002). The effect of Question Order on Responses to Multiple-choice Questions. 2002 Physics Education Research Conference.
- Gwinnett, C. & Cassella, J. (2011). The trials and tribulations of designing and utilising MCQs in HE and for assessing forensic practitioner competency. *New Directions in the Teaching of Physical Sciences*, vol. 7, pp. 72-78.
- Ha, M. Nehm, R. Urban-Lurain, M & Merrill, J.E. (2011). Applying computerized scoring models of written biological explanations across courses and colleges: prospects and limitations. *CBE-Life Sciences Education*, 10, 379-393.
- Haenlein, M., & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding statistics*, 3(4), 283-297.
- Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*: Sage Publications.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing theory and Practice*, 19(2), 139-152. doi:10.2753/MTP1069-6679190202
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2012). Editorial-Partial Least Squares: The Better Approach to Structural Equation Modeling? *Long Range Planning*, 45(5-6), 312-319.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). *Partial least squares structural equation modeling: rigorous applications, better results and higher acceptance* | NOVA. The University of Newcastle's Digital Repository.

- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 66(1), 64-74.
- Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American journal of Physics*, 53(11), 1043-1055.
- Halloun, I., Hake, R., Mosca, E. and Hestenes, D. (1995). Force Concept Inventory (revised 1995), in Mazur, E. (1997) *Peer Instruction: A Users Manual*, Prentice-Hall, New Jersey.
- Harlen, W., & James, M. (1997). Assessment and Learning: Differences and Relationships Between Formative and Summative Assessment. *Assessment in Education*, 4(3), 365-379.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2012). Using partial least squares path modeling in advertising research: basic concepts and recent issues. *Handbook of research on international advertising*, 252.
- Hepplestone, S. & Helm, P. (2003). Strategies for coping with unexpected uptake of CBA. *Proceedings of the 7th International Computer Assisted Assessment Conference*, Loughborough, pp. 59-167. <http://caaconference.co.uk/pastConferences/2003/proceedings/hepplestone.pdf>
- Hernández, R., Fernández, C., & Baptista, P. (2010). *Metodología de la Investigación* (5ta. ed.). México: McGraw-Hill.
- Hernández-Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta* (Mc Graw Hill (ed.); 6a ed., Vol. 1). <https://doi.org/10.1017/CBO9781107415324.004>
- Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller. *Physics Teacher*, 33(8), 502-504.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, 30(3), 141-158.
- Howarth, P. (2015). The opportunities and challenges faced in utilizing e-Based assessment. <http://www.educationalrc.org/oldconf/old/pdf/Paul%20Howarth%20-%20Beirut%20Presentation.pdf>

- Hox, J., & Bechger, T. (1998). An introduction to structural equation modelling. *Family Science Review*, 11(354-373).
- Hu, P. J., Chau, P. Y. K., Sheng, O. R. L., & Tam, K. Y. (1999). Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of Management Information Systems*, 16(2), 91–112.
- Hudson, R. D. (2010). Multiple-choice questions compared to short-answer responses: Which assess understanding of chemistry more effectively? PhD dissertation, Curtin University, Australia.
- Huffman, D., & Heller, P. (1995). What Does the Force Concept Inventory Actually Measure?. *Physics Teacher*, 33(3), 138-43.
- Hughes, D. L., Dwivedi, Y. K., & Rana, N. P. (2017). Mapping IS failure factors on PRINCE2® stages: An application of interpretive ranking process (IRP). *Production Planning & Control*, 28(9), 776-790.
- Hughes, D. L., Dwivedi, Y. K., Rana, N. P., & Simintiras, A. C. (2016). Information systems project failure-analysis of causal links using interpretive structural modelling. *Production Planning & Control*, 27(16), 1313-1333.
- Hughes, D. L., Rana, N. P., & Dwivedi, Y. K. (2020). Elucidation of IS project success factors: an interpretive structural modelling approach. *Annals of Operations Research*, 285(1), 35-66.
- Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic management journal*, 20(2), 195-204.
- Hunt, T. J. (2012). Computer-marked assessment in Moodle: Past, present and future. In *Proceedings of the 2012 International Computer Assisted Assessment (CAA) Conference*, Southampton, 10th-11th July 2012.
- Hussein, M. A., Hassan, H., & Nassef, M. (2019). Automated language essay scoring systems: A literature review. *PeerJ Computer Science*, 5, Article e208. <https://doi.org/10.7717/peerj-cs.208>
- Imtiaz, M. A., & Maarop, N. (2014). A Review of Technology Acceptance Studies in the Field of Education. *Jurnal Teknologi*, 69(2). doi:10.11113/jt.v69.3101
- Isaias, P., Miranda, P., & Pífano, S. (2017). Framework for the analysis and comparison of e-assessment systems. In *ASCILITE 2017-Conference Proceedings-34th*

- International Conference of Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education (pp. 276-283). Australasian Society for Computers in Learning in Tertiary Education (ASCILITE).
- Itza-Ortiz, S. F., Rebello, N. S. and Zollman, D. (2003). Students' Models of Newtons Second Law in Mechanics and Electromagnetism. *European Journal of Physics*, vol. 25.
- Jamil, M, Tariq, R, & Shami, P (2012). Computer-Based vs Paper-Based Examinations: Perceptions of University Teachers. *Turkish Online Journal of Educational Technology-TOJET*, 11(4), 371-381.
- JISC (Joint Information Systems Committee). (2006). Effective use of VLEs: e-assessment. [www.jiscinfonet.ac.uk](http://www.jiscinfonet.ac.uk)
- JISC (Joint Information Systems Committee). (2007). Effective practice with e-assessment: An overview of technologies, policies and practice in further and higher education.
- Johnstone, A. H., & Mughol, A. R. (1976). Concepts of Physics at Secondary Level. *Physics Education*, 11(7), 466-469.
- Jordan, S. & Mitchell, T. (2009). E-Assessment for learning? The potential of short-answer free-text questions with tailored feedback. *British Journal of Educational Technology*, vol. 40, no. 2, pp. 371-385.
- Jordan, S. (2011). Using Interactive Computer-based Assessment to Support Beginning Distance Learners of Science. *Open Learning: The Journal of Open, Distance and e-Learning*, 26(2):147-164.
- Jordan, S. (2013). E-assessment: past, present, and future. *New Directions in the Teaching of Physical Sciences*, vol. 9, no. 1, pp. 87-106.
- Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43(4), 443-477.
- Jöreskog, K. G. (1982). The LISREL approach to causal model-building in the social sciences. *Systems under indirect observation, Part I*, 81-100.
- Jorion, N., Gane, B. D., James, K., Schroeder, L., Dibello, L. V. and Pellerino, J. W. (2015). An Analytic Framework for Evaluating the Validity of Concept Inventory inventory claims. *Journal of Engineering Education*, 104(4), 454-496.



- Jurafsky, D. and Martin, J. "Speech and Language Processing" Prentice Hall, 2 edition, 2008.
- Kaldi, A., Aghaie, A., & Khoshalhan, F. (2008, December). KMS adoption in organizations. In 2008 IEEE International Conference on Industrial Engineering and Engineering Management (pp. 37-41). IEEE.
- Karahanna, E., & Straub, D. W. (1999). The psychological origins of perceived usefulness and ease of use. *Information and Management*, 35, 237–250.
- Kaur, G. (2014). Usage of regular expressions in NLP. *International Journal of Research in Engineering and Technology IJERT*, 3(01), 7.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Keys, C. W. (1999). Revitalizing instruction in scientific genres: Connecting knowledge production with writing to learn in science. *Science Education*, 83(2), 115-130.
- Kiryakova, G. (2021). E-assessment-beyond the traditional assessment in digital environment. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1031, No. 1, p. 012063). IOP Publishing.
- Kitchenham, B., & Pfleeger, S. L. (2002). Principles of survey research: part 5: populations and samples. *SIGSOFT Softw. Eng. Notes*, 27(5), 17-20. doi:10.1145/571681.571686
- Klein, R., Kryrilov, A. & Tokman, M. (2011). Automated assessment of short free-text responses in computer science using latent semantic analysis. *Proceedings of the 16th annual joint conference on innovation and technology in computer science education*, pp. 158-162.
- Korpelainen, E. (2011). Theories of ICT system implementation and adoption: A critical review.
- Kuechler, W.L. & Simkin, M.G. (2010). Why is performance on multiple choice tests and constructed response tests not more closely related? Theory and an empirical test. *Decision Sciences Journal of Innovative Education*, 8, 55-73.
- Kuh, G. D. (2008). Excerpt from *High-Impact Educational Practices: What They Are, Who Has Access to Them, and Why They Matter*. *Association of American Colleges and Universities*, 14(3), 28-29.

- Kuikka, M., Kitola, M., & Laakso, M. J. (2014). Challenges when introducing electronic exam. *Research in Learning Technology*, 22.
- Kundu, A., & Bej, T. (2020). Experiencing e-assessment during COVID-19: An analysis of Indian students' perception. *Higher Education Evaluation and Development*, 15(2), 114-134. <https://doi.org/10.1108/heed-03-2021-0032>
- Lachner, A., Burkhart, C., & Nückles, M. (2017). Formative Computer-Based Feedback in the University Classroom: Specific Concept Maps Scaffold Students' Writing. *Computers in Human Behavior*, 72(Supplement C), 459-469.
- Landry, B. J. L., Griffeth, R., & Hartman, S. (2006). Measuring student perceptions of blackboard using the technology acceptance model. *Decision Sciences Journal of Innovative Education*, 4(1), 87-99.
- Lasry, N., Rosenfield, S., Dedic, H., Dahan, A., & Reshef, O. (2011). The puzzling reliability of the Force Concept Inventory. *American Journal of Physics*, 79(9), 909-912.
- Leacock, C. & Choddorow, M. (2003). C-rater: Automated Scoring of Short-Answer Questions. *Computers and Humanities*, vol. 37, no. 4, pp. 389-405.
- Lee, Y. C. (2008). The role of perceived resources in online learning adoption. *Computers and Education*, 50(4), 1423-1438.
- Liao, H. L., & Lu, H. P. (2008). The role of experience and innovation characteristics in the adoption and continued use of e-learning websites. *Computers and Education*, 51(4), 1405-1416.
- Lindell, R. S., Peak, E. and Foster, T. M. (2007). Are they all created equal? A comparison of different concept inventory development methodologies. *AIP Conference Proceedings*, pp. 14-17
- Lu, J., Liu, C., Yu, C., & Wang, K. (2008). Determinants of accepting wireless mobile data services in China. *Information & Management*, 45(1), 52-64.
- Lu, J., Yu, C., Liu, C., & Yao, J. E. (2003). Technology acceptance model for wireless internet. *Internet Research: Electronic Networking Applications and Policy*, 13, 206-222.
- Madsen, A., McKagan, S. B., & Sayre, E. C. (2017). Best practices for administering concept inventories. *The Physics Teacher*, 55(9), 530-536.

- Majchrzak, T. A., & Usener, C. A. (2011). Evaluating the Synergies of Integrating E-Assessment and Software Testing. In Proceedings of Information Systems Development Conference (ISD2011). Springer, 2011.
- Malone, T. W. (1981a). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, 4, 333–369.
- Malone, T. W. (1981b). What makes computer games fun? *Byte*, December, 258–276.
- Maqableh, M., & Mohammed, A. B. (2015). The acceptance and use of computer based assessment in higher education. *Journal of Software Engineering and Applications*, 8(10), 557-574.
- Marangunic, N., & Granic, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society*, 14, 81-95. <https://doi.org/10.1007/s10209-014-0348-1>
- Martinez, M. (1991). A comparison of multiple-choice and constructed response figural items. *Journal of Educational Measurement*, 28(2), 131-145.
- Martinez, M. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34(4), 207-218.
- Marzano, R. J. (1993). How classroom teachers approach the teaching of thinking. *Theory Into Practice*, 32(3), 154-160.
- Mathews, J. (2006). Just whose idea was all this testing?, *The Washington Post*, 14th November 2006.
- Mazur, E. (1997, March). Peer instruction: Getting students to think in class. In AIP conference proceedings (pp. 981-988). IOP INSTITUTE OF PHYSICS PUBLISHING LTD.
- McCloskey, M. (1983). Intuitive Physics, *Scientific American*, vol. 248, no. 4, pp.122-130.
- Michael, J. (2007). What makes physiology hard for students to learn? Results of a faculty survey. *Advances in Physiology Education*, 31, 34-40.
- Mitchell, T., Russell, T., Broomhead, P. & Aldridge, N. (2002). Towards robust computerised marking of free-text responses. 6th International CAA Conference, Loughborough, United Kingdom.
- Monecke, A., & Leisch, F. (2012). semPLS: structural equation modeling using partial least squares.

- Mooi, E., & Sarstedt, M. (2011). *A Concise Guide to Market Research The Process, Data, and Methods Using IBM SPSS Statistics*: Berlin: Springer Google Scholar.
- Moon, J., & Kim, Y. (2001). Extending the TAM for a world-wide-web context. *Information and Management*, 38(4), 217-230.
- Moore, G. & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, vol. 2, núm. 3.
- Morton, J. (2007). Authenticity in the IELTS Academic Module Writing Test: a Comparative Study of Task 2 Items and University Assignments. *IELTS Collected Papers: Research in Speaking and Writing Assessment*, 1, 197-248.
- Muñiz, J., Elosua, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25(2), 151-157.
- Nehm, R.H. & Haertig, H. (2012) Human vs. computer diagnosis of students' natural selection knowledge: testing the efficacy of text analytic software. *Journal of Science Education and Technology*, 21, 56-73.
- Newell, G. (2006). Writing to learn: How alternative theories of school writing account for student performance. In *Handbook of Writing Research*. Eds. pp. 235-246.
- Ngai, E. W. T., Poon, J. K. L., & Chan, Y. H. C. (2007). Empirical examination of the adoption of WebCT using TAM. *Computers and Education*, 48(2), 250-267.
- Nicholls, J. G. (1984). Achievement motivation: conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91, 328–346.
- Nicol, D. (2007) 'E-assessment by design: using multiple-choice tests to good effect', *Journal of Further and Higher Education*, vol. 31, no. 1, 53-64.
- Nikou, S. A., & Economides, A. A. (2017). Mobile-based assessment: Investigating the factors that influence behavioral intention to use. *Computers & Education*, 109, 56-73.
- Noorbehbahani, F. & Karden, A. A. (2011). The automatic assessment of free text answers using a modied BLEU algorithm. *Computers and Education*, vol. 56, no. 2, pp. 337-345.

- Ong, C., & Lai, J. (2006). Gender differences in perceptions and relationships among dominants of e-learning acceptance. *Computers in Human Behaviour*, 22(5), 816-829.
- Ong, C.S., Lai, J.Y. and Wang, Y.S. (2004) Factors Affecting Engineers' Acceptance of Asynchronous E-Learning Systems in High-Tech Companies. *Information and Management*, 41, 795-804. <http://dx.doi.org/10.1016/j.im.2003.08.012>
- Padilla-Melendez, A., Garrido-Moreno, A., & Del Aguila-Obra, A. R. (2008). Factors affecting e-collaboration technology use among management students. *Computers and Education*, 51(2), 609-623.
- Page, E.B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan*, 47(5), 238-243.
- Papamitsiou, Z. K., Terzis, V., & Economides, A. A. (2014, March). Temporal learning analytics for computer based testing. In *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 31-35).
- Papineni, K., Roukas, S., Ward, T. & Zhu, W. (2001). BLEU: a method for automatic evaluation of machine translation. Technical Report RC22176 (WO109-022), IBM Research Division.
- Parker, M. A. J. (2020). Establishing Physics Concept Inventories Using Free-Response Questions. Open University (United Kingdom).
- Pérez, D., Alfonseca, E. & Rodriguez, P. (2004a). Application of the BLEU method for evaluating free-text answers in an e-learning environment. *Language Resources and Evaluation Conference (LREC-2004)*, Lisbon, Portugal.
- Pérez, D., Alfonseca, E. & Rodriguez, P. (2004b). Upper bounds and extensión of the BLEU algorithm applied to assessing student essays. *International Association for Educational Assessment Conference (IAEA-2004)*, Philadelphia, United States.
- Pérez-Marín, D., Pascual-Nieto, I., & Rodriguez, P. (2009). Computer-assisted assessment of free-text answers. *The knowledge engineering review*, vol. 24, no. 4, pp.353-374.
- Porter, L., Taylor, C. and Webb, K. (2014). Leveraging open source principles for flexible concept inventory development, *Proceedings of the 2014 conference on innovation technology in computer science education*, pp. 243-248.

- Prather, J. P. (1985). Philosophical Examination of the Problem of the Unlearning of Incorrect Science Concepts.
- Pulman, S. & Sukkarieh, J. (2005). Automatic Short Answer Marking. Proceedings of the 2nd Workshop on Building Educational Applications using NLP.
- Puthiaparampil, T., & Rahman, M. M. (2020). Very short answer questions: a viable alternative to multiple choice questions. *BMC medical education*, 20, 1-8.
- Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: a systematic literature review. *Artificial Intelligence Review*, 55(3), 2495-2527.
- Rebello, N. & Zollman, D. (2004). The effect of distractors on student performance on the force concept inventory. *American Journal of Physics*, vol. 72.
- Reed-Rhoads, T., & Imbrie, P. K. (2008). Concept inventories in engineering education. In *Proceedings from Evidence on Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education Workshop (Vol. 2, pp. 13-14)*.
- Reynolds, J. A., Thaiss, C., Katkin, W., & Thompson Jr, R. J. (2012). Writing-to-Learn in Undergraduate Science Education: a Community-Based, Conceptually Driven Approach. *CBE (Cell Biology Education)—Life Sciences Education*, 11(1), 17-25.
- Richardson, C. T. and O'Shea, B. W. (2013). Assessing gender differences in response system questions for an introductory physics course, *American Journal of Physics*, vol. 81.
- Ringle, C. M., Wedne, S., & Becker, J.-M. (2015). SmartPLS 3: SmartPLS GmbH. Recuperado de: <http://www.smartpls.com>
- Rivard, L. O. P. (1994). A Review of Writing to Learn in Science: Implications for Practice and Research. *Journal of Research in Science Teaching*, 31(9), 969-983.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., & Gould, R. R. (2010). The astronomy and space science concept inventory: development and validation of assessment instruments aligned with the k-12 national science standards. *Astronomy Education Review*, 8(1), 010111.
- Sangwin, C. J. (2013) *Computer aided assessment of mathematics*, Oxford: Oxford University Press.

- Sarstedt, M., Ringle, C. M., Smith, D., Reams, R., & Hair, J. F. (2014). Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. *Journal of Family Business Strategy*, 5(1), 105-115.
- Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson education.
- Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, 128, 13-35. <https://doi.org/10.1016/j.compedu.2018.09.009>
- Scott, T. F. & Schumayer, D. (2017). Conceptual coherence of non-Newtonian worldviews in Force Concept Inventory data. *Physical Review Physics Education Research*, 13(1), 010126.
- Scott, T. F., Schumayer, D. & Gray, A. R. (2012). Exploratory Factor Analysis of a Force Concept Inventory Data Set, *Physical Review Special Topics Physics Education Research*, vol. 8, no. 2.
- Shahbaz, M., McMinn, P., & Stevenson, M. (2012, August). Automated discovery of valid test strings from the web using dynamic regular expressions collation and natural language processing. In 2012 12th International Conference on Quality Software (pp. 79-88). IEEE.
- Shee, D. Y., & Wang, Y.-S. (2008). Multi-criteria evaluation of the web-based e-learning system: a methodology based on learner satisfaction and its applications. *Computer & Education*, 50(3), 894-905.
- Shih, H. (2008). Using a cognitive-motivation-control view to assess the adoption intention for Web-based learning. *Computer & Education*, 50(1), 327-337.
- Shuhidan, S., Hamilton, M. and DSouza, D. (2010). Instructors perspectives of multiple-choice questions in summative assessment for novice programmers, *Computer Science Education*, vol. 20, no. 3, pp. 229-259.
- Sieber, J. E. (1992). *Planning Ethically Responsible Research: A Guide for Students and Internal Review Boards*. Sage Publications.

- Simon & Snowdon, S. (2014). Multiple-choice vs free-text code-explaining examination questions, *Proceedings of the 14th Koli Calling International Conference on computing education research*, pp. 91-97.
- Smith, J. I. & Tanner, K. (2010). The problem of revealing how students think: Concept inventories and beyond. *CBE life sciences education*, vol. 9, no. 1, pp. 1-5.
- Smith, P. J., Murphy, K. L., & Mahoney, S. E. (2003). Towards identifying factors underlying readiness for online learning: an exploratory study. *Distance Education.*, 24(1), 57-67.
- Sousa, V., Driessnack, M., & Mendes, I. (2007). An overview of research designs relevant to nursing: Part 1: quantitative research designs. *Revista Latino-Americana de Enfermagem*, 15(3), 502-507. <https://doi.org/10.1590/S0104-11692007000300022>
- Stadler, F. (2001). *The Vienna Circle: Studies in the Origins, Development, and Influence of Logical Empiricism*. Springer.
- Stöberg, U. (2012). A research review of e-assessment. *Assessment & Evaluation in Higher Education*, 37(5), 591-604.
- Stoen, S. M., McDaniel, M. A., Frey, R. F., Hynes, K. M., & Cahill, M. J. (2020). Force Concept Inventory: More than just conceptual understanding. *Physical Review Physics Education Research*, 16(1), 010105.
- Sukkarieh, J., Pulman, S. & Raikes, N. (2003). Auto-marking: using computational linguistics to score short, free-text responses. 29th International Association for Educational Assessment (IAEA) Annual Conference, Manchester, United Kingdom.
- Sun, P., Tsai, R. J., Finger, G., Chen, Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education*, 50, 1183-1202.
- Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: a test of competing models. *Information Systems Research*, 6(2), 144-176.
- Teo, T. (2009) Modeling Technology Acceptance in Education: A Study of Pre-Service Teachers. *Computers & Education*, 52, 302-312. <http://dx.doi.org/10.1016/j.compedu.2008.08.006>



- Teo, T. (2011). Factors influencing teachers' intention to use technology: Model development and test. *Computers & Education*, 57(4), 2432-2440.
- Teo, T., Lee, C. B., & Chai, C. S. (2008). Understanding pre-service teachers' computer attitudes: applying and extending the technology acceptance model. *Journal of Computer Assisted Learning*, 24(2), 128-143.
- Terzis, V., & Economides, A. A. (2011). The acceptance and use of computer-based assessment. *Computers & Education*, 56(4), 1032-1044.
- Terzis, V., Moridis, C. N., & Economides, A. A. (2012). How student's personality traits affect Computer Based Assessment Acceptance: Integrating BFI with CBAAM. *Computers in Human Behavior*, 28(5), 1985-1996.
- Terzis, V., Moridis, C. N., & Economides, A. A. (2012). The effect of emotional feedback on behavioral intention to use computer based assessment. *Computers & Education*, 59(2), 710-721.
- Terzis, V., Moridis, C. N., & Economides, A. A. (2013). Continuance acceptance of computer based assessment through the integration of user's expectations and perceptions. *Computers & Education*, 62, 50-61.
- Terzis, V., Moridis, C. N., & Economides, A. A. (2013). Measuring instant emotions based on facial expressions during computer-based assessment. *Personal and ubiquitous computing*, 17, 43-52.
- Thompson, R., Higgins, C. & Howell, J. (1991). Personal Computing: Toward a Conceptual Model of Utilization, *MIS Quarterly*, vol. 15, núm. 1.
- Thongsri, N., Shen, L., & Bao, Y. (2019). Investigating factors affecting learner's perception toward online learning: evidence from ClassStart application in Thailand. *Behaviour & Information Technology*, 38(12), 1243-1258.
- Timmis, S, Broadfoot, P, Sutherland, R, & Oldfield, A (2016). Rethinking assessment in a digital age. *British Educational Research Journal*, 42(3), 454-476.
- Triandis, H. C. (1977). *Interpersonal behaviour*. Monterey, CA: Brooke/Cole.
- Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. In *Advances in experimental social psychology* (Vol. 29, pp. 271-360). Academic Press.

- Van Raaij, E. M., & Schepers, J. J. L. (2008). The acceptance and use of a virtual learning environment in China. *Computers and Education*, 50(3), 838-852.
- Venkatesh, V. & Davis, F. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, vol. 46, núm. 2, pp. 186.
- Venkatesh, V. (1999). Creation of favorable user perceptions: Exploring the role of intrinsic motivation. *MIS quarterly*, 239-260.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
- Vinzi, V. E., Trinchera, L., & Amato, S. (2010). PLS path modeling: from foundations to recent developments and open issues for model assessment and improvement *Handbook of partial least squares* (pp. 47-82): Springer.
- Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.
- Wallace, C. S., & Bailey, J. M. (2010). Do concept inventories actually measure anything. *Astronomy Education Review*, 9(1).
- Wang, Y. (2003). Assessment of learner satisfaction with asynchronous electronic learning systems. *Information & Management*, 41(1), 75-86.
- Wang, Y.-S., Wu, M.-C., & Wang, H.-Y. (2009). Investigating the determinants and age and gender differences in the acceptance of mobile learning. *British Journal of Educational Technology*, 40(1), 92-118.
- Wedlock, B. C., & Trahan, M. P. (2019). Revisiting the Unified Theory of Acceptance and the Use of Technology (UTAUT) Model and Scale: An Empirical Evolution of Educational Technology. *Research Issues in Contemporary Education*, 4(1), 6-20.
- White, B. Y. (1983). Sources of difficulty in understanding Newtonian dynamics. *Cognitive science*, 7(1), 41-65.
- Wieman, C. E. (2014). Large-scale comparison of science teaching methods sends clear message. *Proceedings of the National Academy of Sciences*, 111(23), 8319-8320.
- Williams, M. D., Dwivedi, Y. K., Lal, B., & Schwarz, A. (2009). Contemporary trends and issues in IT adoption and diffusion research. *Journal of Information Technology*, 24(1), 1-10.

- Williams, M. D., Rana, N. P., & Dwivedi, Y. K. (2015). The unified theory of acceptance and use of technology (UTAUT): a literature review. *Journal of Enterprise Information Management*, 28(3), 443-488.
- Williamson, D.M., Xi, X., Breyer, F.J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, 31(1), 2-13.
- Winkley, J. (2010). E-assessment and innovation. A Becta report. <http://www.becta.org.uk/>
- Wold, H. (1980). Model construction and evaluation when theoretical knowledge is scarce. *Evaluation of econometric models* (pp. 47-74): Academic Press.
- Wold, H. (1985). Partial least squares. *Encyclopedia of statistical sciences*.
- Wong, K. K.-K. (2013). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. *Marketing Bulletin*, 24(1), 1-32.
- Woodford, K., & Bancroft, P. (2005, January). Multiple choice questions not considered harmful. In *Proceedings of the 7th Australasian conference on Computing Education-Volume 42* (pp. 109-116).
- Yi, M.Y. and Hwang, Y. (2003) Predicting the Use of Web-Based Information Systems: Self-Efficacy, Enjoyment, Learning Goal Orientation, and the Technology Acceptance Model. *International Journal of Human Computer Studies*, 59, 431-449. [http://dx.doi.org/10.1016/S1071-5819\(03\)00114-9](http://dx.doi.org/10.1016/S1071-5819(03)00114-9)
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1-27.
- Zehner, F., Salzer, C. & Goldhammer, F. (2016). Automatic Coding of Short Text Responses via clustering in Educational Assessment. *Educational and Psychological Measurement*, vol. 76, no. 2, pp. 280-303.

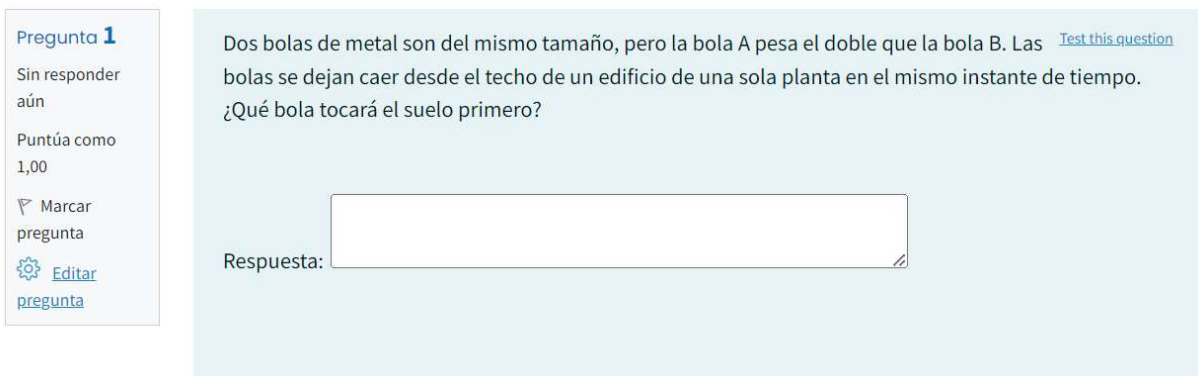
## Apéndice A: Traducción del Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM)

Constructo	Definición operacional	Ítems
Utilidad Percibida	Se define como el grado en que una persona cree que usar un sistema en particular mejorará su rendimiento	1. El uso de la Evaluación basada en computadora (EBC) mejorará mi trabajo 2. El uso de la Evaluación basada en computadora (EBC) mejorará mi eficacia 3. El uso de la Evaluación basada en computadora aumentará mi productividad
Facilidad de uso percibida	Se define como el grado en que una persona cree que usar el sistema estaría libre de esfuerzo	4. Mi interacción con el sistema es clara y comprensible 5. Es fácil para mí llegar a ser hábil en el uso del sistema 6. El sistema fácil de usar
Autoeficacia Informática	Se define como las percepciones del individuo sobre su capacidad para utilizar computadoras	7. Podría completar un trabajo o tarea usando la computadora 8. Podría completar un trabajo o una tarea usando la computadora si alguien me mostrara cómo hacerlo primero 9. Puedo navegar fácilmente por la Web para encontrar cualquier información que necesite 10. Era completamente capaz de usar la computadora e Internet antes de comenzar a usar la evaluación basada en computadora (EBC)
Influencia Social	El efecto de la opinión de otras personas, la influencia superior y la influencia de los compañeros	11. Las personas que influyen en mi comportamiento piensan que debería usar la EBC 12. Las personas que son importantes para mí piensan que debería usar la EBC 13. Los estudiantes de último año de mi universidad han sido de ayuda en el uso de la EBC 14. En general, mi universidad ha apoyado el uso de la EBC
Condiciones Facilitadoras	Son factores que influyen en la creencia de un	15. Cuando necesito ayuda para usar la EBC, alguien está ahí para ayudarme

	individuo para realizar un procedimiento	16. Cuando necesito ayuda para aprender a usar la EBC, el soporte de ayuda está ahí para enseñarme
Contenido	Es uno de los determinantes de la satisfacción del alumno al usar sistemas de aprendizaje	17. Las preguntas de la EBC fueron claras y comprensibles 18. Las preguntas de la EBC fueron fáciles de responder 19. Las preguntas de la EBC eran relativas al temario del curso 20. Las preguntas de la EBC fueron útiles para mi curso
Expectativa de Meta	Es una variable que influye en la creencia de un individuo que está preparado adecuadamente para usar la EBC	21. La preparación de los cursos fue suficiente para la EBC 22. Mi preparación personal fue suficiente para la EBC 23. Mis expectativas de desempeño fueron acordes a los resultados de la EBC
Percepción lúdica	Se define por tres dimensiones: concentración, curiosidad y disfrute del usuario al interactuar con el sistema	24. Usar EBC me mantiene feliz para realizar mi tarea 25. Usar EBC me hace disfrutar de mi aprendizaje 26. Usando EBC, mi curiosidad se estimula 27. El uso de EBC conducirá a mi exploración
Intención de utilizar la EBC	Se refiere a la medida en que una persona tiene la intención de utilizar una tecnología en particular	28. Tengo la intención de utilizar EBC en el futuro 29. Predigo que usaría EBC en el futuro 30. Planeo usar EBC en el futuro

## Apéndice B. Instrumento de evaluación en formato de respuesta corta

Basado en el Inventario de conceptos de fuerza (Force Concept Inventory, FCI) (Hestenes et al., 1992)



The screenshot shows a question interface. On the left, a sidebar contains the following information: 'Pregunta 1', 'Sin responder aún', 'Puntúa como 1,00', 'Marcar pregunta', and 'Editar pregunta'. The main area contains the question text: 'Dos bolas de metal son del mismo tamaño, pero la bola A pesa el doble que la bola B. Las bolas se dejan caer desde el techo de un edificio de una sola planta en el mismo instante de tiempo. ¿Qué bola tocará el suelo primero?' followed by a 'Respuesta:' label and a text input field.

Figura B.1. Pregunta 1 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
    match_mw (ninguna)
    match_mw (ambas)
    match_mw (identica*)
    match_mw (juntas)
    match_mw (igual)
    match_mw (mismo)
    match_mw (simultáneamente)
    match_mw (simultanea)
)

#Respuesta incorrecta
match_any (
    match_mw (mayor peso)
    match_mw (una)
    match_mw (no lo sé)
    match_mw (A)
)
```

Figura B.2. Reglas de calificación para pregunta 1 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 2**

Sin responder aún

Puntúa como 1,00

🚩 Marcar pregunta

⚙️ [Editar pregunta](#)

Dos bolas de metal son del mismo tamaño, pero la bola A pesa el doble que la bola B. [Test this question](#)

Ambas ruedan sobre una mesa horizontal con la misma velocidad. Indique qué bola tocará inicialmente el suelo más cerca de la mesa.

Respuesta:

Figura B.3. Pregunta 2 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mw (ninguna)
  match_mw (ambas)
  match_mw (identica*)
  match_mw (juntas)
  match_mw (igual)
  match_mw (mismo)
  match_mw (simultáneamente)
  match_mw (simultanea)
)

#Respuesta incorrecta
match_any (
  match_mw (doble)
  match_mw (mayor peso)
  match_w (una)
  match_w (cualquiera)
  match_mw (no lo sé)
  match_mw (A)
  match_mw (más lejos)
  match_mw (más cerca)
  match_mw (mitad)
)

```

Figura B.4. Reglas de calificación para pregunta 2 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 3**

Sin responder aún

Puntúa como 1,00

🚩 Marcar pregunta

⚙️ Editar pregunta

[Test this question](#)

Se deja caer una piedra desde el techo de un edificio de una sola planta hasta la superficie de la tierra. Indique qué le sucederá a la velocidad de la Piedra mientras está en vuelo, antes de que golpee el suelo.

Nota. Recuerde ignorar los efectos de la resistencia del aire cuando responda esta pregunta.

Respuesta:

Figura B.5. Pregunta 3 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mw (más rápido)
  match_mw (mayor*)
  match_mw (crece)
  match_mw (crecera)
  match_mw (aumenta*)
  match_mw (incrementa*)
  match_mw (acelera*)
)

#Respuesta incorrecta
match_any (
  match_mw (terminal)
  match_mw (max*)
  match_mw (no lo sé)
)
```

Figura B.6. Reglas de calificación para pregunta 3 del instrumento de evaluación en formato de respuesta corta.



**Pregunta 4**

Sin responder aún

Puntúa como 1,00

🚩 Marcar pregunta

⚙️ [Editar pregunta](#)

Un camión grande choca de frente con un automóvil pequeño. Compare la fuerza ejercida [Test this question](#) por el automóvil sobre el camión con la fuerza ejercida por el camión sobre el automóvil durante la colisión. ¿Qué fuerza es mayor?

Respuesta:

Figura B.7. Pregunta 4 del instrumento de evaluación en formato de respuesta corta.

```

#Respuesta correcta
match_any (
  match_mw (equivalente*)
  match_mw (mismo valor)
  match_mw (mayor*)
  match_mw (ninguna más grande*|mayor*|más alta)
  match_mw (ninguna)
  match_mw (igual*)
)

#Respuesta incorrecta
match_any (
  match_w (cualquiera)
  match_mw (más grande*|mayor*|más alta)
  match_mw (no lo sé)
  match_mw (no iguales)
)

```

Figura B.8. Reglas de calificación para pregunta 4 del instrumento de evaluación en formato de respuesta corta.

Pregunta 5

Sin responder aún

Puntúa como 1,00

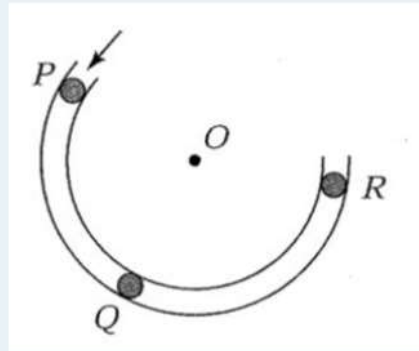
🚩 Marcar pregunta

⚙ Editar pregunta

La figura adjunta muestra un canal sin fricción en la forma de un segmento de un círculo con centro en  $O$ . El canal ha sido anclado a una mesa horizontal sin fricción. Estás mirando hacia la mesa. Se lanza una pelota a gran velocidad al canal en  $P$  y sale en  $R$ .

¿Cuáles de las siguientes fuerzas actúan sobre la bola cuando está en el canal sin fricción en el punto  $Q$ ?

Seleccione una opción o más:



- a. Una fuerza de gravedad hacia abajo.
- b. Una fuerza que apunta de  $Q$  a  $O$ .
- c. Una fuerza en la dirección del movimiento.
- d. Una fuerza que apunta de  $O$  a  $Q$ .
- e. Una fuerza hacia arriba de la mesa.

Figura B.9. Pregunta 5 del instrumento de evaluación en formato de respuesta corta.

Pregunta 6

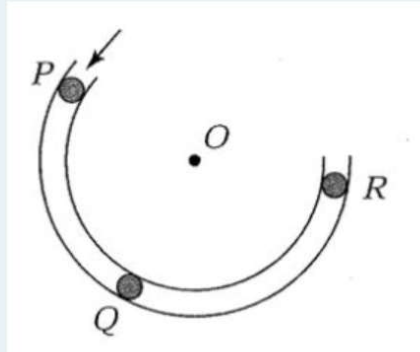
Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

Mencione la fuerza o fuerzas que actúan sobre la pelota después de que emerge de la pista [Test this question](#) en R.



Respuesta:

Figura B.10. Pregunta 6 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mow (gravedad*|peso*|gravitacional*|mesa gravitación)
  match_mow (gravedad*|peso*|gravitacional*|soporte gravitación)
  match_mow (gravedad*|peso*|gravitacional*|contacto gravitación)
  match_mow (gravedad*|peso*|gravitacional*|hacia arriba
gravitación)
  match_mow (gravedad*|peso*|gravitacional*|reacción gravitación)
  match_mow (gravedad*|peso*|gravitacional*|normal gravitación)
  match_mow (gravedad*|peso*|gravitacional*|superficie gravitación)
)

#Respuesta incorrecta
match_any (
  match_mw (centrifug*)
  match_mw (empuje)
  match_mw (no lo sé)
  match_mow (movimiento)
  match_mw (momentum)
)
```

Figura B.11. Reglas de calificación para pregunta 6 del instrumento de evaluación en formato de respuesta corta.

Pregunta 7

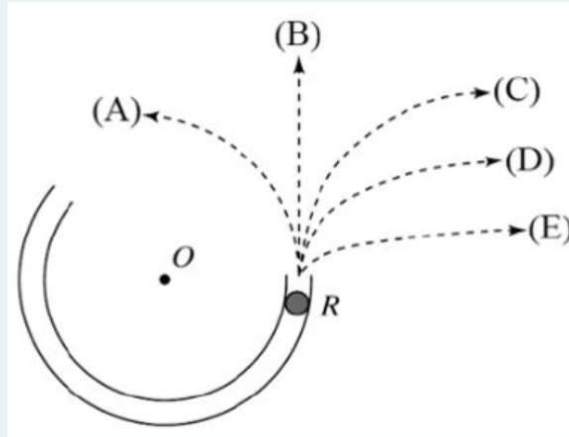
Sin responder aún

Puntúa como 1,00

⚑ Marcar pregunta

⚙ Editar pregunta

¿Qué camino en la siguiente figura seguiría la pelota después de que sale del canal en R y se mueve a través de la superficie de la mesa sin fricción? [Test this question](#)



Respuesta:

Figura B.12. Pregunta 7 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
  match_mw (B)

#Respuesta incorrecta
match_any (
  match_mw (A)
  match_mw (C)
  match_mw (D)
  match_mw (no sé)
)
```

Figura B.13. Reglas de calificación para pregunta 7 del instrumento de evaluación en formato de respuesta corta.

Pregunta 8

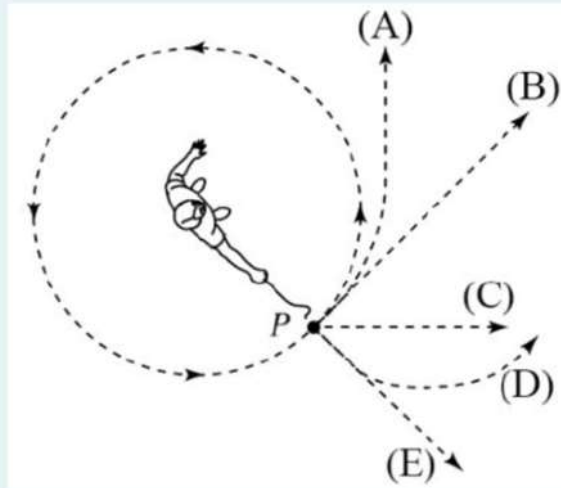
Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

Una bola de acero está unida a una cuerda y se balancea en una trayectoria circular en un [Test this question](#) plano horizontal como se ilustra en la figura anterior. En el punto A indicado en la figura, la cuerda se rompe repentinamente cerca de la pelota. Si estos eventos se observan directamente desde arriba como en la figura, ¿Qué camino seguiría más de cerca la pelota después de que se rompiera la cuerda?



Respuesta:

Figura B.14. Pregunta 8 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
    match_mw (B)
#Respuesta incorrecta
match_any (
    match_mw (A)
    match_mw (C)
    match_mw (D)
    match_mw (no sé)
)
```

Figura B.15. Reglas de calificación para pregunta 8 del instrumento de evaluación en formato de respuesta corta.

Pregunta 9

Sin responder aún

Puntúa como 1,00

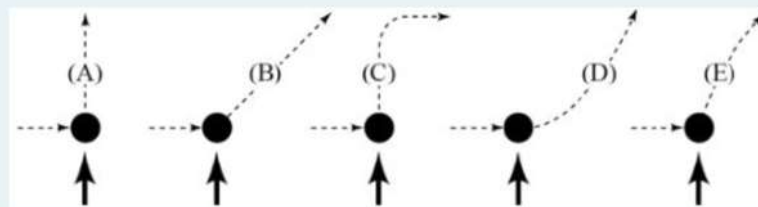
Marcar pregunta

Editar pregunta

La figura representa un disco de hockey sobre hielo deslizando con velocidad constante [Test this question](#)  $r$  en línea recta desde el punto P hasta el punto Q sobre una superficie sin fricción. Tu estas observando directamente desde arriba como en la figura. Cuando el disco alcanza el punto Q, recibe una rápida patada horizontal en la dirección de la flecha. Si el disco hubiera estado en reposo en el punto Q, la patada lo habría puesto en movimiento horizontal con una velocidad  $k$  en la dirección de la patada.



¿Cuál de los siguientes caminos seguiría el disco más de cerca después de recibir la patada?



Respuesta:

Figura B.16. Pregunta 9 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
  match_mw (B)

#Respuesta incorrecta
match_any (
  match_mw (A)
  match_mw (C)
  match_mw (D)
  match_mw (no sé)
)
```

Figura B.17. Reglas de calificación para pregunta 9 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 10**

Sin responder aún

Puntúa como 1,00

🚩 Marcar pregunta

⚙ Editar pregunta

Compare cualitativamente la velocidad del disco justo después de recibir la patada con las [Test this question](#) velocidades  $r$  y  $k$ . Por ejemplo, ¿la velocidad es mayor que  $r$  pero menor que  $k$ , mayor que ambas o menor que ambas?

Respuesta:

Figura B.18. Pregunta 10 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mw (mayor*|grande*|mayor que r y k)
  match_mw (mayor*|grande*|mayor que k y r)
  match_mw (raíz cuadrada)
  match_mw (mayor*|grande*|mayor ambas*)
)

#Respuesta incorrecta
match_any (
  match_mw (menor)
  match_mw (igual*|mism*|sin cambio)
  match_mw (no sé)
  match_mw (disminu*)
)
```

Figura B.19. Reglas de calificación para pregunta 10 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 11**

Sin responder aún

Puntúa como 1,00

🚩 Marcar pregunta

⚙️ [Editar pregunta](#)

¿ Qué pasará con la velocidad del disco cuando se mueva a lo largo de la trayectoria sin [Test this question](#) fricción después de recibir la patada ?

Respuesta:

Figura B.20. Pregunta 11 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mw (no cambia)
  match_mw (sin cambio)
  match_mw (nada)
  match_mw (constante*)
  match_mw (igual*|permanece*|mantiene*)
)

#Respuesta incorrecta
match_any (
  match_mw (disminu*)
  match_mw (acelera*)
  match_mw (no se)
  match_mw (incremen*)
)
```

Figura B.21. Reglas de calificación para pregunta 11 del instrumento de evaluación en formato de respuesta corta.



**Pregunta 12**

Sin responder aún

Puntúa como 1,00

🚩 Marcar pregunta

⚙ Editar pregunta

[Test this question](#)

Mencione la fuerza o fuerzas principales que actúan sobre el disco después de haber recibido la patada y todavía se está moviendo a lo largo de la trayectoria sin fricción.

Respuesta:

Figura B.22. Pregunta 12 del instrumento de evaluación en formato de respuesta corta.

```

#Respuesta correcta
match_any (
  match_mow (gravedad*|peso*|gravitacional*|soporte gravitación)
  match_mow (gravedad*|peso*|gravitacional*|contacto gravitación)
  match_mow (gravedad*|peso*|gravitacional*|hacia arriba
gravitación)
  match_mow (gravedad*|peso*|gravitacional*|reacción gravitación)
  match_mow (gravedad*|peso*|gravitacional*|normal gravitación)
  match_mow (gravedad*|peso*|gravitacional*|superficie gravitación)
  match_mow (gravedad*|peso*|gravitacional*|hielo gravitación)
)

#Respuesta incorrecta
match_any (
  match_mw (patada)
  match_mw (empuje)
  match_mw (no lo sé)
  match_mow (movimiento)
  match_mw (momentum)
)

```

Figura B.23. Reglas de calificación para pregunta 12 del instrumento de evaluación en formato de respuesta corta.

Pregunta 13

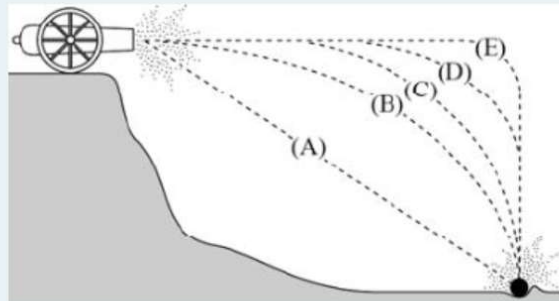
Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

Una bala es disparada por un cañón desde lo alto de un acantilado como se muestra en la [figura de arriba](#). ¿Cuál de los caminos seguiría más de cerca la bala de cañón?



Respuesta:

Figura B.24. Pregunta 13 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
  match_mw (B)

#Respuesta incorrecta
match_any (
  match_mw (A)
  match_mw (C)
  match_mw (D)
  match_mw (no sé)
)
```

Figura B.25. Reglas de calificación para pregunta 13 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 14**

Sin responder aún

Puntúa como 1,00

▼ Marcar pregunta

[Editar pregunta](#)

Un niño lanza una bola de acero hacia arriba. ¿Qué fuerza o fuerzas actúan sobre la bola después de que sale de la mano del niño? [Test this question](#)

Respuesta:

Figura B.26. Pregunta 14 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mw (peso)
  match_mw (gravedad)
  match_mw (gravitacional)
)

#Respuesta incorrecta
match_any (
  match_mwop[] (gravedad y peso)
  match_mwop[] (peso y gravedad)
  match_mw (fuerza aplicada)
  match_mw (fuerza resultante)
  match_mw (tensión)
  match_mw (masa)
  match_mw (reacción)
  match_mw (energía)
  match_mw (fuerza inicial)
  match_mw (lanza*)
  match_mw (fuerza hacia arriba)
  match_mw (arrastre)
  match_mw (no sé)
  match_mw (fricción)
  match_mw (potencial)
  match_mw (empuje)
  match_mw (cinética)
)
```

Figura B.27. Reglas de calificación para pregunta 14 del instrumento de evaluación en formato de respuesta corta.

Pregunta 15

Sin responder aún

Puntúa como 1,00

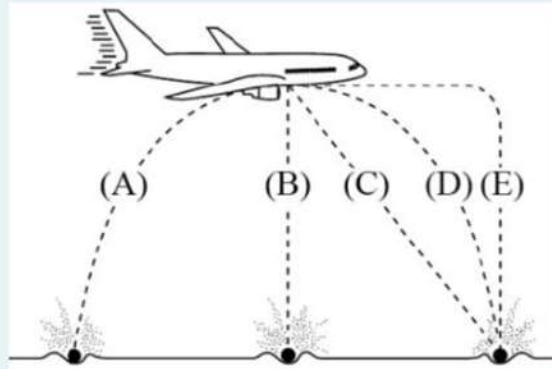
Marcar pregunta

Editar pregunta

14. Una bola de boliche se cae accidentalmente de la bodega de carga de un avión comercial mientras vuela en dirección horizontal.

[Test this question](#)

Según lo observado por una persona parada en el suelo y mirando el avión como en la figura siguiente, ¿Qué camino seguiría más de cerca la bola de boliche después de dejar el avión?



Respuesta:

Figura B.28. Pregunta 15 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_mw (D)

#Respuesta incorrecta
match_any (
  match_mw (A)
  match_mw (C)
  match_mw (B)
  match_mw (E)
  match_mw (no sé)
)
```

Figura B.29. Reglas de calificación para pregunta 15 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 16**

Sin responder aún

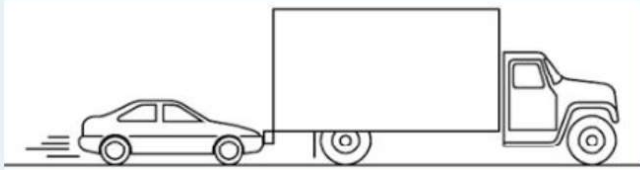
Puntúa como 1,00

⚑ Marcar pregunta

⚙ Editar pregunta

[Test this question](#)

Un camión grande se descompone y un automóvil pequeño lo empuja de vuelta a la ciudad, como se muestra en la siguiente figura.



Mientras el automóvil, que empuja al camión, acelera, ¿Cómo se compara la fuerza que ejerce el automóvil sobre el camión con la fuerza que ejerce el camión sobre el automóvil?

Respuesta:

Figura B.30. Pregunta 16 del instrumento de evaluación en formato de respuesta corta.

```

#Respuesta correcta
match_any (
  match_mw (igual*|mism*|identicas)
)

#Respuesta incorrecta
match_any (
  match_mw (más grande*|mayor*|increment*)
  match_mw (no lo sé)
)

```

Figura B.31. Reglas de calificación para pregunta 16 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 17**  
Sin responder aún  
Puntúa como 1,00  
Marcar pregunta  
Editar pregunta

Después de que el automóvil alcanza la velocidad constante a la que su conductor desea empujar el camión, ¿Cómo se compara la fuerza que ejerce el automóvil sobre el camión con la fuerza que ejerce el camión sobre el automóvil? [Test this question](#)

Respuesta:

Figura B.32. Pregunta 17 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mw (igual*|mism*|identicas)
)

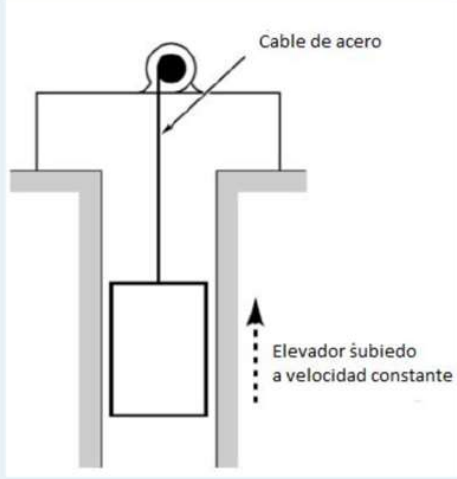
#Respuesta incorrecta
match_any (
  match_mw (más grande*|mayor*|increment*)
  match_mw (no lo sé)
)
```

Figura B.33. Reglas de calificación para pregunta 17 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 18**  
Sin responder aún  
Puntúa como 1,00  
Marcar pregunta  
Editar pregunta

En el siguiente diagrama, un ascensor está siendo levantado por un eje a una velocidad constante por un cable de acero. ¿Qué te dice esto acerca de las fuerzas que actúan sobre el ascensor?

[Test this question](#)



Respuesta:

Figura B.34. Pregunta 18 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mw (ceros|ninguna|nada|nula*)
  match_mw (identicas)
  match_mw (cancelan)

  match_mw (no hay fuerza neta)
  match_mw (no hay fuerza resultante)
  match_mw (balance*)
  match_mw (igual*|equivalente|equilib*)
)

#Respuesta incorrecta
match_any (
  match_mw (menor)
  match_mw (menor)
  match_mw (desbalanceadas)
  match_mw (no lo sé)
  match_mw (mayor)
)
```

Figura B.35. Reglas de calificación para pregunta 18 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 19**

Sin responder aún


Puntúa como 1,00

🚩 Marcar pregunta

⚙ Editar pregunta

[Test this question](#)

La figura muestra un péndulo que se balancea en una cuerda, comenzando en un punto más alto que P.



Mencione la fuerza o fuerzas que actúan sobre el péndulo cuando está en la posición P.

Respuesta:

Figura B.36. Pregunta 19 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta

match_any (
  match_mow (peso fuerza cuerda)
  match_mow (grav* jala* cuerda)
  match_mow (peso jala* cuerda)
  match_mow (grav* tension)
  match_mow (peso tensión)
  match_mow (grav* fuerza cuerda)
)

#Respuesta incorrecta

match_any (
  match_mw (aplicada)
  match_mw (acelera*)
  match_mw (centri*)
  match_mw (no lo sé)
  match_mw (empuje)
)
```

Figura B.37. Reglas de calificación para pregunta 19 del instrumento de evaluación en formato de respuesta corta.



**Pregunta 20**

Sin responder aún

Puntuación como 1,00

Marcar pregunta

Editar pregunta

Un cohete se desplaza lateralmente en el espacio exterior desde el punto P hasta el punto Q, como se muestra a continuación. El cohete no está sujeto a fuerzas externas. Comenzando en la posición Q, el motor del cohete se enciende e inmediatamente produce un empuje constante (fuerza sobre el cohete) en ángulo recto con la línea PQ. El empuje constante se mantiene hasta que el cohete alcanza un punto R en el espacio.

¿Cuál de los siguientes caminos representa mejor el camino del cohete entre los puntos Q y R?

Respuesta:

Figura B.38. Pregunta 20 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_mw (E)

#Respuesta incorrecta
match_any (
  match_mw (A)
  match_mw (C)
  match_mw (D)
  match_mw (B)
  match_mw (no sé)
)
```

Figura B.39. Reglas de calificación para pregunta 20 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 21**  
Sin responder aún  
Puntúa como 1,00  
Marcar pregunta  
Editar pregunta

A medida que el cohete se mueve de la posición Q a la posición R, ¿aumenta, disminuye o permanece igual su velocidad? [Test this question](#)

Respuesta:

Figura B.40. Pregunta 21 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta  
  
match_any (  
  match_mw (aumenta)  
  match_mw (acelera)  
  match_mw (incremen*)  
  
)  
  
#Respuesta incorrecta  
  
match_any (  
  match_mw (disminuye)  
  match_mw (igual)  
  match_mw (constante)  
  match_mw (no lo sé)  
  match_mw (menor)  
  
)
```

Figura B.41. Reglas de calificación para pregunta 21 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 22**  
Sin responder aún  
Puntúa como 1,00  
Marcar pregunta  
Editar pregunta

En el punto R, el motor del cohete se apaga y el empuje cae inmediatamente a cero. ¿Cuál de los siguientes caminos seguirá el cohete más allá del punto R? [Test this question](#)

Respuesta:

Figura B.42. Pregunta 22 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta

    match_mw (B)

#Respuesta incorrecta
match_any (
    match_mw (A)
    match_mw (C)
    match_mw (D)
    match_mw (E)
    match_mw (no sé)
)
```

Figura B.43. Reglas de calificación para pregunta 22 del instrumento de evaluación en formato de respuesta corta.

Pregunta 23

Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

Test this question

¿Qué sucede con la velocidad del cohete más allá de la posición R?

Respuesta:

Figura B.44. Pregunta 23 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mw (igual*|mantien*|permane*)
  match_mw (constante)
  match_mw (sin cambio)
  match_mw (no cambia)
)

#Respuesta incorrecta
match_any (
  match_mw (disminuye)
  match_mw (decrementa)
  match_mw (desacelera)
  match_mw (no lo sé)
  match_mw (incremen*)
  match_mw (aumenta)
)
```

Figura B.45. Reglas de calificación para pregunta 23 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 24**

Sin responder aún

Puntúa como 1,00

🚩 Marcar pregunta

⚙️ [Editar pregunta](#)

Una mujer ejerce una fuerza horizontal constante sobre una caja grande. Como resultado, [Test this question](#) la caja se mueve sobre el piso horizontal a velocidad constante. ¿Qué te dice esto sobre las fuerzas que actúan sobre la caja?

Respuesta:

Figura B.46. Pregunta 24 del instrumento de evaluación en formato de respuesta corta.

```

#Respuesta correcta
match_any (
  match_mw (igual*|equilibr*|equivalente*)
  match_mw (ceros*|ningun*|nada*|nula*|anulan)
  match_mw (mism*|identico*)
  match_mw (cancelan)
  match_mw (sin fuerza neta)
  match_mw (sin fuerza resultante)
  match_mw (balan*|balanceadas)
)

#Respuesta incorrecta
match_any (
  match_mwp (no están equilibradas)
  match_mwp (no son iguales)
  match_mw (desbalanceadas)
  match_mw (gravedad)
  match_mw (reacción)
  match_mw (no sé)
  match_mw (más grande)
  match_mw (más pequeña)
  match_mw (menor)
  match_mw (mayor)
)

```

Figura B.47. Reglas de calificación para pregunta 24 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 25**

Sin responder aún

Puntúa como 1,00

🚩 Marcar pregunta

⚙ Editar pregunta

[Test this question](#)

Si la fuerza horizontal externa constante que ejerce la mujer sobre la caja, ahora se duplica mientras empuja la caja sobre el mismo piso horizontal, ¿Qué sucede con la velocidad de la caja?

Respuesta:

Figura B.48. Pregunta 25 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mw (mayor)
  match_mw (aumenta)
  match_mw (alta)
  match_mw (incrementa)
  match_mw (acelera*)
)

#Respuesta incorrecta
match_any (
  match_mw (raiz)
  match_mw (no lo sé)
  match_mw (doble|dos|2|veces)
)
```

Figura B.49. Reglas de calificación para pregunta 25 del instrumento de evaluación en formato de respuesta corta.

Pregunta 26

Sin responder aún

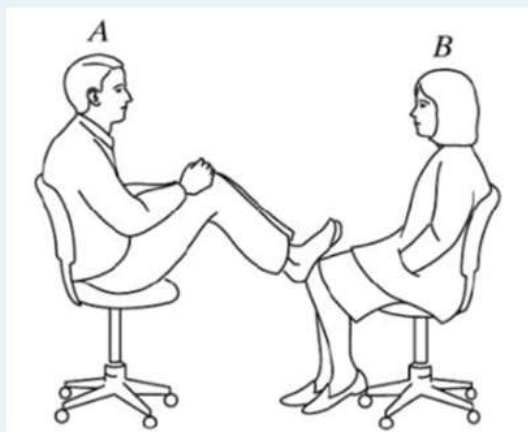
Puntúa como 1,00

Marcar pregunta

Editar pregunta

[Test this question](#)

En la siguiente figura, el estudiante A tiene una masa de 95 kg y el estudiante B tiene una masa de 77 kg. Se sientan en sillas de oficina idénticas una frente a la otra. El estudiante A coloca sus pies sobre las rodillas del estudiante B, como se muestra. Luego, el estudiante A empuja repentinamente hacia afuera con los pies, lo que hace que ambas sillas se muevan.



¿Qué puedes decir sobre la cantidad de fuerza que cada estudiante ejerce sobre el otro durante el empujón?

Respuesta:

Figura B.50. Pregunta 26 del instrumento de evaluación en formato de respuesta corta.

```
#Respuesta correcta
match_any (
  match_mw (igual*|mism*|idénticas|balanceadas|parejas|emparejadas|)
)

#Respuesta incorrecta
match_any (
  match_mw (más grande*|mayor*|increment*)
  match_mw (menor*|menos)
  match_mw (no lo sé)
)
```

Figura B.51. Reglas de calificación para pregunta 26 del instrumento de evaluación en formato de respuesta corta.

**Pregunta 27**

Sin responder aún

Puntúa como 1,00

🚩 Marcar pregunta

⚙️ [Editar pregunta](#)

[Test this question](#)

Una silla de oficina vacía está en reposo en el suelo. ¿Qué fuerza o fuerzas actúan sobre la silla de oficina?

Respuesta:

Figura B.52. Pregunta 27 del instrumento de evaluación en formato de respuesta corta.

```

#Respuesta correcta
match_any (
  match_mow (grav* empuj* piso|superficie)
  match_mow (peso empuj* piso|superficie)
  match_mw (grav* fuerza piso|superficie)
  match_mw (fuerza piso|superficie grav*)
  match_mow (peso fuerza piso|superficie)
  match_mow (peso resistencia piso|superficie)
  match_mow (grav* resistencia piso|superficie)
  match_mow (grav* reacción)
  match_mow (peso reacción)
  match_mow (grav* normal)
  match_mow (peso normal)
  match_mow (grav* contacto)
  match_mow (peso contacto)
)

#Respuesta incorrecta
match_any (
  match_mw (no lo sé)
)

```

Figura B.53. Reglas de calificación para pregunta 27 del instrumento de evaluación en formato de respuesta corta.



## Apéndice C: Instrumento de evaluación en formato de opción múltiple

Basado en el Inventario de conceptos de fuerza (Force Concept Inventory, FCI) (Hestenes et al., 1992).

**Pregunta 1**

Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

Dos bolas de metal son del mismo tamaño, pero la bola A pesa el doble que la bola B. Las bolas se dejan caer desde el techo de un edificio de una sola planta en el mismo instante de tiempo. ¿Qué bola tocará el suelo primero?

- A. Aproximadamente la mitad para la bola más liviana que para la bola más pesada.
- B. Considerablemente menor para la bola más pesada, pero no necesariamente la mitad.
- C. Considerablemente menor para la bola más liviana, pero no necesariamente la mitad.
- D. Aproximadamente la mitad para la bola más pesada que para la bola más liviana.
- E. Aproximadamente el mismo para ambas bolas.

Figura C.1. Pregunta 1 del instrumento de evaluación en formato de opción múltiple.

**Pregunta 2**

Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

Dos bolas de metal son del mismo tamaño, pero la bola A pesa el doble que la bola B. Ambas ruedan sobre una mesa horizontal con la misma velocidad. Indique qué bola tocará inicialmente el suelo más cerca de la mesa. En esta situación:

- A. La bola mas pesada golpea el suelo considerablemente mas cerca de la base de la mesa que la bola mas liviana, pero no necesariamente a la mitad de la distancia horizontal.
- B. Ambas bolas golpean el suelo aproximadamente a la misma distancia horizontal de la base de la mesa.
- C. La bola mas liviana golpea el suelo aproximadamente a la mitad de la distancia horizontal de la base de la mesa que la bola mas pesada.
- D. La bola mas liviana golpea el suelo considerablemente mas cerca de la base de la mesa que la bola mas pesada, pero no necesariamente a la mitad de la distancia horizontal.
- E. La bola mas pesada golpea el suelo aproximadamente a la mitad de la distancia horizontal de la base de la mesa que la bola más liviana.

Figura C.2. Pregunta 2 del instrumento de evaluación en formato de opción múltiple.

**Pregunta 3**

Sin responder aún

Puntúa como 1,00

🚩 Marcar pregunta

⚙️ [Editar pregunta](#)

Se deja caer una piedra desde el techo de un edificio de una sola planta hasta la superficie de la tierra. Indique qué le sucederá a la velocidad de la piedra mientras está en vuelo, antes de que golpee el suelo.

Nota. Recuerde ignorar los efectos de la resistencia del aire cuando responda esta pregunta.

- A. Caer debido a la tendencia natural de todos los objetos a descansar sobre la superficie de la tierra.
- B. Aumenta su velocidad porque una fuerza de gravedad casi constante actúa sobre ella.
- C. Caer debido a los efectos combinados de la fuerza de la gravedad, empujándola hacia abajo, y la fuerza del aire, también empujándola hacia abajo.
- D. Aumenta su velocidad mientras cae porque la atracción gravitatoria se hace considerablemente mayor cuanto más se acerca la piedra a la tierra.
- E. Alcanza un máximo de velocidad muy pronto después de ser soltada y desde entonces cae con una velocidad constante.

Figura C.3. Pregunta 3 del instrumento de evaluación en formato de opción múltiple.

**Pregunta 4**  
Sin responder aún  
Puntúa como 1,00  
🚩 Marcar pregunta  
⚙ Editar pregunta

Un camión grande choca de frente con un automóvil pequeño. Compare la fuerza ejercida por el automóvil sobre el camión con la fuerza ejercida por el camión sobre el automóvil durante la colisión. ¿Qué fuerza es mayor?

- A. Ninguno ejerce una fuerza sobre el otro, el auto es aplastado simplemente porque se interpone en el camino del camión.
- B. La intensidad de la fuerza que el automóvil ejerce sobre el camión es mayor que la de la fuerza que el camión ejerce sobre el auto.
- C. El camión ejerce una fuerza sobre el automóvil pero el auto no ejerce ninguna fuerza sobre el camión.
- D. La intensidad de la fuerza que el camión ejerce sobre el automóvil es mayor que la de la fuerza que el auto ejerce sobre el camión.
- E. El camión ejerce una fuerza de la misma intensidad sobre el auto que la que el auto ejerce sobre el camión.

Figura C.4. Pregunta 4 del instrumento de evaluación en formato de opción múltiple.

Pregunta 5

Sin responder aún

Puntúa como 1,00

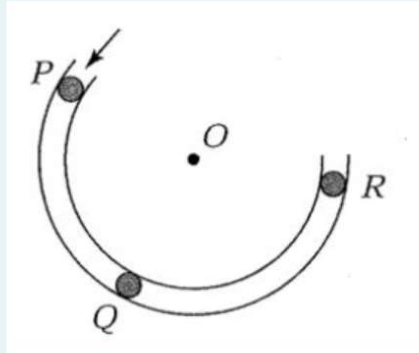
Marcar pregunta

Editar pregunta

La figura adjunta muestra un canal sin fricción en la forma de un segmento de un círculo con centro en  $O$ . El canal ha sido anclado a una mesa horizontal sin fricción. Estás mirando hacia la mesa. Se lanza una pelota a gran velocidad al canal en  $P$  y sale en  $R$ .

¿Cuáles de las siguientes fuerzas actúan sobre la bola cuando está en el canal sin fricción en el punto  $Q$ ?

Seleccione una opción o más:



- A. Una fuerza que apunta de  $O$  a  $Q$ .
- B. Una fuerza que apunta de  $Q$  a  $O$ .
- C. Una fuerza de gravedad hacia abajo.
- D. Una fuerza en la dirección del movimiento.
- E. Una fuerza hacia arriba de la mesa.

Figura C.5. Pregunta 5 del instrumento de evaluación en formato de opción múltiple.

Pregunta 6

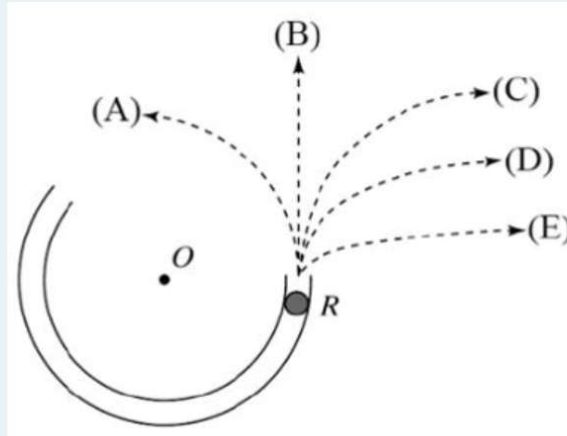
Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

¿Qué camino en la siguiente figura seguiría la pelota después de que sale del canal en R y se mueve a través de la superficie de la mesa sin fricción?



- seguirá el camino (D)
- seguirá el camino (E)
- seguirá el camino (C)
- seguirá el camino (A)
- seguirá el camino (B)

Figura C.6. Pregunta 6 del instrumento de evaluación en formato de opción múltiple.

**Pregunta 7**

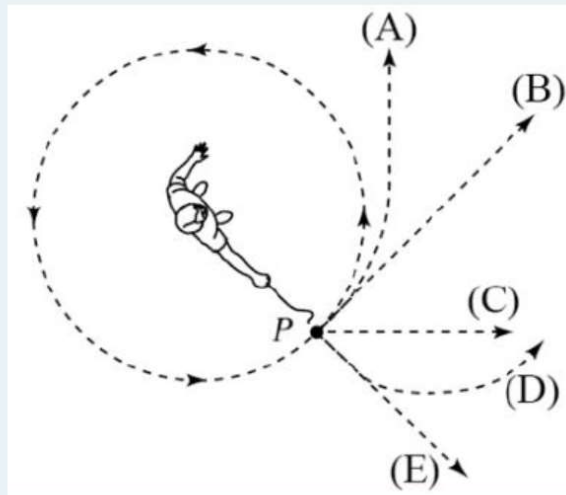
Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

Una bola de acero está unida a una cuerda y se balancea en una trayectoria circular en un plano horizontal como se ilustra en la figura anterior. En el punto A indicado en la figura, la cuerda se rompe repentinamente cerca de la pelota. Si estos eventos se observan directamente desde arriba como en la figura, ¿Qué camino seguiría más de cerca la pelota después de que se rompiera la cuerda?



- Seguirá el camino (A)
- Seguirá el camino (E)
- Seguirá el camino (B)
- Seguirá el camino (D)
- Seguirá el camino (C)

Figura C.7. Pregunta 7 del instrumento de evaluación en formato de opción múltiple.

Pregunta 8

Sin responder aún

Puntúa como 1,00

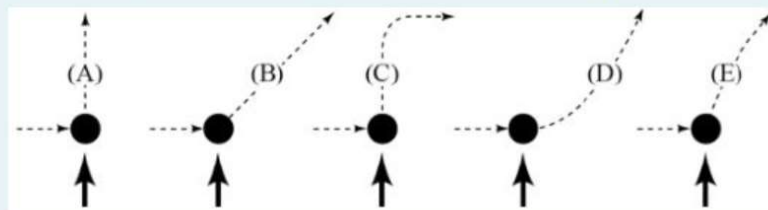
Marcar pregunta

Editar pregunta

La figura representa un disco de hockey sobre hielo deslizándose con velocidad constante  $r$  en línea recta desde el punto P hasta el punto Q sobre una superficie sin fricción. Tu estas observando directamente desde arriba como en la figura. Cuando el disco alcanza el punto Q, recibe una rápida patada horizontal en la dirección de la flecha. Si el disco hubiera estado en reposo en el punto Q, la patada lo habría puesto en movimiento horizontal con una velocidad  $k$  en la dirección de la patada.



¿Cuál de los siguientes caminos seguiría el disco más de cerca después de recibir la patada?



- Seguirá el camino (E)
- Seguirá el camino (B)
- Seguirá el camino (D)
- Seguirá el camino (C)
- Seguirá el camino (A)

Figura C.8. Pregunta 8 del instrumento de evaluación en formato de opción múltiple.

Pregunta 9

Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

La velocidad del disco justo después de recibir la patada es:

- A. Mayor que cualquiera de las velocidades " $v_r$ " o " $v_k$ ", pero menor que la suma aritmética de estas dos velocidades.
- B. igual a la suma aritmética de las velocidades " $v_r$ " y " $v_k$ ".
- C. Igual a la velocidad " $v_k$ " resultante del golpe e independiente de la velocidad " $v_r$ ".
- D. Igual a la velocidad " $v_r$ " que tenía antes de recibir el golpe.
- E. Menor que cualquiera de las velocidades " $v_r$ " o " $v_k$ ".

Figura C.9. Pregunta 9 del instrumento de evaluación en formato de opción múltiple.

Pregunta 10

Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

¿ Qué pasará con la velocidad del disco cuando se mueva a lo largo de la trayectoria sin fricción después de recibir la patada ?

- A. Aumenta durante un rato y después disminuye.
- B. Disminuye continuamente.
- C. Es constante durante un rato y después disminuye.
- D. Es constante.
- E. Aumenta continuamente.

Figura C.10. Pregunta 10 del instrumento de evaluación en formato de opción múltiple.



**Pregunta 11**

Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

Mencione la fuerza o fuerzas principales que actúan sobre el disco después de haber recibido la patada y todavía se está moviendo a lo largo de la trayectoria sin fricción.

- A. Una fuerza hacia abajo debida a la gravedad y una fuerza horizontal en la dirección del movimiento.
- B. Una fuerza hacia abajo debida a la gravedad, una fuerza hacia arriba ejercida por la superficie y una fuerza horizontal en la dirección del movimiento.
- C. Una fuerza hacia abajo debida a la gravedad
- D. Una fuerza hacia abajo debida a la gravedad y una fuerza hacia arriba ejercida por la superficie.
- E. Ninguna. (No actúa ninguna fuerza sobre el disco).

Figura C.11. Pregunta 11 del instrumento de evaluación en formato de opción múltiple.

**Pregunta 12**

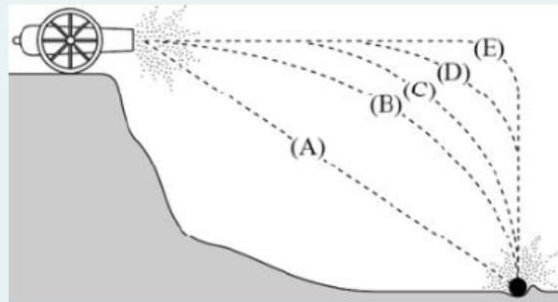
Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

Una bala es disparada por un cañón desde lo alto de un acantilado como se muestra en la figura de arriba. ¿Cuál de los caminos seguiría más de cerca la bala de cañón?



- Seguirá el camino (D)
- Seguirá el camino (A)
- Seguirá el camino (B)
- Seguirá el camino (E)
- Seguirá el camino (C)

Figura C.12. Pregunta 12 del instrumento de evaluación en formato de opción múltiple.

Pregunta 13

Sin responder  
aún

Puntúa como  
1,00

🚩 Marcar  
pregunta

⚙️ [Editar  
pregunta](#)

Un niño lanza una bola de acero hacia arriba. ¿Qué fuerza o fuerzas actúan sobre la bola después de que sale de la mano del niño? (suponga que las fuerzas ejercidas por el aire son despreciables)

- A. Una fuerza hacia abajo debida a la gravedad junto con una fuerza hacia arriba que disminuye continuamente.
- B. Una fuerza hacia arriba que disminuye continuamente desde el momento en que la bola abandona la mano del chico hasta que alcanza su punto mas alto; en el camino de descenso hay una fuerza hacia abajo debida a la gravedad que aumenta continuamente a medida que el objeto se acerca progresivamente a la tierra
- C. Ninguna de las anteriores. La bola cae al suelo por su tendencia natural a descansar sobre la superficie de la tierra.
- D. Solo una fuerza hacia abajo, prácticamente constante, debida a la gravedad.
- E. Una fuerza hacia abajo prácticamente constante debida a la gravedad junto con una fuerza hacia arriba que disminuye continuamente hasta que la bola alcanza su punto mas alto; en el camino de descenso solo hay una fuerza constante hacia abajo debida a la gravedad

Figura C.13. Pregunta 13 del instrumento de evaluación en formato de opción múltiple.

Pregunta 14

Sin responder aún

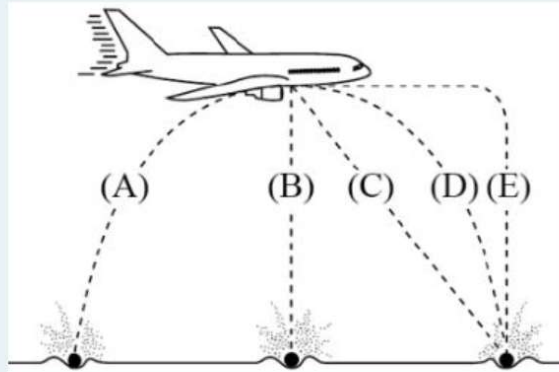
Puntúa como 1,00

⚑ Marcar pregunta

⚙ Editar pregunta

Una bola de boliche se cae accidentalmente de la bodega de carga de un avión comercial mientras vuela en dirección horizontal.

Según lo observado por una persona parada en el suelo y mirando el avión como en la figura siguiente, ¿Qué camino seguiría más de cerca la bola de boliche después de dejar el avión?



- Seguirá el camino (C)
- Seguirá el camino (D)
- Seguirá el camino (A)
- Seguirá el camino (E)
- Seguirá el camino (B)

Figura C.14. Pregunta 14 del instrumento de evaluación en formato de opción múltiple.

Pregunta 15

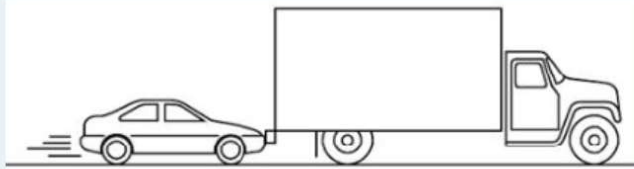
Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

Un camión grande se descompone y un automóvil pequeño lo empuja de vuelta a la ciudad, como se muestra en la siguiente figura.



Mientras el automóvil, que empuja al camión, acelera, ¿Cómo se compara la fuerza que ejerce el automóvil sobre el camión con la fuerza que ejerce el camión sobre el automóvil?

- A. La intensidad de la fuerza que el automóvil aplica sobre el camión es menor que la de la fuerza que el camión aplica sobre el auto.
- B. Ni el camión ni el automóvil ejercen fuerza alguna sobre el otro. El camión es empujado hacia adelante simplemente porque esta en el camino del automóvil.
- C. Dado que el motor del automóvil esta en marcha, este puede empujar al camión, pero el motor del camión no esta funcionando, de modo que el camión no puede empujar al auto. El camión es empujado hacia adelante simplemente porque esta en el camino del automóvil.
- D. La intensidad de la fuerza que el automóvil aplica sobre el camión es igual a la de la fuerza que el camión aplica sobre el auto.
- E. La intensidad de la fuerza que el automóvil aplica sobre el camión es mayor que la de la fuerza que el camión aplica sobre el auto.

Figura C.15. Pregunta 15 del instrumento de evaluación en formato de opción múltiple.

Pregunta 16

Sin responder  
aún

Puntúa como  
1,00

🚩 Marcar  
pregunta

⚙ Editar  
pregunta

Después de que el automóvil alcanza la velocidad constante a la que su conductor desea empujar el camión, ¿Cómo se compara la fuerza que ejerce el automóvil sobre el camión con la fuerza que ejerce el camión sobre el automóvil?

- A. Dado que el motor del automóvil esta en marcha, este puede empujar al camión, pero el motor del camión no esta funcionando, de modo que el camión no puede empujar al auto. El camión es empujado hacia adelante simplemente porque esta en el camino del automóvil.
- B. La intensidad de la fuerza que el automóvil aplica sobre el camión es igual a la de la fuerza que el camión aplica sobre el auto.
- C. Ni el camión ni el automóvil ejercen fuerza alguna sobre el otro. El camión es empujado hacia adelante simplemente porque esta en el camino del automóvil.
- D. La intensidad de la fuerza que el automóvil aplica sobre el camión es mayor que la de la fuerza que el camión aplica sobre el auto.
- E. La intensidad de la fuerza que el automóvil aplica sobre el camión es menor que la de la fuerza que el camión aplica sobre el auto.

Figura C.16. Pregunta 16 del instrumento de evaluación en formato de opción múltiple.

Pregunta 17

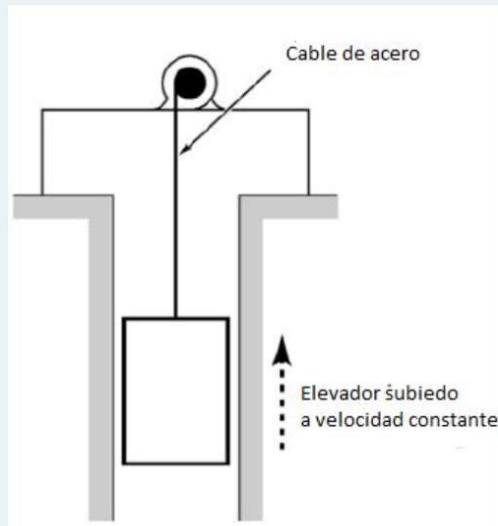
Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

En el siguiente diagrama, un ascensor está siendo levantado por un eje a una velocidad constante por un cable de acero. ¿Qué te dice esto acerca de las fuerzas que actúan sobre el ascensor?



- A. Ninguna de las anteriores. (El ascensor sube porque el cable se está acortando, no porque el cable ejerza una fuerza hacia arriba sobre el ascensor).
- B. La fuerza hacia arriba ejercida por el cable es mayor que la fuerza hacia abajo debida a la gravedad.
- C. La fuerza hacia arriba ejercida por el cable es mayor que la suma de la fuerza hacia abajo debida a la gravedad y una fuerza hacia abajo debida al aire.
- D. La fuerza hacia arriba ejercida por el cable es menor que la fuerza hacia abajo debida a la gravedad.
- E. La fuerza hacia arriba ejercida por el cable es igual a la fuerza hacia abajo debida a la gravedad.

Figura C.17. Pregunta 17 del instrumento de evaluación en formato de opción múltiple.

Pregunta 18

Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

La figura muestra un péndulo que se balancea en una cuerda, comenzando en un punto más alto que P.



Considérense las siguientes fuerzas:

- i. Una fuerza hacia abajo debida a la gravedad.
- ii. Una fuerza ejercida por la cuerda dirigida de P hacia O.
- iii. Una fuerza en la dirección del movimiento del péndulo.
- iv. Una fuerza en la dirección de O hacia P.

Mencione la fuerza o fuerzas que actúan sobre el péndulo cuando está en la posición P.

- i y ii.
- i,iii y iv.
- i, ii y iii.
- solo la i.
- i y iii

Figura C.18. Pregunta 18 del instrumento de evaluación en formato de opción múltiple.

Pregunta 19

Sin responder aún

Puntúa como 1,00

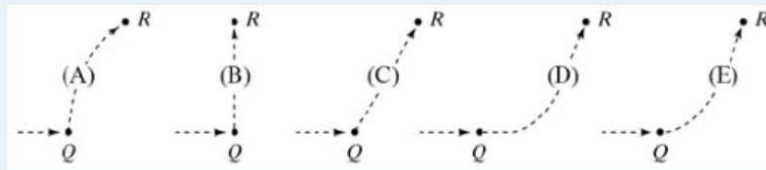
Marcar pregunta

Editar pregunta

Un cohete se desplaza lateralmente en el espacio exterior desde el punto P hasta el punto Q, como se muestra a continuación. El cohete no está sujeto a fuerzas externas. Comenzando en la posición Q, el motor del cohete se enciende e inmediatamente produce un empuje constante (fuerza sobre el cohete) en ángulo recto con la línea PQ. El empuje constante se mantiene hasta que el cohete alcanza un punto R en el espacio.



¿Cuál de los siguientes caminos representa mejor el camino del cohete entre los puntos Q y R?



- El camino (C)
- El camino (A)
- El camino (D)
- El camino (E)
- El camino (B)

Figura C.19. Pregunta 19 del instrumento de evaluación en formato de opción múltiple.



Pregunta 20

Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

A medida que el cohete se mueve de la posición Q a la posición R, la magnitud de su velocidad es:

- A. Continuamente decreciente.
- B. Constante.
- C. Continuamente creciente.
- D. Constante durante un rato y después decreciente.
- E. Creciente durante un rato y después constante.

Figura C.20. Pregunta 20 del instrumento de evaluación en formato de opción múltiple.

Pregunta 21

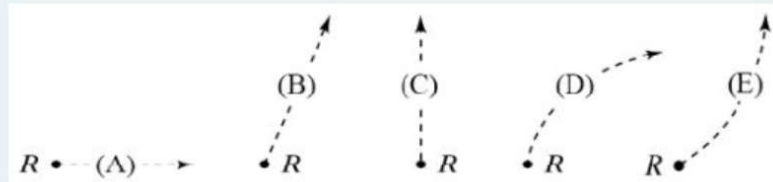
Sin responder aún

Puntúa como 1,00

Marcar pregunta

Editar pregunta

En el punto R, el motor del cohete se apaga y el empuje cae inmediatamente a cero. ¿Cuál de los siguientes caminos seguirá el cohete más allá del punto R?



- El camino (C)
- El camino (B)
- El camino (D)
- El camino (E)
- El camino (A)

Figura C.21. Pregunta 21 del instrumento de evaluación en formato de opción múltiple.

**Pregunta 22**  
Sin responder aún  
Puntúa como 1,00  
Marcar pregunta  
Editar pregunta

A partir de la posición R la velocidad del cohete es:

- A. Continuamente creciente.
- B. Constante.
- C. Continuamente decreciente.
- D. Creciente durante un rato y después constante.
- E. Constante durante un rato y después decreciente.

Figura C.22. Pregunta 22 del instrumento de evaluación en formato de opción múltiple.

**Pregunta 23**  
Sin responder aún  
Puntúa como 1,00  
Marcar pregunta  
Editar pregunta

Una mujer ejerce una fuerza horizontal constante sobre una caja grande. Como resultado, la caja se mueve sobre el piso horizontal a velocidad constante. La fuerza horizontal constante aplicada por la mujer:

- A. Es mayor que la fuerza total que se opone al movimiento de la caja.
- B. Es mayor que el peso de la caja.
- C. Tiene la misma magnitud que el peso de la caja.
- D. Es mayor que el peso de la caja y también que la fuerza total que se opone a su movimiento
- E. Tiene la misma magnitud que la fuerza total que se opone al movimiento de la caja.

Figura C.23. Pregunta 23 del instrumento de evaluación en formato de opción múltiple.

Pregunta 24

Sin responder  
aún

Puntúa como  
1,00

🚩 Marcar  
pregunta

⚙️ [Editar  
pregunta](#)

Si la fuerza horizontal externa constante que ejerce la mujer sobre la caja, ahora se duplica mientras empuja la caja sobre el mismo piso horizontal, La caja se moverá:

- A. Con una velocidad que es constante y mayor que la velocidad de la pregunta anterior durante un rato, y después con una velocidad que aumenta progresivamente.
- B. Con una velocidad constante que es mayor que la velocidad de la pregunta anterior, pero no necesariamente el doble.
- C. Con una velocidad constante que es el doble de la velocidad de la pregunta anterior.
- D. Con una velocidad continuamente creciente.
- E. Con una velocidad creciente durante un rato, y después con una velocidad constante.

Figura C.24. Pregunta 24 del instrumento de evaluación en formato de opción múltiple.

Pregunta 25

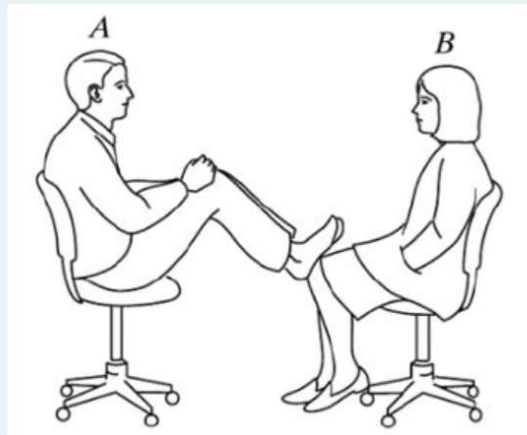
Sin responder aún

Puntúa como 1,00

▼ Marcar pregunta

⚙ Editar pregunta

En la siguiente figura, el estudiante A tiene una masa de 95 kg y el estudiante B tiene una masa de 77 kg. Se sientan en sillas de oficina idénticas una frente a la otra. El estudiante A coloca sus pies sobre las rodillas del estudiante B, como se muestra. Luego, el estudiante A empuja repentinamente hacia afuera con los pies, lo que hace que ambas sillas se muevan.



Durante el empujón, mientras los estudiantes están aún en contacto:

- A. Ambos estudiantes ejercen la misma cantidad de fuerza sobre el otro.
- B. Ambos estudiantes ejercen una fuerza sobre el otro, pero "B" ejerce una fuerza mayor.
- C. El estudiante "A" ejerce una fuerza sobre el estudiante "B", pero "B" no ejerce ninguna fuerza sobre "A".
- D. Ambos estudiantes ejercen una fuerza sobre el otro, pero "A" ejerce una fuerza mayor.
- E. Ninguno de los estudiantes ejerce una fuerza sobre el otro.

Figura C.25. Pregunta 25 del instrumento de evaluación en formato de opción múltiple.

Pregunta 26

Sin responder  
aún


Puntúa como  
1,00

🚩 Marcar  
pregunta

⚙ Editar  
pregunta

Una silla de oficina vacía está en reposo en el suelo. Considérense las siguientes fuerzas:

- i. Una fuerza hacia abajo debida a la gravedad.
- ii. Una fuerza hacia arriba ejercida por el suelo.
- iii. Una fuerza neta hacia abajo ejercida por el aire.

Cual(es) de estas fuerzas actúa  sobre la silla de oficina?

- A. ii y iii.
- B. Solo la i.
- C. i, ii y iii.
- D. i y ii.
- E. Ninguna de las fuerzas. (Puesto que la silla esta en reposo no hay ninguna fuerza actuando sobre ella).

Figura C.26. Pregunta 26 del instrumento de evaluación en formato de opción múltiple.

## Apéndice D: Resultados de la prueba piloto sobre la confiabilidad del CBAAM

Para evaluar la confiabilidad de la versión en español de las subescalas del instrumento (CBAAM), se obtuvo un coeficiente de consistencia interna con la prueba de alpha de Cronbach con los 30 reactivos de 0.9190, que resultó alto. El coeficiente puede oscilar entre 0 y 1, en donde 0 significa nula confiabilidad y 1, confiabilidad total.

Se utilizó el programa minitab 21.3.1 para efectuar el cálculo; los coeficientes de consistencia interna se encontraron en un rango 0.6838 a 0.9325. Únicamente, la subescala de expectativa de meta está por debajo de 0.70, mínimo que se considera adecuado. En la siguiente tabla se muestran los puntajes en cada subescala del instrumento.

### Confiabilidad de la versión en español de las subescalas del Modelo de aceptación de evaluación basada en computadora (CBAAM)

	Instrumento CBAAM en español
Subescala	Coficiente alpha de Cronbach
1. Utilidad Percibida	0.7184
2. Facilidad de uso percibida	0.8501
3. Autoeficacia Informática	0.8076
4. Influencia Social	0.8065
5. Condiciones Facilitadoras	0.7619
6. Contenido	0.7386
7. Expectativa de Meta	0.6838
8. Percepción Lúdica	0.8763
9. Intención de utilizar la EBC	0.9325

## Apéndice E: Variables del modelo CBAAM

Variable	Definición conceptual	Definición operacional	Tipo de variable	Instrumento de recolección de datos
Percepción lúdica (PL)	<p>el constructo percepción lúdica, que se define por tres dimensiones:</p> <ul style="list-style-type: none"> <li>•Concentración: determina si el usuario está concentrado en la actividad.</li> <li>•Curiosidad: establece si el sistema despertó la curiosidad cognitiva del usuario.</li> <li>•Disfrute: estipula si el usuario está disfrutando o no de la interacción con el sistema.</li> </ul>	PL se define como el grado de concentración, curiosidad y disfrute de una persona al interactuar con el sistema de evaluación basada en computadora	Cuantitativa	Ítems 24-27 en escala Likert, del Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM)
Utilidad percibida (UP)	Se define como el grado en que una persona cree que usar un sistema en particular mejorará su rendimiento.	El grado en que un individuo cree que usar el sistema de evaluación basada en computadora mejoraría su rendimiento.	Cuantitativa	Ítems 1-3 en escala Likert, del Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM)
Facilidad de uso percibida (FDUP)	Se define como el grado en que una persona cree que usar el sistema estaría libre de esfuerzo.	El grado en que un individuo cree que usar el sistema de evaluación basada en computadora sería fácil.	Cuantitativa	Ítems 4-6 en escala Likert, del Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM)
Autoeficacia informática (AI)	Se define como las percepciones del individuo	El grado en que un individuo cree sobre su	Cuantitativa	Ítems 7-10 en escala Likert, del Modelo de

	sobre su capacidad para utilizar computadoras.	capacidad para utilizar que usar el sistema de evaluación basada en computadora sería fácil.		Aceptación de Evaluación Basada en Computadora (CBAAM)
Influencia Social (IS)	El efecto de la opinión de otras personas, la influencia superior y la influencia de los compañeros.	El grado en que la opinión de los compañeros sobre el sistema de evaluación basada en computadora influye en un individuo.	Cuantitativa	Ítems 11-14 en escala Likert, del Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM)
Condiciones Facilitadoras (CF)	Son factores que influyen en la creencia de un individuo para realizar un procedimiento.	El grado en que los factores que inciden en la convicción de un individuo para llevar a cabo una evaluación basada en computadora.	Cuantitativa	Ítems 15-16 en escala Likert, del Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM)
Expectativa de Meta (EM)	Es una variable que influye en la creencia de un individuo que está preparado adecuadamente para usar la evaluación basada en computadora.	El grado en que un individuo cree que está preparado adecuadamente para usar la evaluación basada en computadora.	Cuantitativa	Ítems 21-23 en escala Likert, del Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM)
Contenido (C)	Es uno de los determinantes de la satisfacción del alumno al usar sistemas de aprendizaje	El grado en que un individuo cree que está satisfecho al usar la evaluación basada en computadora	Cuantitativa	Ítems 17-20 en escala Likert, del Modelo de Aceptación de Evaluación Basada en Computadora (CBAAM)
Intención de utilizar la EBC	Se refiere a la medida en que una persona tiene la intención de utilizar una tecnología en particular	El grado en que un individuo tiene la intención de usar la evaluación basada en computadora	Cuantitativa	Ítems 28-30 en escala Likert, del Modelo de Aceptación de Evaluación Basada en



				Computadora (CBAAM)
--	--	--	--	------------------------

Fuente: *elaboración propia con base en el modelo CBAAM de Terzis y Economides (2011).*