

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
Instituto de Investigación y Desarrollo Educativo



MAESTRÍA

“Diseño, desarrollo y validación de contenido de escalas para medir motivación, actitudes y autoeficacia en el aprendizaje de las ciencias en estudiantes universitarios”

Tesis
Que para obtener el grado de

Maestra en Ciencias Educativas

Presenta

Marla Lucía Márquez Beltrán

Ensenada, Baja California, México

Diciembre, 2022



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo
Maestría en Ciencias Educativas



**“Diseño, desarrollo y validación de contenido de escalas para
medir motivación, actitudes y autoeficacia en el aprendizaje de
las ciencias en estudiantes universitarios”**

TESIS

Que para obtener el grado de

MAESTRA EN CIENCIAS EDUCATIVAS

Presenta

María Lucía Márquez Beltrán

APROBADO POR:

Dr. Juan Carlos Rodríguez Macías
Director de tesis

Dr. Rodolfo García Galván
Sinodal

Dr. Luis Horacio Pedroza Zúñiga
Sinodal

Dra. Citlali Sánchez Álvarez
Sinodal





UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
Instituto de Investigación y Desarrollo Educativo

Ensenada, B.C., a 25 de noviembre de 2022

ASUNTO: Voto aprobatorio al trabajo de tesis para el grado de Maestra en Ciencias Educativas

Dr. Sergio Gerardo Málaga Villegas
Coordinador de Investigación y Posgrado
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por la C. María Lucía Márquez Beltrán, me permito comunicarle que he dado mi **VOTO APROBATORIO** al mencionado trabajo.

Con base en lo anterior, dicho documento se considera listo para su defensa en el examen de grado de Maestría en Ciencias Educativas, con el trabajo titulado:

"Diseño, desarrollo y validación de contenido de escalas para medir motivación, actitudes y autoeficacia en el aprendizaje de las ciencias en estudiantes universitarios".

Esperando reciba el presente de conformidad, quedo de usted.

Atentamente

Dr. Juan Carlos Rodríguez Macías



UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
Instituto de Investigación y Desarrollo Educativo

Ensenada, B.C., a 25 de noviembre de 2022

ASUNTO: Voto aprobatorio al trabajo
de tesis para el grado de Maestra en Ciencias Educativas

Dr. Sergio Gerardo Málaga Villegas
Coordinador de Investigación y Posgrado
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por la C. Marla Lucía Márquez Beltrán, me permito comunicarle que he dado mi VOTO APROBATORIO al mencionado trabajo.

Con base en lo anterior, dicho documento se considera listo para su defensa en el examen de grado de Maestría en Ciencias Educativas, con el trabajo titulado:

"Diseño, desarrollo y validación de contenido de escalas para medir motivación, actitudes y autoeficacia en el aprendizaje de las ciencias en estudiantes universitarios".

Esperando reciba el presente de conformidad, quedo de usted.

Atentamente

Dra. Citlalli Sánchez Álvarez



UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
Instituto de Investigación y Desarrollo Educativo

Ensenada, B.C., a 24 de noviembre de 2022

ASUNTO: Voto aprobatorio al trabajo de tesis para el grado de Maestra en Ciencias Educativas

Dr. Sergio Gerardo Málaga Villegas
Coordinador de Investigación y Posgrado
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por la **C. Marla Lucía Márquez Beltrán**, me permito comunicarle que he dado mi **VOTO APROBATORIO** al mencionado trabajo.

Con base en lo anterior, dicho documento se considera listo para su defensa en el examen de grado de Maestría en Ciencias Educativas, con el trabajo titulado:

"Diseño, desarrollo y validación de contenido de escalas para medir motivación, actitudes y autoeficacia en el aprendizaje de las ciencias en estudiantes universitarios".

Esperando reciba el presente de conformidad, quedo de usted.

Atentamente

Dr. Rodolfo García Galván



UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
Instituto de Investigación y Desarrollo Educativo

Ensenada, B.C., a 25 de noviembre de 2022

ASUNTO: Voto aprobatorio al trabajo de tesis para el grado de Maestra en Ciencias Educativas

Dr. Sergio Gerardo Málaga Villegas
Coordinador de Investigación y Posgrado
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por la **C. María Lucía Márquez Beltrán**, me permito comunicarle que he dado mi **VOTO APROBATORIO** al mencionado trabajo.

Con base en lo anterior, dicho documento se considera listo para su defensa en el examen de grado de Maestría en Ciencias Educativas, con el trabajo titulado:

"Diseño, desarrollo y validación de contenido de escalas para medir motivación, actitudes y autoeficacia en el aprendizaje de las ciencias en estudiantes universitarios".

Esperando reciba el presente de conformidad, quedo de usted.

Atentamente

Dr. Luis Horacio Pedroza Zúñiga

Tabla de contenido

Índice de tablas	vii
Índice de figuras	ix
AGRADECIMIENTOS	x
RESUMEN	xi
I. INTRODUCCIÓN	1
1.1. Antecedentes de medición en el aprendizaje de las ciencias.....	3
1.2. Planteamiento del problema.....	14
1.3. Objetivos de investigación	19
1.4. Justificación	20
II. MARCO REFERENCIAL	24
2.1. Fundamentos teóricos de la medición de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias.....	24
2.1.1. Motivación en el aprendizaje de las ciencias	29
2.1.2. Actitudes hacia la ciencia	31
2.1.3. Autoeficacia Percibida en el Aprendizaje de la Ciencia e investigación	36
2.2. Diseño, desarrollo y validación de instrumentos de medición	42
2.2.1. Programas para el diseño, desarrollo y validación de instrumentos de medición.....	42
2.2.2. Evidencias de validez de constructo en el aspecto del contenido de los instrumentos de medición	47
2.2.3. Evidencias de validez e imparcialidad de las pruebas.....	52
III. MÉTODO.....	56
3.1. Tipo de estudio.....	56
3.2. Participantes del estudio.....	57
3.3. Instrumentos y materiales	57
3.4. Procedimiento	59
Etapa I. Diseño de escalas para medir de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias en estudiantes de educación superior.....	61
Etapa II. Desarrollo de escalas para la medición de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias	63

Etapa III. Recopilación de evidencias de validez del constructo del contenido y de la imparcialidad de las escalas	64
IV. RESULTADOS.....	67
4.1. Ficha Técnica, Tablas de contenido y Escalas	67
4.2. Validación del aspecto del contenido por escala.....	78
4.3. Análisis de la imparcialidad por escala	89
V. DISCUSIÓN Y CONCLUSIONES	106
5.1. Discusión de los resultados	106
5.2. Limitaciones del estudio.....	113
5.3. Recomendaciones	115
REFERENCIAS	118
APÉNDICES.....	131
Apéndice I.	131
Apéndice II.....	141

Índice de tablas

Tabla 1. Etapas para el desarrollo de instrumentos de medición	43
Tabla 2. Tabla resumen del Manual de procedimientos de la recopilación de evidencias de validez de las escalas	59
Tabla 3. Modelo de diseño, desarrollo y validación de las escalas para explorar la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias en estudiantes de Educación Superior	60
Tabla 4. Ficha Técnica de las escalas para medir la motivación, actitudes y autoeficacia percibida en el aprendizaje de las ciencias.	68
Tabla 5. Escalas para medir la Motivación, Actitudes y Autoeficacia en el Aprendizaje de las Ciencias en estudiantes de Educación Superior	69
Tabla 6. Especificaciones de la Escala de Motivación en el Aprendizaje de las Ciencias (EMAC)	70
Tabla 7. Especificaciones de Escala de Actitudes hacia la Ciencia (EAC)	71
Tabla 8. Especificaciones de la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC).....	73
Tabla 9. Escala de Motivación en el Aprendizaje de las Ciencias (EMAC).....	73
Tabla 10. Escala de Actitudes hacia la Ciencia (EAC)	75
Tabla 11. Escala de Autoeficacia Percibida en el Aprendizaje de las Ciencias (EAPAMC).....	77
Tabla 12. Porcentaje de acuerdo entre jueces especialistas (PAJ) de la EMAC	78
Tabla 13. Resultados del cálculo del IVC de la EMAC.....	80
Tabla 14. Porcentaje de Acuerdo entre Jueces (PAJ) de la EAC por ítem.....	81
Tabla 15. Resultados del cálculo del IVC de la EAC por ítem	85
Tabla 16. Porcentaje de acuerdo entre jueces especialistas (PAJ) de la EAPAMC	87
Tabla 17. Resultados del cálculo del IVC de la EAPAMC por ítem	88
Tabla 18. Numero de comentarios por ítem y por criterio de la EDU en la Escala de Motivación en el Aprendizaje de las Ciencias (EMAC)	89
Tabla 19. Número de comentarios por ítem y por criterio de la EDU en la Escala de Actitudes hacia la Ciencia (EAC)	92

Tabla 20. Número de comentarios por ítem y por criterio de la EDU en la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC)	96
Tabla 21. Ítems que requieren modificaciones mayores de la EMAC	99
Tabla 22. Ítems que requieren modificaciones mayores de la EAC	99
Tabla 23. Escala de Motivación en el Aprendizaje de las Ciencias (EMAC-II).....	103
Tabla 24. Escala de Actitudes hacia la Ciencia (EAC-II)	104

Índice de figuras

<i>Figura 1.</i> Modelo de medición de Actitud hacia la ciencia en la prueba PISA.....	35
<i>Figura 2.</i> Variables relacionadas a la autoeficacia en investigación de acuerdo con la Teoría Social Cognitiva (TSC).....	40

AGRADECIMIENTOS

Agradezco...

A la vida. Por ponerme en el lugar, en el momento y con las personas indicadas.

A mis docentes, a todos.

A los que con paciencia y esmero me han hecho amar la investigación.

Al Consejo Nacional de Ciencia y Tecnología por impulsar y fomentar las vocaciones científicas.

Al Dr. Juan Carlos Pérez Morán por ayudarme a confiar en mi misma, por tantas enseñanzas, por mostrarme el maravilloso mundo de la psicometría.

A mi director de tesis el Dr. Juan Carlos Rodríguez Macías, a mi comité de tesis la Dra. Alicia Chaparro (QEPD), la Dra. Citlalli Sánchez, el Dr. Horacio Pedroza y el Dr. Rodolfo García, mi más profunda admiración.

A Ángel, por todo tu apoyo. Por tu amor a la ciencia que fue mi inspiración.

A mi familia y amigos, por siempre estar ahí.

Gracias.

Marla Lucía Márquez Beltrán

RESUMEN

El interés por el estudio y medición de las variables relevantes en el fomento, desarrollo y la formación científica de los estudiantes creció en las últimas décadas. Las naciones alrededor del mundo y los organismos internacionales como la UNESCO y la OCDE han expresado gran interés en el diseño de políticas nacionales e internacionales que impulsen un mayor nivel de alfabetización científica en la ciudadanía. El propósito de este estudio consistió en diseñar, desarrollar y validar tres escalas para medir la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias en estudiantes de Educación Superior (ES) nivel licenciatura. El procedimiento de la investigación constó de dos etapas: la primera consistió en el diseño y desarrollo de los instrumentos de medición. En la segunda, se recopilaron evidencias de validez de constructo en el aspecto del contenido y de la imparcialidad de las escalas. Se contó con la participación de un panel de expertos para la evaluación de los ítems de la EMAC, la EAC y la EAPAMC. Las tres escalas obtuvieron puntajes globales favorables en el cálculo de diversos índices que proporcionan información cuantitativa del acuerdo entre jueces. En el PAJ ($\geq .80$), la V de Aiken ($\geq .69$), el RVC clásico de Lawshe ($\geq .56$) y el RVC' de Tristán-López ($\geq .58$). En torno a la imparcialidad, las principales observaciones emitidas por los jueces en los ítems fueron con respecto a la *comprensibilidad* e *inclusión poblacional* de los reactivos. Se sugiere seguir con el proceso de recopilación de evidencias de validez de acuerdo con las directrices de los *Estándares* de la AERA, APA y NCME (2018).

Palabras clave: evaluación educativa, validez de constructo, evidencias del aspecto del contenido, formación científica, actitud hacia la ciencia, autoeficacia en investigación, motivación, estudiantes universitarios.

I. INTRODUCCIÓN

Organismos internacionales como la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), el Foro Abierto de Latinoamérica y el Caribe y la Organización para la Cooperación y Desarrollo Económicos (OCDE) advierten la necesidad de contar con ciudadanos científicamente preparados para enfrentar los desafíos por venir en todo el mundo. Tanto aquellos de orden social, como los relacionados con el cambio climático y la salud pública. Además, destacan la importancia de contar con estudiantes capaces de aprender a atender las necesidades sociales emergentes y a resolver problemas concretos utilizando competencias y conocimientos tanto científicos como tecnológicos (UNESCO, 2021).

En esta vía, se reconoce también la relevancia de la educación científica para elevar y mantener el interés del alumnado en las ciencias, dentro y fuera de las aulas. Por ello, la UNESCO pretende convocar a una renovación urgente de la educación científica, ya que es innegable su potencial para inculcar en los estudiantes el gusto por las actividades científicas (UNESCO, 2021). Así pues, el estudio y medición de las variables relevantes en la formación científica de los estudiantes de todos los niveles ha sido y es de gran interés para investigadores educativos alrededor del mundo (Munby, 1997; Siegel y Raney, 2003; Glynn et al. 2011; Velayutham et. al. 2011; Blanco-Blanco et al., 2016 y Summers y Abd-Khalick, 2018).

Algunas de las variables que son objeto de estudio de los especialistas en el campo de la enseñanza y el aprendizaje de las ciencias (retomadas del *Handbook of Research on Science Education*) son: el dominio afectivo del aprendizaje de las ciencias, la docencia en ciencias, el currículum y la evaluación en pequeña y gran escala, entre otras.

Dicho lo anterior, es relevante contar con instrumentos de medición con evidencias de validez que permitan el estudio de las variables relacionadas con el aprendizaje de las ciencias en estudiantes de ES nivel licenciatura. Se pretende que estos instrumentos contribuyan en la recopilación de información sobre el dominio afectivo del aprendizaje de las ciencias en ES, ya que en este nivel educativo las habilidades investigativas y la consideración de la ciencia como una opción laboral están en consolidación (Criollo, Romero y Fontaines-Ruiz, 2017). Así también, es necesario contar con instrumentos que se desarrollen desde un enfoque integrador de las ciencias adecuado para su aplicación en estudiantes de ciencias naturales y sociales o también llamadas *físicas y psíquicas* (Pierce, 1998 como se citó en Midtgarden, 2020) en idioma español y para población mexicana.

Los instrumentos desarrollados y validados en el aspecto del contenido en la presente tesis tienen el propósito de favorecer la toma de decisiones informada de estudiantes de licenciatura que deseen mejorar sus aprendizajes y trabajar en la exploración de su carrera, así como de su especialidad en el futuro. Asimismo, se pretende aportar un instrumento e información de su calidad técnica a investigadores educativos y formadores de profesionistas para su estudio y aplicación en evaluaciones diagnósticas, requeridas como insumo para el diseño de programas de fomento y desarrollo de vocaciones científicas en estudiantes de licenciatura.

1.1. Antecedentes de medición en el aprendizaje de las ciencias

El estudio y medición de las variables relacionadas al aprendizaje de las ciencias de los estudiantes es un tema relevante para los investigadores educativos a nivel internacional desde hace varias décadas (Munby, 1997; Glynn et al. 2011; Velayutham et. al. 2011; Stone, 2014; Blanco-Blanco et al., 2016; Abd-El-Khalick et al. 2015 y Summers y Abd-Khalick, 2018). Organismos internacionales interesados en el desarrollo de la alfabetización científica de los ciudadanos, actores educativos interesados en la enseñanza y aprendizaje de las ciencias, y estudiosos de las decisiones vocacionales de los estudiantes a lo largo de su trayecto educativo, reconocen la importancia de estudiar este fenómeno debido a la pérdida de interés en la ciencia que los estudiantes experimentan conforme avanza su escolaridad (Vázquez-Alonso y Mannasero-Mas, 2009b). Además de la elección de carreras científicas (Osborne et al. 2003) y la ciencia como opción laboral (Vázquez-Alonso y Mannasero-Mas, 2009c).

Para ahondar en ello, es necesario indicar que la formación científica del estudiantado es un tema de estudio que se aborda desde diversas líneas, enfoques y teorías. Algunas de ellas son: el vínculo enseñanza-aprendizaje de la ciencia (Tuan et al., 2005; Vázquez-Alonso y Manassero-Más, 2009a; Hodson, 2014; Cuevas Romo et al. 2016), la formación científica de los estudiantes (Ramírez et al., 2011), la vocación científica (Vázquez Alonso y Manassero Mas, 2009a; Blanco-Blanco, 2016), la competencia o entrenamiento científico (Mallinckrodt y Gelso, 2002), la alfabetización científica (Benjamin et al., 2015; Lee y Kim, 2018), la formación de investigadores (Fontaines-Ruiz y Urdaneta, 2009) el abordaje desde la ciencia ciudadana (Bonney et al., 2009, 2015), la mente del científico (Fugelsang et al., 2004; Feist, 2006a, 2006b), la cultura científica (Jegade, 1997; Godin y Gingras, 2000; Feuer et al., 2002) y la elección de una carrera científica (Lent et al., 2005, 2018).

El tema que nos compete se puede abordar desde sus componentes cognitivos, o *science skills* como lo son las habilidades de pensamiento o razonamiento científico, la observación, la generación de

hipótesis, la experimentación, la redacción científica (Crotwell et al., 2011; Etkina et al. 2006; Pérez-Morán, 2010; Davidescu et al., 2011; Gobert et al. 2012; Stone, 2014; Romine et al., 2016; Kruit et al. 2018) y el pensamiento crítico (Ossa-Cornejo et al., 2017).

Por otro lado, también es estudiado desde el dominio afectivo del aprendizaje: la motivación para el aprendizaje de la ciencia (Glynn et al., 2011; Velayutham et al., 2014; Ardura y Pérez-Bitrián, 2018), las actitudes hacia la ciencia (Willson, 1983; Munby, 1997; Osborne et al., 2003; Blalock et al., 2008; Vásquez Alonso y Manassero Más, 2009a, 2009b; Summers y Abd-El-Khalick, 2018; Toma y Lederman, 2020), los intereses científicos (Feist, 2006; Buccheri et al., 2011) la identidad científica de los estudiantes (Stets et al., 2016) e incluso su sentido de pertenencia a la comunidad científica y la relación con la autoeficacia de los estudiantes (Trujillo y Tanner, 2014), entre otros.

Estos componentes afectivos del aprendizaje fueron considerados afines para el presente estudio, ya que resulta relevante considerar que los estudiantes (o bien, los ciudadanos) tienen: (1) actitudes y experiencias positivas hacia la ciencia, y (2) habilidades científicas básicas, independientemente de si cursan o no una carrera de tipo científica. Por lo tanto, cuando se habla de formación científica o la formación de científicos se recoge una diversa variedad de constructos de tipo cognitivo y afectivo que impactan en la experiencia del aprendizaje de las ciencias.

Algunos de estos constructos se encuentran operacionalizados en instrumentos métricos. Por ejemplo, uno de los antecedentes más relevantes en la medición de los componentes relacionados a la formación científica es la *Competencia Científica*, una de las tres competencias medidas en la prueba Programme for International Student Assessment (PISA) aplicada a estudiantes de 15 años. El componente de *Competencia Científica* de la prueba PISA evalúa las siguientes sub-competencias: (a) identificar asuntos o temas científicos, (b) explicar fenómenos científicamente y (c) usar evidencia científica (OECD, 2022).

En la misma vía, Benjamin et al. (2015) desarrollaron la Encuesta de Alfabetización Científica para la preparación universitaria, e incluyó un instrumento construido para estudiantes de carreras dentro de las *Science, Technology, Engineering and Math* (STEM). En este estudio, la *alfabetización científica* es evaluada en tres dimensiones: los dominios de actitud y comportamiento de la alfabetización científica, el conocimiento del contenido y de conceptos científicos, y las habilidades de razonamiento científico.

En cuanto a la formación científica vista desde el ángulo vocacional o de elección profesional, se pueden mencionar dos principales antecedentes: el SDS Búsqueda Autodirigida de Holland et al. (2005) y la Batería Socio Cognitiva para la elección de vocaciones científico-matemáticas de Blanco-Blanco et al. (2016). El SDS explora la medida en que los rasgos de personalidad, los intereses y las aptitudes tienen congruencia con la elección de una carrera científica o de un trabajo en investigación científica. Por otro lado, en el trabajo de Blanco-Blanco et al. (2016) consistió en la identificación, traducción y adaptación de pruebas para medir el interés de los estudiantes de secundaria en ocupaciones científico-matemáticas. Esta batería mide la autoeficacia, los intereses, las expectativas de resultados, las consideraciones ocupacionales, así como los apoyos y las barreras sociales percibidas por los estudiantes de secundaria para elegir la carrera de Matemáticas.

Ahora bien, entre los antecedentes de la medición de los componentes cognitivos de la formación científica o del aprendizaje de las ciencias encontramos el trabajo de Crotwell et al. (2011), en el que se diseñó una rúbrica universal para la evaluación de las habilidades de razonamiento científico por medio de la redacción académica en estudiantes de licenciatura. Otro ejemplo es el trabajo de Etkina et. al (2006) que tuvo como resultado un conjunto de actividades y rúbricas para la evaluación de habilidades científicas en estudiantes de Física.

También se puede mencionar la aportación de Davidescu et al. (2011) en el que se produjo un instrumento de evaluación para la habilidad en el diseño y la experimentación científica en estudiantes de Física. Es destacable también el trabajo de Gobert et al. (2012), quienes desarrollaron el *Science*

Assessment, que evalúa por medio de un entorno interactivo y de simulación, la indagación y la habilidad para generar hipótesis pertinentes en los estudiantes. Stone (2014) por su parte, diseñó una rúbrica para la evaluación de habilidades en experimentación en estudiantes de biología en nivel secundaria.

Por último, se mencionan otros tres ejemplos de la medición del componente cognitivo del aprendizaje de las ciencias: Romine et al. (2016) con el diseño de la prueba *Quantitative Assessment of Socio-Scientific Reasoning* (QuASSR) que mide cuatro componentes de este tipo de razonamiento: complejidad, toma de perspectiva, indagación y escepticismo. El trabajo de Kruit et al. (2018) que realizaron una prueba para la medición de pensamiento científico y metacognición en estudiantes de educación básica. Y, en contexto nacional, Pérez-Morán (2010) con el diseño del *Examen de Trayecto del Área Metodológica* (ETAM) que tiene como propósito evaluar las habilidades o competencias en metodología de la investigación en estudiantes de último año de la licenciatura en Psicología.

En cuanto a los antecedentes del estudio y medición de los componentes afectivos del aprendizaje de las ciencias, se puede mencionar el trabajo de Glynn et al. (2007, 2011) que se interesaron en medir la motivación para el aprendizaje de las ciencias en estudiantes de ES de carreras científicas y no científicas. Producto de ambos estudios, diseñaron la prueba *Science Learning Motivation Questionnaire* (SLMQ-II) bajo el marco de la Teoría Social Cognitiva (TSC) de Bandura (1986). El objetivo del SLMQ-II es conocer los factores tanto intrínsecos como extrínsecos que permiten que los estudiantes mantengan una motivación en el aprendizaje de contenidos científicos, de esta manera, los docentes pueden realizar ajustes apropiados dentro del aula para mantener la motivación de los estudiantes en los contenidos.

Otro ejemplo es el trabajo de Velayutham et al. (2011), quienes desarrollaron una prueba para medir la motivación y la autorregulación de los estudiantes para el aprendizaje de las ciencias. El instrumento se aplicó a 1360 estudiantes de secundaria y recopilaron información cualitativa de docentes con experiencia en la enseñanza de las ciencias. De acuerdo con los desarrolladores del instrumento, la

información extraída de la prueba se puede utilizar para reorientar los contenidos y evaluar la efectividad de las intervenciones.

En la misma línea, Ardura y Pérez-Bitrián (2018) adaptaron el SMQL-II (Glynn et al., 2011) al castellano para estudiantes de secundaria. El cuestionario fue adaptado específicamente para medir la motivación de los estudiantes para aprender contenidos de física y química. El *Cuestionario de Motivación Científica II* está organizado en cinco subescalas similares a las utilizadas en el cuestionario original: motivación profesional, autoeficacia, autodeterminación, motivación intrínseca y motivación de grado.

Otra variable relevante que ha sido foco de atención en la formación científica de los estudiantes a lo largo de décadas es la actitud de los estudiantes hacia la ciencia (Osborne et al., 2003). Las actitudes hacia la ciencia cobraron relevancia en el campo de la investigación educativa debido a la disminución de ingresos a carreras científicas (Osborne et al., 2003; Vázquez-Alonso y Manassero-Mas, 2015). Desde entonces, las actitudes hacia la ciencia han sido un importante objeto de estudio que ha resultado en la creación de pruebas psicométricas para su evaluación (Munby, 1997; Vázquez-Alonso y Manassero-Mas, 1997; Blalock et al., 2008; Summers y Abd-El-Khalick, 2018).

Uno de los antecedentes más relevantes en la medición de las actitudes hacia la ciencia es el *Test of Science Related Attitudes* (TOSRA) de Fraser (1982). Este instrumento fue diseñado para estudiantes de nivel secundaria, y es uno de los instrumentos más utilizados para la medición de las actitudes hacia la ciencia ya que se basa en el concepto de Klopfer (1973), mismo que es utilizado en la prueba PISA para la evaluación del dominio afectivo del aprendizaje de las ciencias. El TOSRA se compone de siete subescalas: implicaciones sociales de la ciencia, normalidad de los científicos, actitud hacia la investigación, adopción de actitudes científicas, el disfrute de las experiencias de aprendizaje de las ciencias, el interés en el aprendizaje de las ciencias en tiempo de ocio, y el interés por una carrera científica.

Otro instrumento con varias décadas de antigüedad es la adaptación al castellano realizada por Vázquez-Alonso y Manassero-Mas (1997) de la prueba *Wareing Attitudes Toward Science Protocol*, originalmente diseñado en inglés por Wareing (1982). *El Protocolo de Actitudes Relacionadas con la Ciencia* o PAC por sus siglas, consta de 50 ítems organizado en cuatro subescalas: imagen global, aspectos sociales, aspectos escolares y características científicas. Posteriormente, Rodríguez et al. (2007), adaptaron el PAC para su uso en estudiantes universitarios colombianos. En esta versión, las subescalas son: enseñanza de la ciencia, imagen de la ciencia, incidencia social de la ciencia y características de la ciencia.

Con respecto a la modificación de las actitudes hacia la ciencia por medio de la intervención educativa, se puede mencionar el trabajo de Siegel y Ranney (2003) en estudiantes de bachillerato. Este estudio tenía como objetivo medir el impacto de dos programas educativos en las actitudes hacia la ciencia de los estudiantes, uno basado en la toma de decisiones de cuestiones socio-ambientales y el otro orientado al desarrollo de la argumentación lógico-científica. Para el logro de este objetivo, se diseñaron tres instrumentos: uno para la aplicación antes del programa, después del programa y el tercero para medir el mantenimiento de las actitudes de los estudiantes, posterior a recibir la intervención.

Otros ejemplos de la medición de las actitudes hacia la ciencia, son la prueba *Asian Student Attitudes Toward Science Class Survey* (ASATSCS), el *Dimensions of Attitude toward Science* (DAS) y el *Preschool Teacher Attitudes and Beliefs toward Science Teaching Questionnaire* (P-TABS). El primero fue diseñado y validado por Wang y Berlin (2010) con el objetivo de evaluar las actitudes hacia la ciencia de estudiantes de educación básica. La prueba tiene 30 ítems y consta de tres subescalas: el disfrute de la ciencia, la confianza en la ciencia, y la importancia de la ciencia. Por otro lado, el DAS (Van Aalderen-Smeets y Walma van der Molen, 2013) mide las actitudes hacia la enseñanza de la ciencia en docentes de educación básica, cuenta con 28 ítems que evalúan las creencias cognitivas sobre la ciencia, los estados afectivos y el control percibido. Por último, el P-TABS diseñado por Maier et al.

(2013) mide la actitud hacia la ciencia de los docentes de preescolar. Esta prueba consta de 35 ítems y se organiza en tres subescalas: comodidad del docente, beneficio para el infante y desafíos.

Otro ejemplo es el de Unfried et al. (2015) en donde se desarrollaron dos instrumentos, uno para medir las actitudes hacia las STEM y el otro para evaluar el interés de los estudiantes de educación básica por elegir una carrera dentro de las STEM. Este instrumento se compone de cuatro subescalas: el interés por las matemáticas, el interés por las ingenierías, el interés por la tecnología y las habilidades del siglo XXI.

Otras dos propuestas de medición de las actitudes hacia la ciencia son la de Navarro et al. (2016) y Hillman et al. (2016). En el estudio de Navarro et al. (2016) se adaptó la *Prueba de Actitudes Relacionadas con la Ciencia* basada en el TOSRA de Fraser (1982) para medir la actitud hacia la ciencia en estudiantes de secundaria de habla hispana. El estudio tuvo como objetivo analizar las propiedades psicométricas del instrumento tales como la estructura interna, la validez discriminante y concurrente y la confiabilidad de la prueba. Por otro lado, Hillman et al. (2016) desarrolló el instrumento My Attitudes Toward Science (MATS), para estudiantes de educación básica. El MATS tiene 40 ítems organizados en cuatro subescalas: la actitud hacia la asignatura de ciencia; el deseo de convertirse en científico; el valor de la ciencia para la sociedad y la percepción de los científicos. Los investigadores contaron con la participación de 32 profesionales científicos y docentes en la recolección de evidencias de validez en el aspecto del contenido.

Se puede mencionar también el trabajo de Hartman et al. (2017) en donde se desarrolló la Escala de Credibilidad de la Ciencia (CoSS por sus siglas en inglés) una escala de seis ítems para medir la actitud de los ciudadanos hacia la credibilidad de la ciencia con el objetivo de predecir opiniones de los encuestados en creencias científicas específicas.

Por último, es importante hablar del trabajo de Summers y Abd-El-Khalick (2018) con el diseño del cuestionario en inglés *Behaviors, Related Attitudes, and Intentions toward Science* (BRAINS) para

la medición de actitudes hacia la ciencia en estudiantes de secundaria. Este cuestionario surge a partir de las principales revisiones de instrumentos para medir las actitudes hacia la ciencia y sus propiedades psicométricas. Además de hacer una recopilación de estos instrumentos, Summers y Abd-El-Khalick (2018) señalan los principales problemas detectados en las mediciones de la actitud hacia la ciencia. Entre ellos, la ausencia de una teoría que explique el constructo a medir. El estudio involucra el diseño y validación del cuestionario BRAINS haciendo uso de la Theory of Planned Behaviour (TPB) de Ajzen y Fishbein (1977).

En contexto nacional, es relevante mencionar la Encuesta sobre la Percepción Pública de la Ciencia y la Tecnología (ENPECYT). La ENPECYT en primera instancia fue una iniciativa del Consejo Nacional de Ciencia y Tecnología (CONACYT) con la finalidad de conocer el nivel de conocimiento y entendimiento de la población mexicana con respecto a las actividades científicas y tecnológicas realizadas en el país (INEGI, 2021). La ENPECYT tenía como objetivo conocer el grado de interés que tienen los ciudadanos en áreas específicas de la ciencia como la física, la química, la biología, las ciencias de la salud, entre otras. Asimismo, se buscaba saber el nivel de cultura científica de los ciudadanos, así como la percepción del papel social de la ciencia y la tecnología, el conocimiento en investigación básica, el papel de la ética de la ciencia y la percepción de los gastos del gobierno en materia de inversión en desarrollo científico y tecnológico.

Esta encuesta también buscó conocer las opiniones sobre las aplicaciones de la ciencia moderna y las pseudociencias en la población mexicana. Se abordaron variables como la religión, la fe, la creencia en la astrología, la medicina alternativa, las creencias paranormales, etc. Sin embargo, aunque la ENPECYT se aplicó cada dos años casi desde su surgimiento, la última edición registrada en el portal del INEGI es la del 2017.

Otro esfuerzo nacional por medir el grado de cultura científica y la percepción de los mexicanos ante la ciencia fue el estudio realizado por Padilla et al. (2020) en conjunto con la Sociedad Mexicana

para la Divulgación de la Ciencia y la Técnica A. C. (SOMEDICYT) en donde se aplicó una encuesta a investigadores, divulgadores científicos, tomadores de decisiones en materia de CyT y población general sobre el conocimiento científico que necesitan los ciudadanos en el marco de la cultura científica. Es relevante mencionar que en el marco conceptual de lo que en este estudio se denomina Cultura Científica, hay elementos denominados axiológicos que integran componentes como las actitudes, intereses y valores, como lo son la comprensión de la ciencia y la percepción de la ciencia.

Por último, en cuanto a la medición de las actitudes hacia la ciencia, es menester mencionar que hay tres grandes antecedentes de revisiones sistemáticas de instrumentos para la medición de las actitudes hacia la ciencia: la de Munby (1997), Blalock (2008) y Toma y Lederman (2020). Estas revisiones tienen como objetivo la búsqueda y sistematización de los instrumentos existentes para la medición de las actitudes hacia la ciencia. En rasgos generales, las problemáticas principales que los investigadores han detectado en la medición de las actitudes hacia la ciencia, son (1) la ausencia de teoría que respalde el instrumento y (2) la falta de reportes métricos de la recopilación de evidencias de validez.

Otro constructo de creciente interés para los investigadores educativos interesados en el aprendizaje de las ciencias y la psicología vocacional es la autoeficacia en el aprendizaje de las ciencias. Así pues, se distinguen dos principales constructos similares: la autoeficacia para la ciencia (Knaggs y Sondegeld, 2015) y la autoeficacia para la investigación (Bieschke et al., 1996; Hemming y Kay, 2010). Para mencionar algunos ejemplos de la medición de la autoeficacia en investigación: en el estudio de Swenson-Britt y Berndt (2013) se diseñó, desarrolló y validó una prueba para medir la autoeficacia en investigación en estudiantes de enfermería. En español, un antecedente es la escala de Creencias Personales para la Investigación diseñada por Criollo, Romero y Fontaines-Ruiz (2017). En este estudio de metodología mixta se diseñó el instrumento antes mencionado, que se aplicó en estudiantes de licenciaturas dentro y fuera del área científica y se complementó con información cualitativa sobre la percepción de los estudiantes con respecto al trabajo de investigación científica.

En México, es importante mencionar el trabajo de Reyes Cruz y Gutiérrez Arcero (2015) el cual tuvo como objetivo comprender el sentido de autoeficacia en la investigación de estudiantes de maestría. Si bien este estudio es cualitativo, se señala que la autoeficacia en investigación se ha explorado con más profundidad en estudiantes de doctorado. Esto es, los investigadores apuntan que la autoeficacia en actividades de investigación científica ha sido menos estudiada en estudiantes de nivel licenciatura y maestría.

Asimismo, aunque la presente tesis no se centrará en las variables asociadas del aprendizaje de las ciencias de los estudiantes, es menester mencionar algunas de ellas. Dentro de estas variables asociadas se reconocen variables propias del estudiante (o individuales) como el género (Lippa, 1998; Osborne et al., 2003; Buccheri et al., 2011), la identidad científica (Chemers et al., 2011; Trujillo y Tanner, 2014; Stets et al., 2016), el sentido de pertenencia con la comunidad científica (Good et al., 2012), y la ansiedad hacia la ciencia (Udo et al., 2004) por mencionar algunos ejemplos. Por otro lado, en las variables asociadas de contexto se pueden señalar la relevancia de los docentes como formadores y motivadores del interés científico (Osborne et al., 2003), la influencia de los pares (Stets et al., 2016), la educación y actitud hacia la ciencia de los padres (Germann, 1994; Breakwell y Robertson, 2001; Perera, 2014) y los medios de comunicación científica (Garza-Almanza, 2016; Stekolschik et al., 2007; Kawamoto et al, 2013; Yuan, et al., 2019).

En resumen, existe una diversidad de propuestas de medición, así como de constructos y dimensiones asociados al aprendizaje de las ciencias. Como se pudo constatar, existe en la literatura una diversidad de instrumentos de medición para evaluar este objeto de estudio: encuestas, escalas, entrevistas, exámenes, cuestionarios y ambientes interactivos de simulación. La investigación al respecto aún es un tema que interesa tanto a organismos internacionales dedicados a elevar la alfabetización científica de los ciudadanos de todas las edades, como a investigadores educativos y vocacionales en todos los niveles desde la educación básica hasta la educación superior nivel doctorado, incluyendo

estudiantes y docentes. Así también es de provecho para los interesados en el diseño, desarrollo y validación de instrumentos de medición de las variables relevantes y asociadas al aprendizaje de las ciencias.

1.2. Planteamiento del problema

Los objetivos y fines de la educación científica son el generar interés y despertar la motivación en los estudiantes por carreras científicas y por la formación científica de los ciudadanos que comienza con la educación obligatoria. A su vez esta formación en los estudiantes es una necesidad urgente tanto en su desarrollo personal como social (Macedoi, 2016). De acuerdo con Glynn et al. (2011) quienes desarrollaron el SLMQ-II y colaboraron en los *Handbook of Research on Science Education*, es de suma relevancia conocer cuáles son los factores que elevan o disminuyen la motivación para comprender y adquirir los contenidos relacionados a las ciencias en los estudiantes de licenciatura. De igual manera, dichos autores consideran importante que los docentes conozcan esta información ya que podría permitirles realizar intervenciones educativas más eficientes. Por ello, con el propósito de conocer el estado del dominio afectivo en el aprendizaje de las ciencias en los estudiantes de licenciatura, se vuelve de suma relevancia la medición de estos constructos, ya que permitiría brindar conocimiento a una de las cuestiones más relevantes de la educación científica en los últimos años: *¿por qué los estudiantes no aprenden lo que se les trata de enseñar?* (Anderson, 2009).

En el caso de la educación básica, que es crucial para el desarrollo de la alfabetización científica en los ciudadanos, Macedoi (2016) en el marco del documento de Educación Científica para la UNESCO y el Foro Abierto de Latinoamérica y el Caribe, menciona que una posible explicación es la visión rígida de la ciencia, aquella que concibe el desarrollo científico como una serie de conocimientos acabados, acumulativos y lineales, y que no considera los periodos de crisis y de revoluciones científicas. En consecuencia, esta visión dio como resultado una práctica tradicional de enseñanza de las ciencias que los estudiantes perciben como “difícil, aburrida y destinada a solo a unos pocos estudiantes” (pág 8).

La educación científica debería ser punto de partida para fomentar en los estudiantes las actitudes positivas hacia la ciencia, la motivación y el deseo de aprender, así como la duda y la pregunta constante.

Además de expresar con claridad uno de los principales retos: la superación de los modelos tradicionales de la enseñanza de las ciencias basados en las lógicas propias de las disciplinas científicas. Es importante mencionar que, aunque estas características de la educación científica en latinoamérica brindan una dirección a la cual las políticas educativas de cada país deberían dirigirse, incluso pautas para la evaluación educativa, no se encontró ninguna mención o dirección de la medición de estos componentes esenciales de la educación científica en los principales agentes educativos.

Entre las principales dificultades presentes en la evaluación del dominio afectivo del aprendizaje de las ciencias en estudiantes de ES en México, se observó la ausencia de instrumentos de tipo diagnóstico para medir la motivación, las actitudes hacia la ciencia y la autoeficacia en el aprendizaje de las ciencias con el propósito de mejorar los aprendizajes.

Asimismo, es importante resaltar que se cuenta con más información acerca de la medición de las actitudes hacia la ciencia; por ejemplo, revisiones sistemáticas sobre instrumentos que brindan un panorama más amplio de la medición de este constructo a lo largo de cuatro décadas a diferencia de la medición de la motivación y la autoeficacia en el aprendizaje de las ciencias. A su vez, el componente afectivo del aprendizaje hacia la ciencia se ha evaluado con mayor frecuencia en estudiantes de educación básica, o el equivalente según el sistema educativo por país. En tanto que, los componentes cognitivos se suelen medir mayormente en estudiantes de ES pero pertenecientes a las ciencias naturales y exactas. Se listan a continuación las problemáticas que se encontraron en la revisión de la literatura:

(1) No hay variedad ni fácil acceso a instrumentos en español. La mayoría de los instrumentos son en inglés (Wang y Berlin, 2010; Velayutham, et al., 2011; Van Aalderen-Smeets y Walma van der Molen, 2013; Hartman et al., 2017). Con respecto al idioma en que están escritos los instrumentos, hay más variedad para la medición de las actitudes hacia la ciencia que para la motivación y la autoeficacia en el

aprendizaje de las ciencias. Sin embargo, aquellos que se tradujeron al castellano, no detallan el procedimiento que siguieron para cuidar la fidelidad del constructo.

(2) La mayoría de los instrumentos de actitudes hacia la ciencia y de motivación para el aprendizaje de las ciencias, fueron diseñados para su uso en educación básica (Velayutham et al., 2011; Abd-El-Khalick, 2015; Blanco-Blanco et al., 2016; Navarro et al., 2016) o para docentes de educación básica (Maier et al., 2013; Hilman et al., 2016). Es decir, el dominio afectivo es mayormente explorado en niños y adolescentes que en población adulta. Este punto es retomado en la revisión de instrumentos para la medición de actitudes hacia la ciencia de Toma y Lederman (2020), en donde se menciona que, en los resultados de su búsqueda, solo dos instrumentos para la medición de este constructo tenían como población objetivo a los estudiantes universitarios.

(3) En ES se ha explorado poco el dominio afectivo del aprendizaje de las ciencias, concentrándose en el dominio cognitivo, por ejemplo, la redacción científica, el pensamiento crítico o el pensamiento científico (Crotwell et al., 2013; Etkina et al., 2006; Kruit et al., 2018).

(4) Los rasgos e ítems de estos instrumentos tanto para educación básica como para educación superior se han diseñado de forma restringida y particularizada para ciencias específicas, por ejemplo: la química, la biología o la física (Crotwell et al., 2013; Etkina, 2006; Kruit et al., 2018). Asimismo, aunque se miden componentes como la observación, el diseño de experimentos y la redacción científica, los ítems hacen referencia a actividades propias de estas ciencias, que no podrían ser aplicados a estudiantes de ciencias sociales, educativas y del comportamiento. Por ello, se señala la necesidad de desarrollar y validar instrumentos que contemplen actividades o procesos de la ciencia que se realizan independientemente del campo de aplicación de conocimiento. Esto con la finalidad de contar con pruebas aplicables tanto en estudiantes de ciencias sociales, educativas y del comportamiento, como de ciencias naturales y exactas.

En este punto, sería importante señalar el riesgo de realizar interpretaciones y afirmaciones sobre las actitudes hacia la ciencia, la motivación, la curiosidad científica o el interés por realizar investigación científica, utilizando instrumentos que en los ítems evocan actividades exclusivas de ciencias particulares.

(5) Algunos de los instrumentos encontrados en la revisión de la literatura abordan la medición desde un enfoque más tradicional de la ciencia, en donde lo científico es lo relacionado a las ciencias exactas y naturales, dejando de lado los perfiles de las ciencias sociales y humanas. Si bien epistemológicamente es aceptado que todas las ciencias sin importar su objeto de estudio hacen uso de estrategias investigativas para la generación de conocimiento científico (Bunge, 2015), esta postura no ha permeado del todo en los instrumentos de medición que tienen como objetivo medir el dominio afectivo del aprendizaje de las ciencias.

(6) En cuanto a las cuestiones de tipo metodológicas se observa que, en instrumentos registrados desde el año 1986 hasta el 2020, la mayoría reportan la confiabilidad, pero se recopilan solo algunas evidencias de validez de constructo (Blalock et al., 2008; Toma y Lederman, 2020). Como ejemplo, para los problemas encontrados en la medición de las actitudes hacia la ciencia, se encontró que la mayoría de los instrumentos carecieron de uso, seguimiento y replicación (Blalock et al., 2008). En ese sentido, Toma y Lederman (2020), señalaron la ausencia de reportes de evidencias de validez de contenido y, por lo tanto, se desconoce la relevancia y exhaustividad de los ítems.

(7) La mayoría de los instrumentos relacionados con los constructos de interés de la presente investigación, no presentan evidencias de validez de constructo en el aspecto del contenido y que se adhieran a algún marco metodológico o a los *Estándares* para el diseño, desarrollo y validación de los instrumentos de medición como son los presentados por la AERA, APA y NCME (2018).

Dado los antecedentes anteriores y las problemáticas de investigación aquí planteadas, se puede establecer que son escasos los instrumentos de medición para explorar la motivación, las actitudes y la

autoeficacia en el aprendizaje de las ciencias, así como los estudios de dichas mediciones de los constructos de interés en poblaciones de estudiantes universitarios. En síntesis, con respecto al panorama de la medición de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias se considera lo siguiente: (a) no hay suficientes instrumentos para medir la amplia variedad de constructos relevantes para explorar la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias, en especial para estudiantes de educación superior; (b) no hay suficientes instrumentos en español y que se encuentren adaptados para población mexicana; (c) los instrumentos que se utilizan actualmente se enfocan en las actividades propias de las ciencias naturales y exactas, dejando de lado los perfiles asociados con las ciencias sociales, educativas y del comportamiento; y (d) se reporta mayoritariamente la confiabilidad y la consistencia interna de los instrumentos pero sólo en menor cantidad otras evidencias de validez y (e) se menciona la recopilación de evidencias de validez de constructo en el aspecto del contenido, pero no se ahonda en los procedimientos y resultados obtenidos. Esto en concordancia con lo que menciona Toma y Lederman (2020) sobre el desconocimiento de la relevancia y exhaustividad de los ítems.

1.3. Objetivos de investigación

Para atender la problemática planteada se propone como objetivo general de investigación diseñar, desarrollar y obtener evidencias de validez de contenido para tres escalas para medir la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias desde un enfoque integrador en estudiantes de ES nivel licenciatura. Para el logro de dicho objetivo general se proponen cuatro objetivos específicos:

- Revisar diversos modelos consolidados para la medición de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias.
- Diseñar y desarrollar tres escalas con el propósito de medir la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias en estudiantes universitarios de nivel licenciatura. Dicho objetivo tomando en cuenta modelos de medición consolidados y estándares de calidad técnica para la medición psicológica y educativa.
- Obtener evidencias de validez de constructo en el aspecto del contenido de las escalas desarrolladas para medir la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias en estudiantes universitarios mediante la estimación de distintos índices de concordancia entre jueces.
- Obtener evidencias de la imparcialidad de las escalas para medir la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias desde un enfoque integrador tomando en cuenta los criterios de la Evaluación de Diseño Universal (EDU) por su traducción (Pérez-Morán, 2014) para instrumentos de medición.

1.4. Justificación

El diseño, desarrollo y validación de instrumentos para medir la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias desde un enfoque integral en estudiantes de educación superior presenta una amplia diversidad de aportaciones (teóricas, metodológicas, instrumentales, prácticas, entre otras), y beneficios para el campo de la investigación y la medición educativa, en especial para el fomento y formación de vocaciones científicas.

Contar con instrumentos de medición que den cuenta de las evidencias de validez de constructo obtenidas mediante procedimientos rigurosos y alineados a estándares de calidad favorece la toma de decisiones informada de los usuarios potenciales directos e indirectos. En especial, se beneficia con estos instrumentos a estudiantes, docentes, orientadores vocacionales y autoridades educativas de ES a nivel licenciatura en la recopilación de información relevante para la toma de decisiones informada, que impulse el diseño de políticas educativas, planes de estudio, intervenciones educativas y estrategias pedagógicas, contribuyendo a la mejora educativa, y al desarrollo y fomento de la vocación científica.

Entre las aportaciones teóricas de la presente tesis, se resaltan los modelos de medición fundamentados en la revisión del estado del arte en el campo del dominio afectivo del aprendizaje de las ciencias. Las aportaciones metodológicas de este trabajo de investigación son tres instrumentos de medición y el procedimiento, en donde se detalla su diseño y desarrollo considerando las recomendaciones de los *Estándares* para el desarrollo de pruebas psicológicas y educativas de la AERA, APA y NCME (2018). Se pueden mencionar también las aportaciones instrumentales de esta tesis como el proporcionar a investigadores y especialistas en el campo una adaptación en línea del protocolo para la obtención de evidencias de validez de contenido en formato físico utilizado por Rodríguez de la Vega-Elu (2021) así como un ejemplo de *Manual de procedimientos* para la capacitación del panel de jueces (**Apéndice 2**).

De esta manera, se puede agregar que al contar con estos instrumentos de medición se beneficia a estudiantes de ES, que requieran de uso para la mejora de su aprendizaje en los contenidos relacionados a las ciencias, para ampliar su conocimiento de sí mismos en torno a la motivación, actitudes y autoeficacia en el aprendizaje del método científico, así como para contar con información que auxilie en su proceso de exploración vocacional. De igual modo, su uso y beneficio se puede extender a docentes y orientadores vocacionales para el diseño de estrategias pedagógicas que favorezcan el desarrollo y fomento de las vocaciones científicas. Así también, su uso se podría extender a autoridades educativas que soliciten información relevante para la toma de decisiones informada que impulse el diseño de planes de estudio o políticas educativas relacionadas con el fomento del gusto y persistencia en las actividades científicas e investigativas.

Otro aporte de la presente investigación es el estudio y medición de las variables relevantes en la formación científica de los estudiantes de licenciatura. Específicamente, una contribución al estudio y medición de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias en población mexicana. Las escalas y la información derivada, puede ser de utilidad para los sistemas educativos y desarrolladores de políticas en ciencia y tecnología al momento de impulsar y fomentar las vocaciones científicas en el nivel de educación superior.

En cuanto a la revisión de la literatura, permitirá visualizar un panorama de las variables que más influyen en la formación científica de los estudiantes, así como los modelos de medición más frecuentes en el campo de la educación científica. Aportará una fundamentación teórica y metodológica de la medición de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias, por medio de la sistematización de la información, el uso de palabras clave y la búsqueda de documentos en diferentes bases de datos y revistas indexadas. La revisión de la literatura permitirá también delimitar de manera teórica la definición clara de los constructos de acuerdo con las sugerencias de la AERA, APA y NCME

(2018). También será útil para el desarrollo del documento rector o ficha técnica de las escalas, así como las especificaciones del contenido de los instrumentos de medición.

El diseño y desarrollo de instrumentos para explorar la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias aporta una herramienta integral de calidad técnica al campo de la medición de la formación científica. El uso de esta prueba se podría extender a diversos actores involucrados en la investigación educativa, psicometría y estudio de la psicología vocacional. Asimismo, las escalas podrían ser utilizadas por estudiantes de posgrado que deseen realizar investigaciones en el campo educativo.

En cuanto al diseño y desarrollo de las escalas, se propone que los ítems eviten la exclusión de las ciencias sociales, educativas y del comportamiento. Por lo tanto, se pretende que los reactivos atiendan actividades científicas y de investigación en contextos de aprendizaje en ES, aplicables a distintas ciencias. Asimismo, las escalas permiten recopilar información sobre la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias de los estudiantes universitarios. Esta información podría ser de utilidad para los coordinadores de licenciatura, orientadores de psicopedagógicos y tutores encargados de orientar a los estudiantes en la toma de decisiones informadas sobre la elección de su área de especialización y formación curricular.

El uso de las escalas podría auxiliar también a los comités de posgrado, con la finalidad de tener más información del perfil de los aspirantes que buscan seguir con su formación profesional y científica. Esta herramienta podría ayudar a los investigadores a conocer las características del componente afectivo del aprendizaje de las ciencias de los estudiantes, identificar las principales fortalezas y áreas de oportunidad en la formación de nuevos científicos. Podría ser también útil para el diseño de estrategias pedagógicas para la enseñanza de los procesos investigativos.

Como parte de los objetivos de esta tesis, se desarrollará una propuesta metodológica para la obtención de evidencias de validez de la imparcialidad y del contenido de las escalas. Esta propuesta, la

cual consiste en un panel de expertos colaborativo y asincrónico puede ser utilizada y perfeccionada en trabajos posteriores. Asimismo, la recopilación de evidencias de validez de constructo en el aspecto de la imparcialidad y del contenido de las escalas, puede beneficiar a investigadores, psicólogos dedicados a la evaluación, psicómetras y desarrolladores de ítems; ya que brinda un ejemplo del procedimiento para la obtención de evidencias de validez apropiadas para la interpretación de los puntajes de las escalas. También, se presenta el reporte correspondiente a los resultados del cálculo de distintos índices de validez de contenido, con la finalidad de detallar la toma de decisiones respecto a las escalas, posterior a la obtención de los resultados del proceso de recopilación de evidencias de validez.

En el primer capítulo de la presente tesis se desarrollaron los antecedentes de medición de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias, así como las problemáticas encontradas en la literatura, últimas que dieron pie al esclarecimiento de los objetivos de investigación y la justificación del estudio. En el segundo capítulo se presentarán los fundamentos teóricos en dos grandes apartados, en el primero se presentarán propuestas teóricas de la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias utilizadas para el diseño y desarrollo de las escalas. En el segundo subapartado, se desarrollarán los fundamentos teórico-metodológicos del diseño, desarrollo y validación de contenido sugeridos por los Estándares de la AERA, APA y NCME (2018). En el tercer capítulo se presentará el método, el tipo de estudio, la descripción de los participantes del Comité de Validación de las Escalas (CVE), los instrumentos y materiales utilizados en el trabajo de investigación y el procedimiento del estudio en tres etapas. Posteriormente, en el cuarto capítulo se desarrollarán los resultados, divididos en los resultados del diseño y desarrollo de las escalas, así como la validación en torno al contenido y la imparcialidad de las escalas. Por último, en el quinto capítulo se presentarán las discusiones de los resultados, así como las limitaciones y recomendaciones para estudios posteriores.

II. MARCO REFERENCIAL

El presente capítulo se divide en dos principales apartados. En el primero, se describen los principales fundamentos teóricos del trabajo de investigación, con énfasis en los constructos de interés: la motivación, las actitudes y las creencias de autoeficacia en el aprendizaje de las ciencias. En el segundo apartado se mencionan los antecedentes y fundamentos metodológicos de la validez de constructo en el aspecto de la imparcialidad y del contenido según los Estándares de la AERA, APA y NCME (2018); así como el cálculo del Índice de Validez de Contenido (IVC).

2.1. Fundamentos teóricos de la medición de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias

Con la finalidad de desarrollar los fundamentos teóricos de los constructos relevantes en la presente tesis, es menester aclarar dos principales conceptos: uno tiene que ver con la definición de ciencia desde la que se parte, así como la división tradicional de las ciencias y, por otro lado, el aprendizaje de la ciencia o de las ciencias. Aunque existen en la literatura muchas definiciones de ciencia, se consideró apropiado retomar la definición propuesta por Pérez-Tamayo (2009), en donde se define a la ciencia como “la actividad humana creativa cuyo objetivo es la comprensión de la naturaleza y cuyo producto es el conocimiento, obtenido por medio de un método organizado en forma deductiva y que aspira a alcanzar el mayor consenso” (pag.1).

Ahora bien, tomando en cuenta que las escalas de medición diseñadas y desarrolladas en el presente documento pretenden ser útiles a todas las ciencias independientemente de la naturaleza de su objeto de estudio, es menester declarar que para realizar una clasificación apropiada de las ciencias, es

necesario referirse a teóricos clásicos en Filosofía de la Ciencia como Auguste Comte o Charles Sanders Peirce. En términos del primero, podemos extraer la idea del Positivismo, un sistema filosófico donde se busca desarrollar explicaciones materiales para todo fenómeno, incluidos los eventos humanos (Thyer, 2008). En este acercamiento, Comte supone que la organización del conocimiento científico es necesariamente jerárquica: Existen ciencias “más generales” cuyos contenidos facilitan el quehacer de ciencias “más específicas.”

Peirce, de forma similar, propone una organización jerárquica que comienza con una distinción tripartita: (a) ciencia de descubrimiento, (b) ciencia de revisión y (c) ciencia práctica. Donde la *ciencia de revisión* se refiere a aquella labor científica que concierne a la clasificación (el ordenamiento de objetos de acuerdo con ideas), incluyendo la actividad misma de organizar jerárquicamente las ciencias, así como la historia de la ciencia. La *ciencia práctica* para Peirce se refiere a aquella ciencia utilizada en la vida diaria, aunque por la naturaleza de una sociedad tecnológica, esta distinción es difícil de hacer y, finalmente, la *ciencia de descubrimiento* engloba disciplinas variadas, las matemáticas, la filosofía, y las ciencias empíricas especializadas, que a su vez estarían divididas en las *ciencias naturales* (ciencias físicas) por una parte, y las *ciencias sociales*, humanidades y psicología (ciencias psíquicas), por la otra. (Midtgarden, 2020). Dicho esto, es necesario mencionar que cuando se hace uso de un enfoque integrador en el presente trabajo, hace referencia a las ciencias tanto naturales como sociales de acuerdo con el enfoque de Pierce (1998 como se citó en Midtgarden, 2020).

Ahora bien, el aprendizaje de las ciencias es un tema amplio que ha cobrado relevancia en los últimos años. Entre los principales referentes globales que abonan a la investigación de la educación científica se pueden destacar las investigaciones recopiladas en los *Handbook of Research on Science Education* en sus distintas ediciones que dan cuenta de la evolución y los avances en este objeto de estudio en las últimas décadas. Algunas de las temáticas que se abordan son: el aprendizaje de las ciencias, cultura, género y sociedad en el aprendizaje de las ciencias, la enseñanza de las ciencias, el

currículum y la evaluación en ciencias y la docencia en la educación científica. Sin embargo, antes de entrar de lleno en los constructos de interés es necesario mencionar que en ES el aprendizaje de las ciencias ha sido más estudiado desde su componente cognitivo, es decir, el pensamiento científico. En el trabajo de Vázquez-Alonso y Manassero Mas (2018) se realiza una síntesis de las principales habilidades que constituyen el pensamiento científico, tomando en cuenta las propuestas de Feist (2006) y Kuhn (2012). Se destacan 11 habilidades que son: (1) observar con los sentidos, (2) categorizar lo observado, (3) reconocer patrones en objetos y sucesos, (4) crear y comprobar hipótesis, (5) pensar en causas y efectos, (6) apoyar eficazmente la teoría con pruebas, (7) visualizar el pensamiento científico, (8) ser consciente y controlar el pensamiento propio (9) usar metáforas y analogías, (10) usar la heurística de “confirmar rápido, revocar después” y (11) comunicar, evaluar, compartir, colaborar y pensar la información. En esta propuesta los factores actitudinales como las actitudes hacia la búsqueda y validación del conocimiento (curiosidad, creatividad, integridad, imparcialidad y apertura) forman parte importante del desarrollo del pensamiento científico (Vázquez-Alonso y Manassero Mas, 2018). En esta vía, se puede mencionar también el modelo de medición utilizado por PISA que contempla tanto el dominio cognitivo como el afectivo de la *Competencia Científica*. De este modo, en el dominio afectivo se evalúan constructos como la motivación, la curiosidad, el interés, las actitudes, la autoeficacia, entre otros.

Dicho esto, la investigación sobre el aprendizaje de las ciencias se puede resumir en el estudio del desarrollo de la alfabetización científica y la importancia de su adquisición en el público en general: estudiantes, trabajadores, consumidores, miembros de familia, entre otros. La alfabetización científica se puede definir como los conocimientos, las prácticas y los valores relacionados con las ciencias que se espera que los estudiantes aprendan a lo largo de su trayectoria como actores educativos (Anderson, 2007).

Asimismo, la alfabetización científica incluye una agencia por parte del estudiante en dos vías: *la agencia social*, es decir, los estudiantes exitosos obtienen reconocimiento por sus logros y habilidades que les permiten realizar trabajo útil para su comunidad y *la agencia en el mundo material*, los estudiantes con competencias científicas tienen la habilidad de describir, medir el mundo que los rodea, así como explicar y predecir fenómenos (Anderson, 2007).

De igual manera, es menester señalar que los investigadores del aprendizaje de las ciencias han llegado al acuerdo sobre la ineficacia de la educación científica formal, de manera que las deficiencias en la educación científica de los ciudadanos impactan no solo en el dominio de contenidos científicos específicos, si no que “una educación científica de baja calidad tampoco logrará despertar el interés, el placer y el gusto por aprender ciencias” (Macedoi, 2016, pág. 2).

Asimismo, Anderson (2007) reconoce tres tradiciones en el estudio del aprendizaje de las ciencias: la tradición del *cambio conceptual*, la *sociocultural* y la *tradición crítica* que concibe a la alfabetización científica como una forma de empoderamiento. Sin embargo, por motivos del foco principal de este trabajo de tesis se describirán únicamente las cinco directrices que estas tres tradiciones tienen en común con respecto al aprendizaje de las ciencias: (1) la historia intelectual y disciplinas afines, (2) las ideas sobre la naturaleza de la ciencia, (3) las ideas sobre los estudiantes y su aprendizaje en ciencias, (4) los objetivos y métodos de investigación y (5) las ideas para mejorar el aprendizaje de las ciencias.

La historia intelectual y las disciplinas afines al aprendizaje de las ciencias reconocen que las tres tradiciones provienen de otras disciplinas como la psicología, la sociología, la antropología, entre otras. Sin embargo, difieren en el reconocimiento de cuál disciplina ha sido la de mayor influencia. La directriz dos, *ideas sobre la naturaleza de la ciencia*, es de suma relevancia para la presente tesis ya que reconoce que el aprendizaje de las ciencias, así como la alfabetización científica depende de las ideas que los individuos tienen con respecto a la ciencia. Señalan también que la ciencia, más que un

conjunto de conocimientos y métodos para desarrollar nuevo conocimiento es una subcultura, un lenguaje y un conjunto de valores compartidos por los científicos.

Las ideas sobre los estudiantes de ciencias y su aprendizaje, tiene que ver con la idea general de que el estudiante es un agente independiente y dueño de su propio aprendizaje, que llega a él por medio de su conocimiento, creencias, roles y prácticas culturales. Se menciona también que el aprendizaje de las ciencias se da a partir de las interacciones entre los estudiantes, sus conocimientos y las prácticas de clase. En cuanto a los *Objetivos y métodos de investigación*, se menciona que hasta la fecha el aprendizaje de las ciencias se ha basado más en la adquisición de los contenidos con respecto a los métodos cuantitativos y cualitativos, con énfasis en los métodos cualitativos. Por último, en las ideas para mejorar el aprendizaje de las ciencias, las tres tradiciones están de acuerdo sobre el fracaso de la educación científica formal, sin embargo, también difieren en las barreras percibidas sobre el aprendizaje y las estrategias para corregirlas (Anderson, 2007).

Ahora bien, dentro del dominio afectivo del aprendizaje de las ciencias, la actitud y la motivación se estudian con la finalidad de inferir patrones de pensamientos, emociones y conductas relacionadas con la ciencia. Tanto la actitud como la motivación suelen ser estables y duraderas en las personas, aunque también pueden modificarse con una buena instrucción científica (Koballa y Glynn, 2009). De hecho, como lo mencionan Koballa y Glynn (2009) las actividades científicas en clase como las prácticas, el trabajo en laboratorios, las lecturas, así como los trabajos de campo tienen la finalidad de potenciar las actitudes de los estudiantes hacia la ciencia, así como la motivación para aprenderla.

De este modo, las actitudes y la motivación son constructos del dominio afectivo del aprendizaje de los más relevantes para el aprendizaje de las ciencias. Sin embargo, es probable que por la concepción tradicional de la ciencia como “objetiva” y que no considera las emociones, el dominio afectivo haya recibido menor atención en los estudiantes que el dominio cognitivo del aprendizaje de las

ciencias. Actualmente, es aceptado por la comunidad científica que el componente afectivo es indispensable para que se efectúe un aprendizaje significativo (Koballa y Glynn, 2009).

2.1.1. Motivación en el aprendizaje de las ciencias

La motivación para el aprendizaje de las ciencias, según Schunk, Pintrich y Meece luego retomada por Glynn et al. (2011) se define como un estado interno que despierta, dirige y sostiene la conducta de aprendizaje de la ciencia. Por lo tanto, de acuerdo con dichos autores los estudiantes motivados por aprender ciencias obtienen buen desempeño académico, ya que realizan conductas como la participación en clase y resolución de dudas, o la participación en laboratorios y grupos de estudio.

Clásicamente, la investigación en motivación ha sostenido que se trata de un proceso orientado hacia el logro de metas, sean éstas de origen intrínseco o extrínseco. Potvin y Hasni (2014) destacan, sin embargo, que parece existir una mayor afinidad en el discurso sobre motivación por la exploración del componente intrínseco, atendiendo con menor frecuencia a las influencias orientadas hacia el exterior.

Dicho lo anterior, y de acuerdo con Vasteenkiste et al. (2006), las conductas intrínsecamente motivadas son aquellas que no surgen de necesidades fisiológicas ni sus derivados, y para las cuales la recompensa es la satisfacción asociada a la actividad misma. Estos autores destacan que lo que parece caracterizar a la conducta intrínsecamente motivada es su fundamento en la necesidad de competencia y causación personal, íntimamente relacionándose con la percepción de autonomía. En el contexto del aprendizaje de la ciencia, notaremos entonces que la motivación intrínseca se presenta en aquellos individuos capaces de disfrutar la labor científica.

En el sentido opuesto, las conductas extrínsecamente motivadas serían aquellas que tienen su origen en el deseo de alcanzar un resultado deseable (por ejemplo, el recibir una recompensa o evitar un castigo). No obstante, esto no significa que la motivación extrínseca se trate únicamente sobre actuar

pasivamente en reacción al medio. Como Ryan y Deci señalan (2000), es posible que un individuo adopte voluntariamente la determinación de realizar una tarea externamente impuesta si ésta corresponde con la idea de valor o utilidad que se posee. De esta forma, la motivación extrínseca puede guardar relación con los deseos de adquirir poder o estatus de una persona, vinculándose similarmente con la identidad.

En cuanto a la medición de este constructo, el antecedente más relevante a señalar es el *Science Learning Motivation Questionnaire* desarrollado por Glynn et al. (2007; 2011). El SMLQ-II, basado en la Teoría Social Cognitiva responde a una clara necesidad de saber qué factores motivan a los estudiantes dentro de las aulas y qué pueden realizar los docentes para mejorar estas condiciones. De igual manera, el SMLQ-II busca que los docentes tengan una herramienta para recopilar información de los estudiantes con el propósito de optimizar la efectividad de las clases, así como el uso por parte de los investigadores permite conocer la motivación de los estudiantes para aprender ciencias en las materias dentro de la universidad (Glynn et al. 2011). De acuerdo con la Teoría Social Cognitiva, la motivación es un constructo que se compone de diversos atributos como lo son la motivación intrínseca, la motivación extrínseca, la autodeterminación y la autoeficacia, de manera que estos atributos se relacionan y contribuyen de manera positiva en la dirección y el sostenimiento en el aprendizaje de los estudiantes (Glynn, et al. 2011).

El SLMQ-II está dividido originalmente en las siguientes 5 subescalas: motivación intrínseca, autodeterminación, autoeficacia, motivación por las calificaciones, y motivación por la carrera. Éstas últimas dos, en la primera versión del instrumento, formaban parte de una única subescala denominada motivación extrínseca, pero se han separado para mejorar la precisión con que se pudiesen medir las metas de los estudiantes universitarios. La primera dimensión, motivación intrínseca, hace referencia a la satisfacción misma de aprender ciencia. La segunda dimensión, autodeterminación, se refiere al control que los estudiantes creen tener sobre su aprendizaje de la ciencia. La tercera dimensión, autoeficacia, se refiere a la creencia de los estudiantes de que pueden tener un buen desempeño en la ciencia. Por último,

la cuarta dimensión, la motivación extrínseca, se relaciona con el aprendizaje de la ciencia como un medio para un fin tangible, como lo es finalizar una carrera, u obtener una calificación satisfactoria.

2.1.2. Actitudes hacia la ciencia

El estudio de las actitudes se remonta varias décadas atrás en la historia de la psicología, al igual que existen en la literatura múltiples definiciones de lo que es una actitud. Para retomar el estudio y medición de las actitudes, se acudió al *Handbook of Attitudes* (Albarracín, Johnson y Zanna, 2005) en donde se señala a Allport como uno de los principales antecedentes del estudio de las actitudes desde la psicología social. Allport (1935, como se citó en Albarracín et al., 2005) define las actitudes como un estado mental y neuronal de preparación, organizado a través de la experiencia, que ejerce una influencia directiva o dinámica sobre la respuesta del individuo a todos los objetos y situaciones con los que se relaciona. Por otro lado, de una manera más resumida, Eagly y Chaiken (1993) definen la actitud como una tendencia psicológica que se expresa al evaluar una entidad en particular con cierto grado de favor o desagrado.

En cuanto a la medición de las actitudes, de acuerdo con Meliá (1990) es en la evolución del trabajo de Thurstone en donde se encuentra uno de los antecedentes más importantes para la medición de atributos psicológicos como las actitudes y las opiniones. Dicho autor, propone que las actitudes son un tema complejo que no puede delimitarse a un único índice numérico, sin embargo, no las exime de ser medidas. Para este autor una actitud es la acción potencial hacia el objeto con respecto solo a la pregunta de si el potencial de la acción será favorable o desfavorable hacia el objeto.

Por ende, tal y como lo señalan Albarracín et al. (2005), las actitudes toman una dirección evaluativa, la actitud es favorable o desfavorable de forma específica. Lo relevante de medir las actitudes hacia algo en específico es que éstas se relacionan con el comportamiento. Es decir, la actitud funciona como un indicador de la conducta, sin embargo, no debe confundirse con la ejecución de la conducta en

sí (Hernández-Sampieri et al., 2014). Por ejemplo, si la actitud de una persona hacia una buena alimentación es favorable, no significa que se alimente de manera nutritiva, pero es más probable que adopte posteriormente un régimen de comida saludable.

Crawley y Cobala (1994, como se citó en Ab-El-Khalick, 2015) indican que los investigadores educativos estudiosos de las actitudes, se han inclinado por modelos de psicología social para explicar las actitudes como la Teoría de la Acción Razonada de Ajzen y Fishbein (1977). Actualmente, este modelo que vincula la relación entre creencias, opiniones, valores y actitudes es de los más estudiados, utilizados y aplicados a una gran cantidad de variables en diversos campos como la educación, la salud, tanto física como mental, entre otros (Albarracín et al., 2005).

Como se ha mencionado anteriormente, las actitudes tienen una relación consistente con la realización de una conducta o tarea. En contexto del aprendizaje, el estudio de las actitudes de los estudiantes es relevante ya que permite a investigadores y creadores de políticas educativas desarrollar programas que favorezcan una adecuada experiencia educativa. Similarmente, puede distinguirse un lazo entre el incremento del interés de estudiantes por un cierto contenido y su capacidad para aprender el mismo, por lo que resulta vital para la investigación de prácticas de docencia efectivas; especialmente aquellas que pueden estimular experiencias afectivas positivas en los estudiantes, como la disposición entusiasta (Keller, Neumann y Fischer, 2016).

Asimismo, de acuerdo con Koballa (1988, como se citó en Koballa et al., 2007) el objetivo principal de conocer las actitudes en los estudiantes es poder predecir su conducta. El estudio y medición de las actitudes y la motivación en el aprendizaje de las ciencias es relevante para comprender el impacto de estos constructos en el desempeño de los estudiantes en materias de ciencias, al mismo tiempo que se reconoce a la actitud y la motivación como las variables más importantes en el estudio del componente afectivo del aprendizaje de las ciencias (Koballa et al. 2007).

La actitud hacia la ciencia o actitudes hacia la ciencia, como objeto de estudio, interesa a investigadores educativos y vocacionales desde hace por lo menos cuatro décadas (Tytler y Osborne, 2011, Navarro et al., 2015). Es necesario precisar, que este constructo carece de una definición única ya que dentro de las actitudes hacia la ciencia se integran componentes tan variados como los intereses científicos, así como las actitudes hacia los científicos, las carreras científicas, los métodos de enseñanza de las ciencias, el plan de estudios de ciencias o las materias científicas en el aula (Abd- El- Khalick, 2015).

También es necesario mencionar que, autores como Aiken y Aiken (1969) y Osborne et al. (2003) destacaron los problemas adyacentes a la ausencia de una definición clara de este constructo, como su medición. Actualmente se realizan esfuerzos por clarificar este concepto en trabajos como el de Abd-El-Khalick (2015) y Summers y Abd-El-Khalick (2018).

Es necesario también diferenciar la actitud hacia la ciencia de las actitudes científicas. Las actitudes científicas se definen como la adopción de enfoques específicos de toma de decisiones y resolución de problemas. Es decir, se trata de atributos cuya presencia se reconoce como característica del trabajo científico: el deseo de saber y comprender, una disposición de cuestionamiento ante las afirmaciones, la búsqueda de datos y su significado, la necesidad de verificación; y el respeto por la lógica (Comisión de Políticas Educativas, 1962, en Osborne et al., 2003).

Las actitudes hacia la ciencia, por otro lado, pueden describirse como los sentimientos, creencias y valores que se tienen sobre un objeto que puede ser la realización de la ciencia, la ciencia escolar, el impacto de la ciencia en la sociedad, o los científicos en sí mismos. Para la realización de este estudio, y en concordancia con los objetivos de investigación, se utilizará la definición de Klopfer (1973), quien define la actitud hacia la ciencia como un conjunto de comportamientos afectivos en la educación científica. Klopfer enumera seis categorías para estos comportamientos afectivos, que más tarde Fraser (1982) describe de la forma siguiente:

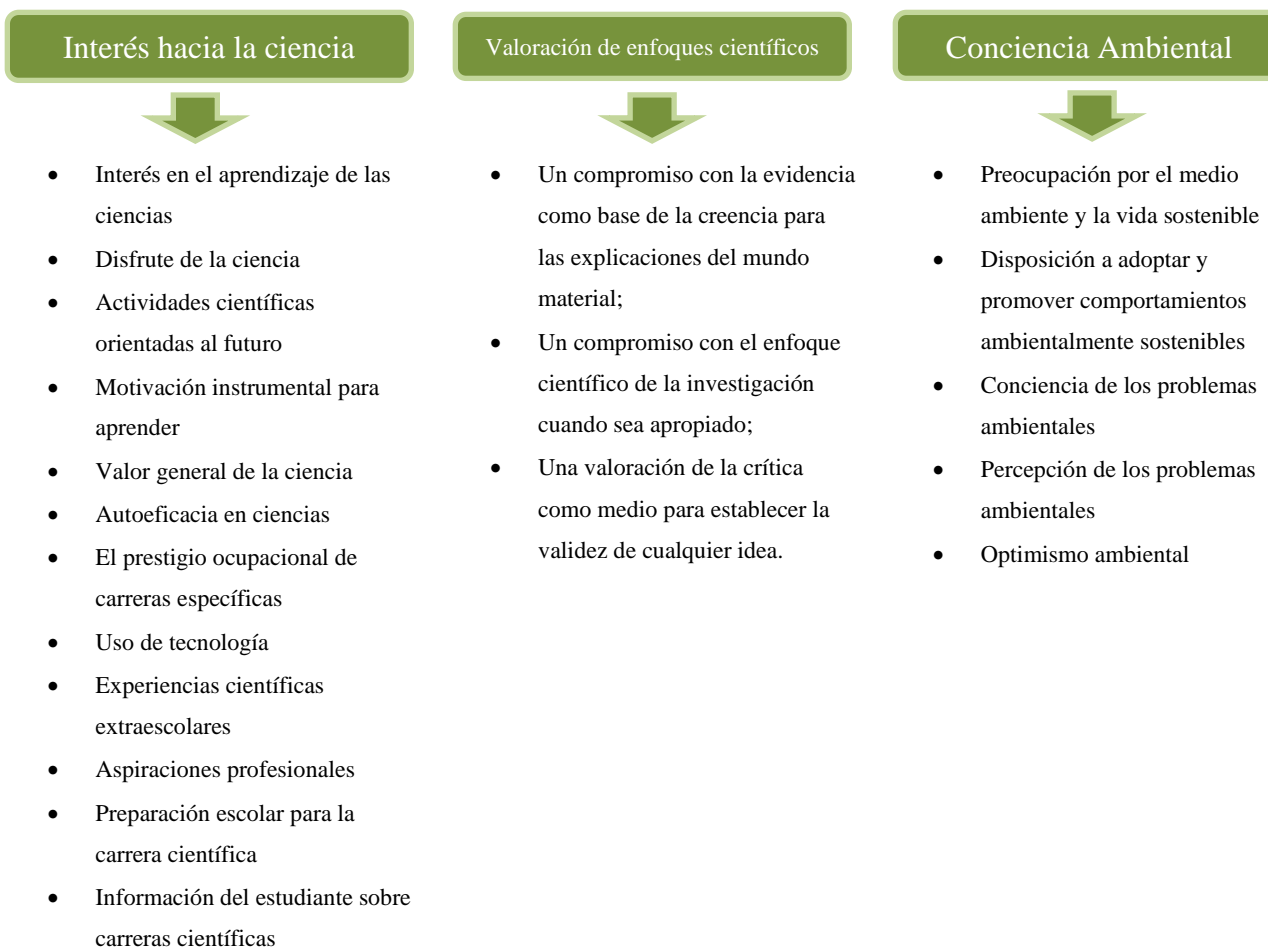
- (1) la manifestación de actitudes favorables hacia la ciencia, la actitud hacia los beneficios sociales y los problemas que acompañan al progreso científico;
- (2) la normalidad de los científicos, la apreciación de que los científicos son personas comunes,
- (3) la aceptación de la investigación científica como forma de pensamiento, la actitud hacia la investigación y experimentación científica como métodos para obtener información respecto del mundo natural,
- (6) la adopción de "actitudes científicas", la disposición para la incorporación de características como la curiosidad, apertura, y la voluntad para modificar las opiniones propias,
- (7) el disfrute de las experiencias de aprendizaje de las ciencias, el disfrute de las actividades escolares relacionadas con la ciencia,
- (8) el desarrollo de intereses en actividades relacionadas con la ciencia; el interés por realizar actividades de ocio relacionadas con la ciencia y,
- (9) el desarrollo de un interés en seguir una carrera o trabajo relacionado con la ciencia, es decir, el interés en la ciencia como carrera profesional.

La definición de Klopfer (1973) de las actitudes hacia la ciencia, es el referente teórico para el dominio afectivo de la prueba PISA (OCDE, 2015). Además, las actitudes hacia la ciencia en este modelo de medición se constituyen por subconstructos: (1) el interés por la ciencia, (2) la valoración de enfoques científicos para la investigación, y (3) la conciencia ambiental. El interés por la ciencia se reconoce como una curiosidad por la ciencia, hacia los problemas y esfuerzos relacionados con la ciencia, así como una voluntad para adquirir tanto los conocimientos como las habilidades científicas. Asimismo, un interés sostenido en la ciencia, en carreras o especializaciones relacionadas a lo científico.

A su vez, estos subconstructos o dimensiones se evalúan a través de otros constructos relacionados, organizados como se detalla en la **Figura 1**.

Figura 1.

Modelo de medición de Actitud hacia la ciencia en la prueba PISA.



Por último, aunque las actitudes siguen siendo un objeto de estudio relevante e interesante para psicólogos e investigadores en distintos campos científicos, como puede constatar el cuerpo de investigación a la fecha, la presencia de actitudes como preferencias y sentimientos expresados hacia un objeto no es necesariamente un predictor de conductas relacionadas con el trabajo científico. Es posible que, la presencia de actitudes con mayor fuerza pudiera alterar el resultado conductual final. Similarmente, si a través del análisis de consecuencias una conducta pudiese llevar a un individuo a una

situación percibida como desfavorable, una actitud positiva hacia la ciencia quizá no resulte suficiente (Osborne et al., 2003).

2.1.3. Autoeficacia Percibida en el Aprendizaje de la Ciencia e investigación

Por último, en el marco del aprendizaje de las ciencias, se puede mencionar la autoeficacia percibida tanto en ciencia (Bieschke, Bishop y Garcia, 1996) como en investigación (Bieschke et al., 1996; Reyes Cruz y Gutiérrez Arceo, 2015). Para ahondar en este constructo se debe definir primeramente qué es la Autoeficacia Percibida y cuáles son los principales indicadores que la componen.

Para definir la Autoeficacia Percibida se retoma de igual manera la Teoría Social Cognitiva (TSC), desarrollada por Albert Bandura (1977). Pajares (1996) señala que este concepto surge en el año 1977 con la publicación de Bandura "*Self-Efficacy: Toward a Unifying Theory of Behavioural Change*". Posterior a su publicación, el aporte de la Autoeficacia Percibida tuvo un buen recibimiento en muchas áreas del conocimiento, popularizándose rápidamente en el ámbito educativo. Actualmente, la Autoeficacia Percibida en educación es aplicada principalmente para explicar la motivación académica y el impacto que tienen las creencias propias en el logro académico de los estudiantes (Pajares, 1996).

En palabras de Bandura (2010), la Autoeficacia Percibida se refiere a las creencias de las personas sobre su capacidad para influir en los eventos que afectan sus vidas. Esta creencia central es la base de la motivación humana, los logros de desempeño y el bienestar emocional. Asimismo, es menester mencionar que, de acuerdo con la Teoría Social Cognitiva, la Autoeficacia Percibida proviene de cuatro fuentes de información principales: las experiencias de dominio, el modelado social, la persuasión social y la corrección en la interpretación de los estados somáticos como el estrés y la ansiedad.

Las experiencias de dominio según Bandura (2010), son clave para elevar la Autoeficacia Percibida. El tener éxito en una tarea solidifica las creencias de autoeficacia, mientras que las

experiencias de fracaso las disminuyen. Por otro lado, el modelado social se refiere a la influencia que ejerce el logro de otras personas en la conducta propia. Es decir, el observar a personas similares realizar una actividad con éxito por medio de la perseverancia eleva la confianza en las capacidades propias.

Las últimas dos fuentes de autoeficacia de acuerdo con dicho autor son la persuasión social y los estados somáticos. La persuasión social se refiere a la retroalimentación entre el aumento realista de la autoeficacia, el esfuerzo invertido para realizar una tarea y el éxito de ésta. Es decir, entre más esfuerzo invertido en una tarea, mayor posibilidad de éxito, lo cual a su vez eleva la Autoeficacia Percibida. Por último, una correcta interpretación de los estados somáticos es otra forma de elevar la autoeficacia. Estados como la fatiga, el estrés y la ansiedad se suelen interpretar como incapacidad para realizar una tarea con éxito (Bandura, 2010).

Continuando con los postulados de Bandura (2010) las creencias de autoeficacia modulan el funcionamiento humano de cuatro maneras principales: en los procesos cognitivos, motivacionales, emocionales y de selección. Los procesos cognitivos se refieren al pensamiento, a los escenarios que las personas imaginan y ensayan sobre situaciones específicas. La autoeficacia influye en estos escenarios, ya que las personas con una baja autoeficacia tenderán a imaginar escenarios de fracaso, mientras que, las personas con una elevada autoeficacia tenderán a recrear situaciones de éxito.

En los procesos motivacionales, la base es que la autoeficacia percibida regula la motivación cognitiva, debido a que las personas se fijan metas de acuerdo con la percepción de sus propias capacidades y de los resultados que creen que obtendrán al realizar una tarea. Desde luego estas percepciones también impactan en los objetivos que las personas se plantean, así como en el esfuerzo requerido para su cumplimiento (Bandura, 2010).

Los procesos afectivos o emocionales también se ven impactados por los niveles de Autoeficacia Percibida, ya que hay tres formas en que los procesos emocionales impactan a las creencias de autoeficacia: creando sesgos de atención, modulando la forma en la que se interpretan ciertos eventos,

afectando el control sobre pensamientos negativos o perturbadores y generando vías de acción para reinterpretar de manera más adaptativa los eventos considerados como amenazantes. En resumen, las personas con niveles de autoeficacia más elevados pueden controlar reacciones emocionales adversas como el estrés o la ansiedad, mientras que personas con un bajo nivel de autoeficacia se concentrarán y preocuparán de manera más intensa por las posibles amenazas del entorno (Bandura, 2010).

Por último, en los procesos de selección Bandura (2010) señala que las creencias de autoeficacia percibida van moldeando el curso de circunstancias que las personas seleccionarán enfrentar o evitar por completo. Es decir, la autoeficacia influye en las decisiones que toman los individuos en torno a qué ambientes y entornos afrontar y cuáles evadir. Dicho autor menciona también que la elección de una carrera profesional puede verse afectada por los procesos de selección ya que una persona con altos niveles de autoeficacia buscará tareas más complejas que impliquen retos, mientras que una persona con un nivel más bajo de autoeficacia evitará tareas difíciles ya que las interpretas como amenazas personales, atribuirá fracasos a su propia falta de habilidades e incluso desertará con mayor rapidez.

Desde luego, el estudio de estos componentes como posibles precursores para la ejecución de una conducta es un campo de investigación de gran interés para los científicos educativos y psicólogos. Desafortunadamente, los modelos de explicación a menudo presentan problemas para determinar de forma precisa de cómo la autoeficacia percibida se relaciona con otros elementos motivacionales y contextuales para convertirse en una intención sobre la cual el individuo pueda actuar (Schwarzer y Hallum, 2008).

Una vez explicados los componentes más relevantes de la teoría de la Autoeficacia Percibida, es menester ahondar en el estudio de este constructo en el contexto del aprendizaje de las ciencias. Hasta el momento en el que se redacta esta tesis se reconocen dos vías para estudiar este fenómeno en educación y son la *Science Self-efficacy* o Autoeficacia en la Ciencia por su traducción y la *Research Self-Efficacy* o Autoeficacia en la Investigación, se ahondará mayormente en el primero.

La autoeficacia científica y su impacto en el logro educativo ha sido objeto de estudio desde hace ya algunas décadas. En 1991, Schunk habría identificado una relación entre la autoeficacia y la motivación académica, aunque distinguiendo que la relación es dependiente de situaciones específicas. Es decir, aún si un estudiante demuestra autoeficacia para ejercicios propios de matemáticas, bien pudiera ser que su autoeficacia le motive a realizar tareas específicas relacionadas con la geometría y no con el cálculo, por ejemplo. Muestra de ello, es el trabajo de Britner y Pajares (2006) en donde se señala que la autoeficacia científica está asociada al rendimiento académico en ciencias y a la elección de carreras y actividades científicas en todos los niveles académicos.

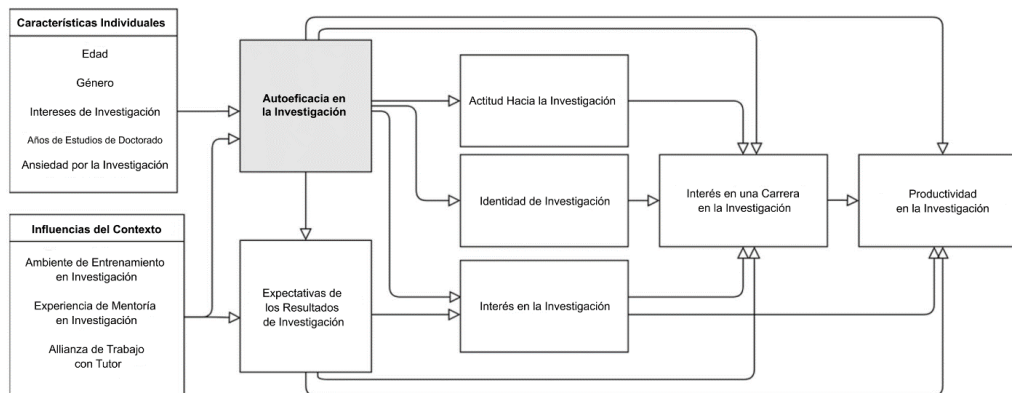
De forma similar, Jansen, Scherer y Schroeders (2015) destacan que la motivación para realizar actividades científicas en el futuro se relaciona simultáneamente con la autoeficacia en ciencia y el autoconcepto científico. Chen y Usher (2013) describen, además, que la idea de cuán flexibles son las habilidades para la ciencia (lo que los autores nombran una teoría implícita de la habilidad) se asocian con una mayor autoeficacia en ciencia, y a mayores estados de afecto positivo relacionados con la ciencia escolar.

Dentro del contexto de estudiar la autoeficacia científica, existe a su vez autoeficacia relacionada específicamente con habilidades de investigación. Phillips & Russell (1994) indican que la participación activa de un estudiante en proyectos de investigación es equivalente al logro directo en este dominio, el más importante factor para determinar las expectativas de autoeficacia. Similarmente, la existencia de modelos apropiados en el proceso de investigación se relacionaría con aprendizaje vicario. Las conclusiones de estos investigadores, además, son que una alta autoeficacia para la investigación se correlaciona positivamente con la productividad en investigación.

En el 2021, Livinți, Gunnesh-Luca y Illiescu emplearon la Teoría de Elección Carrera Social Cognitiva para tratar relacionar las variables específicamente asociadas con la Autoeficacia en Investigación, se observa en la **Figura 2:**

Figura 2.

VARIABLES RELACIONADAS A LA AUTOEFICACIA EN INVESTIGACIÓN DE ACUERDO CON LA TEORÍA SOCIAL COGNITIVA (TSC).



Nota: Recreado de Livinçi et al. (2021).

En el mismo, se puede identificar que tanto variables individuales como de contexto tienen influencia sobre la Autoeficacia de Investigación de una persona en particular. En cuanto a las variables individuales se puede encontrar la edad, el género, los intereses de investigación, los años de estudio en un programa de doctorado, y la ansiedad relacionada a la investigación. En cuanto a las variables de contexto, se puede encontrar que el ambiente de entrenamiento en investigación, la experiencia con mentores de investigación, y la alianza de trabajo con el director pueden impactar tanto la autoeficacia percibida de investigación, como las expectativas del resultado de investigación.

Ahora bien, debido a que la prueba PISA es uno de los referentes principales del diseño y desarrollo de la escala de autoeficacia percibida, es menester definir lo que para la OCDE (2022) es la *Competencia Científica*, ésta se refiere a:

“los conocimientos científicos y el uso que de esos conocimientos haga un individuo para identificar preguntas, adquirir nuevos conocimientos, explicar los fenómenos científicos y sacar conclusiones basadas en evidencias, sobre asuntos relacionados con la ciencia” (p.17).

La *Competencia Científica* tiene tres subcompetencias: identificar asuntos o temas científicos, explicar científicamente los fenómenos y usar evidencia científica. *Identificar asuntos o temas científicos*, se refiere a la capacidad de reconocer asuntos que pueden estudiarse de manera científica, así como el identificar las palabras adecuadas para buscar información apropiada y reconocer los elementos propios de una investigación científica. *Explicar científicamente los fenómenos* involucra el aplicar el conocimiento científico en situaciones diversas, así como el describir, interpretar y predecir de manera apropiada los fenómenos. Por último, el usar la evidencia científica se refiere a la interpretación de la evidencia, la identificación de hipótesis de investigación, extraer la lógica de las conclusiones de un estudio y comunicarla, así como ubicar las implicaciones sociales del desarrollo científico y tecnológico. En este apartado de reactivos, las subcompetencias de la *Competencia Científica* se evalúan a través del dominio del conocimiento y su aplicación de ciencias como la Física, la Química, las Ciencias biológicas, las Ciencias de la tierra y del espacio en contextos diversos como lo son salud, el medio ambiente y la tecnología (OCDE, 2022).

Ahora bien, la medición de la *Competencia Científica* y la medición de la autoeficacia en la prueba PISA se aborda de la siguiente manera: en el caso de la *Competencia Científica*, cada ítem tiene características particulares, por ejemplo: el ítem CS644Q01 que trata sobre los terremotos, se evalúa 1) la subcompetencia de *Interpretar datos y evidencia científica*, conocimiento *procedimental* en un contexto de *peligros o amenazas globales*, en una demanda cognitiva *baja* y con un formato de ítem de *opción múltiple*.

Por otro lado, en la prueba PISA (OCDE, 2022) se define la autoeficacia como el grado en que los individuos creen en su propia capacidad para participar en ciertas actividades y realizar tareas específicas (Bandura, 1977) por lo que se considera relevante solicitar a los estudiantes que evalúen qué tan capaces se sienten para realizar actividades relacionadas con las matemáticas o las ciencias frente a situaciones adversas.

Debido a que a lo largo del documento se utilizaron los mismos conceptos de manera continua, es menester unificar lo que a partir de este capítulo se entenderá por motivación, actitudes y autoeficacia en el aprendizaje de las ciencias. La motivación en el aprendizaje de las ciencias se define como un estado interno que despierta, dirige y sostiene la conducta de aprendizaje de la ciencia (Glynn et al., 2011), por otro lado, la actitud hacia la ciencia se entiende como los sentimientos, creencias y valores que tienen las personas sobre la realización de la ciencia, la ciencia escolar, el impacto de la ciencia en la sociedad, o los científicos en sí mismos (Klopfer, 1973; Fraser, 1982). Por último, la autoeficacia en ciencia se refiere a la creencia de un individuo en su capacidad para realizar con éxito tareas asociadas con la realización de investigaciones científicas (Forester et al., 2004 como se citó en Livinți et al. 2021).

2.2. Diseño, desarrollo y validación de instrumentos de medición

En este apartado se desarrollarán algunas de las propuestas de programas para el diseño, desarrollo y validación de instrumentos de medición, así como las directrices en común y las sugerencias que fueron tomadas en cuenta para la realización de las escalas. Posteriormente, se abordarán los fundamentos teóricos de la recopilación de evidencias de validez de contenido, el panel de expertos, el cálculo del Índice de Validez de Contenido (IVC), así como la imparcialidad de los reactivos de acuerdo con el modelo de Thompson, Johnstone y Thurlow (2002).

2.2.1. Programas para el diseño, desarrollo y validación de instrumentos de medición

Para el desarrollo de instrumentos de medición, existen directrices o pasos sugeridos por autores o instituciones reconocidas en el campo de la psicometría. En la **Tabla 1** se muestran los pasos sugeridos

por los psicómetras Haladyna y Rodríguez (2013), por los *Estándares* de la AERA, APA y NCME (2018), por Nitko (1994) y Contreras (2009).

Tabla 1
Etapas para el desarrollo de instrumentos de medición

Haladyna y Rodríguez (2013)	Estándares de la AERA, APA y NCME (2018)	Nitko (1994)	Contreras (2009)
(a) Introducción y descripción general	Declaración original de la(s) finalidad(es) hasta el producto final:	(a) Definir el dominio de resultados que pretende el currículum	(a) Definición de resultados que pretende el currículum
(b) Definición de contenido o constructo	(a) desarrollo y evaluación de las especificaciones de la prueba;	(b) Analizar el currículum	(b) Análisis del currículum
(c) Estructura de la prueba (ítem y especificaciones de la prueba)	(b) desarrollo, ensayo y evaluación de los ítems;	(c) Desarrollar un plan de evaluación	(c) Desarrollar del plan de evaluación
(d) Formatos	(c) reunión y evaluación de nuevos formularios de la prueba;	(d) Desarrollar las especificación	(d) Producción y validación de ítems
(e) Directrices para escribir ítems	y	(e) Producir y validar ítems	(e) Análisis primario de resultados
(f) Ejemplos de ítems bien y mal escritos en discusión grupal	(d) desarrollo de procedimientos y materiales para administración y calificación.	(f) Ensamblar los exámenes	(f) Análisis secundario de resultados
(g) Formularios para usar al escribir ítems		(g) Establecer estándares (h) Análisis primario	
(h) Sesión de redacción de artículos		(i) Análisis secundario	
(i) Discusión grupal posterior a la sesión sobre los resultados			
(j) Edición y revisión de ítems			
(k) Características de los ítems validados			
(l) Tareas y horario			

Nota. Los modelos de Nitko (1994) y Contreras (2009) son utilizado en evaluación de los aprendizajes.

Uno de los principales referentes en el diseño, desarrollo y validación de instrumentos de medición son los Estándares para el desarrollo de pruebas psicológicas y educativas de la AERA, APA y NCME (2018). En los *Estándares* se encuentran las directrices y los pasos sugeridos para el diseño y desarrollo de pruebas, estos son: la Declaración original de la(s) finalidad(es) hasta el producto final, en este punto se señala que:

“El proceso de desarrollar pruebas educativas y psicológicas debe comenzar con una declaración de la(s) finalidad(es) de la prueba, los usuarios y usos previstos, el constructo o dominio de contenido

sometido a medición, y la población prevista de individuos examinados. Las pruebas del mismo constructo o dominio pueden diferir de maneras importantes porque factores tales como finalidad, usos previstos y población de individuos examinados pueden variar” (AERA, APA y NCME, 2018, pp. 86).

Posteriormente, en el (a) desarrollo y evaluación de las especificaciones de la prueba se deberá atender a la pregunta *¿Qué debe incluirse en la prueba? ¿Cuál es la delimitación de contenidos de la prueba?* Una vez acordado el contenido de la prueba, le seguirán las especificaciones del formato de la prueba, que contienen “el formato de ítems (es decir, tareas o preguntas); el formato de respuesta o condiciones para responder; y el tipo de procedimientos de calificación” (pp. 87). Posteriormente, (b) el desarrollo, ensayo y evaluación de los ítems, en donde el desarrollador o desarrolladores reúnen las preguntas o tareas esenciales para la construcción del instrumento, así como los ensayos necesarios o evaluaciones previas para detectar características irrelevantes o posibles sesgos en los ítems (AERA, APA y NCME, 2018).

Una vez desarrollado el instrumento, la (c) reunión y evaluación de nuevos formularios de la prueba involucra la adaptación de la prueba a diferentes formatos, siguiendo la directriz de las especificaciones de contenido obviando los ítems que pudieran ser repetitivos, así como la aplicación de la prueba en campo para detectar si el comportamiento de los ítems es el esperado en la población meta de la evaluación. Por último y en síntesis, en el (d) desarrollo de procedimientos y materiales para administración y calificación, sostiene que

“la administración de la prueba debe ser imparcial para todos los individuos examinados. Esto significa que las instrucciones para dar la prueba deben ser claras y que las condiciones de administración de la prueba deben ser estandarizadas para todos los individuos examinados” (pp.95).

Así también que, en el proceso de revisión de las pruebas se vuelve necesario determinar si al paso del tiempo una prueba ha dejado de ser óptima para su uso. Por este motivo es necesario someter a

revisión tanto el dominio, las condiciones de uso, las interpretaciones, el contenido o el lenguaje. De ser necesario, se deberán realizar cambios en la prueba ya que puede verse afectada la validez de las interpretaciones propuestas, de igual manera si la población meta de la evaluación ha sufrido cambios a lo largo del tiempo (AERA, APA y NCME, 2018).

Por otro lado, en el trabajo de Haladyna y Rodríguez (2013), “*Developing and validating items*” se propone la redacción de una *guía de redacción de ítems*, el cual es un documento oficial utilizado en la capacitación de los desarrolladores de ítems para auxiliarlos en el proceso de redacción de los reactivos. Esta guía contiene las especificaciones de todos los ítems de la prueba, el formato que deberá usarse, así como pautas para la redacción de los ítems.

Los pasos sugeridos por Haladyna y Rodríguez (2013) para la elaboración de ítems son: la (a) introducción y descripción general, (b) la definición de contenido o constructo, (c) la delimitación de la estructura de la prueba (los ítems y especificaciones de la prueba), (d) el formato de los instrumentos de medición, (e) el desarrollo de las directrices o especificaciones para escribir ítems, (f) la discusión de ejemplos de ítems bien escritos y mal escritos con el grupo de desarrolladores, (g) los formularios para escribir ítems, (h) la sesión de redacción de ítems, (i) la discusión grupal posterior a la sesión sobre los resultados, actividad que dichos autores consideran sumamente relevante ya que escuchar a los desarrolladores hablar de sus ítems y las pautas para redactarlos pueden ser consejos valiosos para aprender a desarrollar mejores reactivos. La (j) edición y revisión de ítems, (k) las características de los ítems validados y por último, (l) las actividades y el horario que se dedicará a las sesiones de redacción de ítems.

Otro modelo para el diseño, desarrollo y validación de instrumentos es la propuesta de Nitko (1994), para el desarrollo de pruebas para la evaluación de los aprendizajes. Dado que el propósito de la medición en este modelo es la evaluación propiamente de los aprendizajes, los primeros pasos involucran la revisión del currículum del cual se desprenderán los contenidos a evaluar. Es por ello que, los primeros

pasos propuestos por dicho autor son: (a) definir el dominio de resultados que pretende el currículum y (b) analizar el currículum. Posteriormente, se (c) desarrollará un plan de evaluación en donde se realiza un muestreo de los aprendizajes a evaluar y el diseño de prototipos de especificaciones, así como otra actividad de suma relevancia: la creación de comités de examen. Una vez conformado este comité, se (d) desarrollarán las especificaciones, así como la (e) producción y validación de ítems, el (f) ensamblaje los exámenes y (g) el establecimiento de estándares de ejecución. Por último, el (h) análisis primario y (i) el análisis secundario. El primario involucra la especificación de estándares de calidad del ítem, el diseño de procedimientos para calificar la prueba y el análisis de resultados. El secundario (i) se refiere a la elaboración del reporte de resultados (Pérez-Morán, 2010).

En el mismo tenor de evaluación de los aprendizajes, se contempló el modelo de Contreras (2009), en donde el primer paso es la (a) definición de resultados que pretende el currículum y el (b) análisis del currículum propiamente. Una vez realizado este análisis, le sigue la etapa de (c) desarrollo del plan de evaluación, así como (d) la producción y validación de los ítems. El modelo de dicho autor, contempla también un (e) análisis primario de resultados y un (f) análisis secundario de resultados. El primario involucra la lectura de hojas de respuesta, la base de datos y su análisis, la interpretación de resultados del examen, la calificación del examen y el análisis final de ítems. El análisis secundario se refiere a los análisis por área curricular, por variable de contexto, la elaboración del reporte y la difusión de la información (Pérez-Morán, 2010).

Dichos modelos, tienen en común una lógica que sigue una serie de pasos de acuerdo con las necesidades y propósitos de la medición. Los pasos que tienen en común los modelos anteriormente expuestos son:

- (a) Establecer al inicio cual es el propósito de la medición
- (b) Revisión de referentes, el currículum, el dominio o la delimitación del contenido de la prueba

- (c) Desarrollo de las especificaciones
- (d) Desarrollo de los ítems
- (e) Proceso de validación de la prueba

La adaptación de estos modelos para el diseño, desarrollo y validación de las escalas para medir la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias para estudiantes de Educación Superior nivel licenciatura se adaptó del modelo de Haladyna y Rodríguez (2013) y los Estándares de la AERA, APA y NCME (2018). El motivo de esta decisión radicó en que el diseño y desarrollo de las escalas del presente estudio tienen propósito de mejora de los aprendizajes y son del dominio afectivo del mismo. El modelo propuesto para el diseño, desarrollo y validación de las escalas se encuentra en el apartado de procedimiento.

2.2.2. Evidencias de validez de constructo en el aspecto del contenido de los instrumentos de medición

La validez en los instrumentos de medición se refiere “al grado en que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba para los usos propuestos de las pruebas” (AERA, APA y NCME, 2018, p. 11). Es decir, el proceso de validez se refiere a la recuperación de evidencias con base científica para la interpretación de los puntajes en los instrumentos de medición. Es importante señalar que cuando una prueba tiene más de una interpretación, es necesario validar cada una de las interpretaciones, por este motivo es incorrecto decir que una prueba es *válida* (AERA, APA y NCME, 2018).

Asimismo, se utilizó la definición de Messik (1995) que define la validez como el grado en que la evidencia empírica y los fundamentos teóricos apoyan la idoneidad y pertinencia de las interpretaciones y acciones sobre la base de los puntajes de las pruebas u otros modos de evaluación, es

decir, las consecuencias sociales derivadas de las puntuaciones de las pruebas. Por este motivo, el concepto de validez de Messik (1995) se conoce como validez unitaria o validez de constructo.

De acuerdo con los *Estándares*, el proceso de validación de un instrumento de medición debe de comenzar con un enunciado en donde se explicita la interpretación propuesta para el instrumento, en conjunto con la relevancia de su uso y las interpretaciones derivadas. La interpretación de la prueba que se propone deberá contar con una definición del constructo que se pretende medir, entendiendo el constructo como el “concepto o característica para cuya medición se diseña una prueba” (AERA, APA y NCME, 2018, p. 11).

Siguiendo con los *Estándares*, “la validez es un concepto unitario” (p.14). Sin embargo, existen fuentes de evidencias de validez según el uso y la interpretación propuesta para un instrumento de medición. En este punto, cabe aclarar que las fuentes de validez representan diferentes aspectos de ésta, y no diferentes tipos de validez. Por este motivo, es importante diferenciar que, aunque históricamente en el concepto de validez se subdivide en tipos de validez, en los *Estándares* el tratamiento del concepto es distinto: hay diferentes tipos de evidencias de validez, no tipos de validez. Dicho esto, las fuentes de evidencias de validez son: la evidencia basada en el contenido de la prueba, la evidencia basada en los procesos de respuesta, la evidencia basada en la estructura interna, la evidencia basada en la relación con otras variables y la evidencia de validación y consecuencias de las pruebas (AERA, APA y NCME, 2018). Para los propósitos de esta tesis se definirá la evidencia basada en el contenido de la prueba.

Esta se refiere a “la relación entre el contenido de una prueba y el constructo que se intenta medir” (AERA, APA y NCME, 2018, p. 14). Por lo tanto, el contenido de una prueba es todo lo que la compone: los temas, ítems, preguntas, la redacción y el formato. Al momento de construir un instrumento de medición, los encargados de esta tarea trabajan con las especificaciones que no es más que el dominio de contenido. Es decir, en una especificación se detallan los temas que forman parte del contenido de una

prueba, así como los tipos de ítems que se desarrollarán a partir del contenido (AERA, APA y NCME, 2018).

Aunque el concepto de validez de contenido ha sido cambiante a lo largo de los años, se considera fundamental para la recolección de evidencias de validez de constructo y, por tanto, para los instrumentos de medición (Sireci, 1998). Para Messik (1995), el contenido de una prueba se refiere a los límites del constructo que se va a medir, es decir, los conocimientos, habilidades, o actitudes que son representativas del dominio del constructo.

La evidencia basada en el contenido puede derivar de análisis lógicos o empíricos de que el contenido de la prueba representa el dominio y la relevancia del constructo que se pretende medir. Asimismo, también puede provenir del juicio de expertos que confirmen la relación entre el dominio del constructo y el contenido de la prueba (AERA, APA y NCME, 2018). Por ejemplo, en el trabajo de Escobar- Pérez y Cuervo- Martínez (2008) se proponen una serie de pasos para la conformación de un panel de jueces expertos de forma sistemática. En primer lugar, es indispensable definir el objetivo del juicio de expertos, es decir, esclarecer la función específica del jueceo. Esta función puede variar, ya que el grupo de expertos puede evaluar la adaptación de una prueba a otro idioma, la adaptación cultural de un constructo, o una prueba realizada por los investigadores.

Posteriormente, se realiza una actividad de suma relevancia: la selección de jueces. Ésta se llevará a cabo respetando los criterios definidos con anterioridad. La elección de expertos toma en cuenta la experiencia, la formación académica y el reconocimiento social. También, como otro criterio de elección el comité deberá tener por lo menos un miembro experto en medición y evaluación, y un experto en lingüística en caso de que la prueba haya sido traducida del idioma original.

Una vez que el panel de expertos está conformado, se deberá explicitar las dimensiones e indicadores que cada ítem de la prueba está midiendo, así como especificar el objetivo de la prueba a los jueces, es decir, para qué será utilizada y qué significan las puntuaciones de la prueba, esto para dar

contexto específico a los expertos sobre el uso e interpretación de los puntajes de la prueba. Otro aspecto que considerar, según Escobar-Pérez y Cuervo-Martínez (2008) es establecer los pesos diferenciales de las dimensiones de la prueba. Esto aplica cuando algún constructo tiene más peso en la prueba, por ejemplo, para la realización de un diagnóstico específico. Se propone también el diseño de planillas conforme a los objetivos de evaluación, así como el cálculo de acuerdo entre jueces. Por último, los autores señalan la elaboración de conclusiones del juicio, mismas que serán desarrolladas en la descripción de la prueba.

A lo largo de las décadas, se han desarrollado índices para calcular el acuerdo entre el panel de jueces. Algunos de ellos, sintetizados en el trabajo de Cardonda et al. (2016) son el Acuerdo de proporción, el Índice de Kappa y el índice de Aiken. El acuerdo de proporción es básicamente un porcentaje, extraído de la evaluación individual que cada juez hace con respecto al dominio del contenido de una prueba (Wynd et al. como se citó en Cardonda et al, 2016). La principal crítica que se le realiza al acuerdo de proporción es que el porcentaje de acuerdo puede ser inflado por factores no considerados (Polit y Beck, 2006 y Cardonda et al., 2016).

Otra de las formas más utilizadas para computar el porcentaje de acuerdo entre jueces, es el cálculo de la “V” Aiken (1985). En el trabajo de García Sedeña y García Tejera (2014) el coeficiente o índice “V” de Aiken, se define como “una de las técnicas que permite cuantificar la relevancia de cada ítem respecto de un dominio de contenido formulado por N jueces” (pág. 50).

La ecuación para calcular el coeficiente de Aiken es la siguiente:

$$V = \frac{\bar{X} - l}{K}$$

En donde V es el coeficiente de Aiken, X es el promedio de calificación de los jueces para un aspecto concreto de cada ítem, l es la calificación más baja con la que un juez puede evaluar un ítem y K es el rango de la escala.

Otro método relevante para el cálculo del acuerdo entre los jueces es la Razón de Validez de Contenido (RVC) propuesta por Lawshe (1975). Este método consiste en la organización de un Panel de Evaluación de Contenido, el cual está integrado por un panel de jueces expertos en el tema. Estos jueces reciben un ejemplar de la prueba y evalúan el contenido de los ítems de acuerdo con tres categorías: esencial, útil pero no esencial y no necesario. Se espera, por tanto, un alto acuerdo entre los jueces en la relevancia del contenido de los ítems para medir el constructo esperado (Tristán-López, 2008). La fórmula para calcular el RVC de describe a continuación:

$$RVC = \frac{n_e - N/2}{N/2}$$

En donde n es el número de jueces que calificaron el ítem como esencial, y N el número total de jueces que evaluaron el contenido de la prueba. Posteriormente, se calcula el promedio de todos los ítems, lo que resultará en el IVC general de la prueba.

Una de las críticas al Índice de Lawshe es la cantidad de jueces que se requieren en el proceso de evaluación de la prueba, aspecto que lo hace poco práctico de usar. Estas observaciones y la respectiva corrección se abordan en el trabajo de Tristán-López (2008). Por lo tanto, Tristán-López propone dos versiones del RVC de Lawshe, una ecuación corregida para el cálculo del RVC por ítem, y una ecuación para el cálculo del RVC total de la prueba. La primera ecuación se describe a continuación:

$$RVC = \frac{RVC + 1}{2}$$

En donde RVC es la razón de validez de cada ítem, y la respectiva corrección propuesta por Tristán-López (2008). Además, se propone una nueva ecuación para el cálculo de RVC total de la prueba. Dicha ecuación es la siguiente:

$$RVI = \frac{\sum_{i=1}^M RVC_i}{M}$$

Estos resultados numéricos, también están sujetos a interpretación.

Pedrosa et al. (2014) señalan que la forma más adecuada de estimar la validez de contenido de un instrumento de medición es la combinación de métodos, tanto cualitativos como cuantitativos. Es decir, el procedimiento más robusto para la recopilación de evidencias de validez de constructo en el aspecto del contenido sería contar con las valoraciones de un panel de expertos que emiten su juicio con respecto a los ítems y, posteriormente, la aplicación de métodos estadísticos para calcular el IVC, en donde el RVC de Lawshe (1975) se muestra como uno de los más completos entre todas las opciones para estimar el cálculo de la validez de contenido.

2.2.3. Evidencias de validez e imparcialidad de las pruebas

De acuerdo con los Estándares para el desarrollo de pruebas psicológicas y educativas de la AERA, APA y NCME (2014) “la imparcialidad es una cuestión de validez fundamental y requiere atención en todas las etapas del desarrollo y uso de las pruebas” (p. 53). El capítulo de imparcialidad en los Estándares aborda el sesgo de medición como una amenaza a la validez de las pruebas, mientras que introduce dos conceptos para potencializar la imparcialidad de los instrumentos de medición: accesibilidad y diseño universal (AERA, APA y NCME, 2018).

La accesibilidad se refiere a “la noción de que todos los examinandos deben tener la oportunidad sin obstáculos de demostrar su situación respecto de los constructos sometidos a medición” (AERA, APA y NCME, 2018, p. 54). Por ejemplo, el tamaño de la letra y algunos formatos electrónicos que pudieran afectar la evaluación de personas con debilidad visual, pero que nada tienen que ver con el constructo que se pretende medir.

Por otro lado, el *diseño universal*, concepto que se agregó a los *Estándares* en 2014, pretende que los ítems y tareas de una prueba se diseñen y desarrollen desde el inicio con la finalidad de minimizar aspectos irrelevantes del constructo y maximizar la accesibilidad de la prueba.

Asimismo, el *diseño universal* de las pruebas:

“requiere que los desarrolladores de la prueba sean claros sobre los constructos sometidos a medición, incluyendo el objetivo de la evaluación, el fin para el que se usarán los puntajes, las inferencias que se harán a partir de los puntajes, y las características de los individuos examinados y los subgrupos de la población prevista de la prueba que podrían influir en el acceso” (AERA, APA, NCME, 2018, p. 54).

Para atender el Estándar de imparcialidad, se utilizó el Modelo de Diseño Universal para el diseño de pruebas desarrollado por Thompson, Johnstone y Thurlow (2002) con traducción de Pérez-Morán (2014). La Evaluación de Diseño Universal (EDU) por su traducción, tiene siete principios que guían el desarrollo de pruebas, se describen a continuación.

El primer principio de diseño de pruebas requiere de un uso equitativo, es decir, cuando se diseña una prueba debe pensarse en el contexto de la población que será evaluada. Aunque en algunas situaciones es necesario acotar la población meta de la prueba, para otro tipo de evaluaciones como la toma de decisiones en materia educativa, es ideal que la prueba sea accesible a todo tipo de estudiantes. Esto no quiere decir que el constructo a medir cambie o se flexibilice, si no que la prueba se desarrolla en distintos diseños, tecnologías y formatos para alcanzar a una mayor cantidad de usuarios a fin de evitar que algunos elementos de la prueba sean engañosos o poco claros para ciertos grupos de estudiantes.

El segundo principio señala la necesidad de contar con una definición precisa de los constructos que se van a medir. Además, es importante que estas definiciones sean accesibles para los actores involucrados en la evaluación, sobre todo si se tomarán decisiones importantes a raíz de dicha medición.

El tercer principio habla sobre la *accesibilidad y sesgo de los ítems*, indicando que éstos tengan calidad en cuanto al contenido, claridad y falta de ambigüedad.

El cuarto principio *Acomodación flexible de los contenidos* señala que las pruebas deben facilitar las adaptaciones posteriores. Es decir, no deben contener gráficos o imágenes irrelevantes, textos diagonales o verticales, ítems que dependan de imágenes para resolverse o imágenes o texto irrelevante para el constructo que se desea medir.

El quinto principio *Procedimientos e instrucciones simples, claras e intuitivas* indica que las instrucciones y procedimientos deben ser fáciles de entender para los participantes independientemente de su dominio del idioma, nivel de desarrollo o nivel de concentración en el que se encuentren.

El sexto principio, de la *Comprensibilidad*, señala que en los ítems se deberá cuidar que las oraciones no sean muy largas o confusas y que las palabras utilizadas sean familiares para quien las lee. Es decir, la respuesta no deberá ser influenciada por la falta de comprensión de las instrucciones del instrumento.

Por último, el principio de *Máxima legibilidad* señala la importancia de que los elementos de la prueba sean fácilmente descifrables (legibles) para los sustentantes. Este elemento es aplicable también para figuras y tablas y formatos en el que se insertan los ítems.

Asimismo, en el apartado de amenazas a las interpretaciones imparciales y válidas de los puntajes de las pruebas se señala que una posible fuente de varianza irrelevante del constructo, puede proceder del propio contenido de la prueba. Esto sucede cuando el contenido de una prueba confunde el objetivo de la medición, al mismo tiempo que puede llegar a favorecer a alguna población en especial (AERA, APA, NCME, 2018). En los *Estándares* se recomienda evitar utilizar palabras o expresiones asociadas con disciplinas u ocupaciones específicas, así como niveles socioeconómicos, grupos raciales o lugares geográficos. Esto con la finalidad de medir apropiadamente el constructo, y no experiencias o conocimientos previos de los examinados (AERA, APA, NCME, 2018).

Así también, el compromiso y el valor motivacional pueden ser factores irrelevantes del constructo. Es decir, que una tarea sea más interesante para un subgrupo de la población que para otro. En este caso, los ítems de las pruebas deberán ser atractivos para la mayor cantidad de examinados de la población destino. Asimismo, se debe evitar que en el contenido de la prueba haya palabras o situaciones que pudieran ser ofensivos para los examinados, más si esto pudiera impedir la capacidad de la persona para responder la prueba. En ambos ejemplos puede no ser aplicable si el interés o la situación ofensiva o perturbadora cae dentro del constructo que se quiere medir (AERA, APA y NCME, 2018).

III. MÉTODO

En este capítulo se describe el método que se utilizó para la realización del proyecto de investigación. Se divide en cuatro secciones: en la primera se describe el tipo de estudio, en la segunda sección se describen las características de los participantes, miembros del panel de jueces. En el apartado tres se describen los instrumentos y materiales utilizados en el estudio. Finalmente, en la última sección de este capítulo, se presenta el procedimiento de la investigación, que consiste en la mención y descripción de todas las actividades realizadas para el logro de los objetivos del estudio.

3.1. Tipo de estudio

La presente tesis consiste en un estudio cuantitativo-psicométrico de diseño, desarrollo y validación de tres escalas para medir la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias desde un enfoque integrador en estudiantes de ES a nivel licenciatura. El propósito de dichos instrumentos es ofrecer a estudiantes, docentes, orientadores vocacionales y autoridades educativas de educación superior a nivel licenciatura información relevante para la toma de decisiones informada que impulse el diseño de políticas educativas, planes de estudio, intervenciones educativas y estrategias pedagógicas que favorezcan la mejora educativa y el desarrollo y fomento de la vocación científica. Para ello, se aplica un modelo de diseño, desarrollo y validación de instrumentos de medición con base en las recomendaciones de Haladyna y Rodríguez (2013), desde un enfoque unitario de la validez de constructo (Messick, 1995; Kane, 2006; Embretson y Gorin, 2001), y atendiendo los Estándares para el desarrollo de pruebas educativas y psicológicas de la AERA, APA y NCME (2018).

3.2. Participantes del estudio

Los participantes se organizaron en dos etapas relacionadas con (a) el diseño y desarrollo de las escalas, y (b) la validación por juicio de expertos. En el diseño y desarrollo de las tres escalas trabajaron dos investigadores con experiencia en el campo de la medición educativa. Por su parte, para la validación de las escalas se integró el Comité de Validación de las Escalas (CVE) el cual fue conformado por 12 jueces con experiencia en el diseño, desarrollo y validación de instrumentos de medición; en evaluación educativa; y en docencia de metodología de la investigación en nivel superior. Todos los jueces del CVE se encuentran adscritos a distintas instituciones de reconocido prestigio en el campo de la evaluación educativa como lo son: la Universidad Autónoma de Baja California (UABC), la Universidad Autónoma Nacional de México (UNAM), la Universidad Veracruzana (UV), el Centro de Enseñanza Técnica y Superior (CETYS) Campus Tijuana, la Comisión Nacional para la Mejora Continua de la Educación (MEJOREDU), la Secretaría de Educación Pública (SEP) y Métrica Educativa A.C. Asimismo, cuentan con una amplia experiencia en investigación y evaluación educativa, y en especial, en el diseño, desarrollo y validación de instrumentos de medición. Por último, nueve jueces cuentan con estudios de doctorado, dos son doctorantes y uno tiene estudios de maestría en Ciencias Educativas. Diez de los 12 jueces cuentan con publicaciones en el campo de la medición educativa en revistas indexadas.

3.3. Instrumentos y materiales

Los instrumentos y materiales que se utilizaron en el estudio fueron: (a) el manual de procedimientos para la recopilación de evidencias de validez, (b) el video tutorial para el uso de los documentos, (c) la ficha técnica de las tres escalas, (d) el sistema de categorías EDU y (e) el Protocolo de validación de contenido de las escalas. Los primeros dos elementos se utilizaron para la capacitación del CVE, la ficha

técnica y el sistema de categorías de la EDU fueron insumos para dicha capacitación. Por último, el protocolo de validación de contenido de las escalas se utilizó como herramienta analítica.

En la **Tabla 2** se muestra un resumen del Manual de procedimientos, que consiste en un documento que consta de cuatro páginas con las instrucciones precisas para que los jueces puedan realizar el procedimiento de recopilación de evidencias de validez de constructo en el aspecto del contenido y de la imparcialidad, de acuerdo con los criterios de la EDU. Por otro lado, el video tutorial es un video de corta duración de elaboración propia en donde se hace un seguimiento puntual en pantalla de todo el procedimiento de recopilación de evidencias de validez para resolver cualquier duda que pudiera surgir por parte de los jueces. Asimismo, la ficha técnica es el documento rector de las escalas, en este documento se concentran los fundamentos teóricos y metodológicos de la medición de los constructos, los nombres de los instrumentos, los usos y propósitos de la medición, así como la estructura y formato de las escalas. En cuanto al sistema de categorías de la EDU, se describieron con más profundidad en el apartado de fundamentos.

Por último, el Protocolo de validación de contenido es un instrumento con tipo de respuesta en escalamiento, que tiene como objetivo conocer la opinión de los jueces con respecto a cada ítem y la escala en general en cuanto a cuatro criterios: relevancia, claridad, simplicidad y precisión. El código de respuesta es 1= nada relevante, 2=poco relevante, 3=algo relevante y 4=muy relevante. Para su uso se adaptó a un formulario de Google Forms.

Tabla 2.

Tabla resumen del Manual de procedimientos de la recopilación de evidencias de validez de las escalas.

Pasos	Contenido
Paso 1	Descripción de los criterios para el cálculo del IVC y de las categorías de la EDU.
Paso 2	Explicación de la carpeta de Google Drive compartida y de la Ficha técnica de las escalas.
Paso 3	Instrucciones para retroalimentar las escalas con las categorías de la EDU.
Paso 4	Instrucciones para responder el <i>Protocolo de validación de contenido</i>
Paso 5	Instrucciones para finalizar el procedimiento de recopilación de evidencias de validez.
Referencias	Artículos y documentos referenciados a lo largo del documento.

3.4. Procedimiento

Para llevar a cabo el presente proyecto de investigación se desarrolló un programa para el diseño, desarrollo y validación de las escalas con base en las recomendaciones de Haladyna y Rodríguez (2013) asociadas con: (1) introducir y describir de forma general el instrumento y sus propósitos, (2) definir del contenido/constructo, (3) la estructura de la prueba (ítem y contenidos de la prueba), (4) los formatos, (5) las directrices para escribir ítems, (6) análisis de ejemplos de ítems bien y mal escritos en discusión en grupo, (8) sesión de redacción de ítems, (9) discusión grupal posterior a la sesión, (10) edición y revisión de elementos y (11) características de los artículos validados. Por cuestiones de tiempo y duración del posgrado, se omitieron los pasos 7 y 12 de la propuesta de Haladyna y Rodríguez (2013) referentes a la agenda y al uso de procesadores de texto para la redacción de los ítems.

Asimismo, se atendieron los Estándares para el desarrollo de pruebas educativas y psicológicas de la AERA, APA y NCME (2018) relacionados con: el diseño de instrumentos (4.0 a 4.14); la obtención de evidencias de validez de constructo (1.9, 1.10), en particular del aspecto del contenido (1.11); y el análisis de la imparcialidad de los instrumentos de medición (3.0 a 3.5 y 3.9).

Para la atención al estándar 3.4 referente a la imparcialidad de los instrumentos de medición se utilizaron las categorías alineadas a la Evaluación de Diseño Universal (EDU) propuestas por Thompson et al. (2002) adaptadas al español por Pérez-Morán (2014) para el estudio del sesgo presente en los instrumentos de medición. Por otro lado, para la recopilación de evidencias de validez en el aspecto del contenido los jueces utilizaron un protocolo de validación para la evaluación de los criterios de relevancia, simplicidad, claridad y precisión de los ítems de las escalas. Posteriormente, para hacer el cálculo del acuerdo entre los jueces se utilizó el Porcentaje de Acuerdo entre Jueces (PAJ), la V de Aiken (1980), la Razón de Validez de Contenido de Lawshe (1975) y la corrección de Tristán-López (2008). En la **Tabla 3** se describe de manera gráfica el procedimiento.

Tabla 3.

Modelo de diseño, desarrollo y validación de las escalas para explorar la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias en estudiantes de Educación Superior. Adaptado de Haladyna y Rodríguez (2013).

Etapas	Actividades
1. Diseño de las escalas	1.1. Constituir el Equipo de Diseño y Desarrollo de las Escalas (EDDE). 1.2. Revisar los antecedentes y fundamentos teóricos de modelos de medición consolidados de los constructos de interés. 1.3. Delimitar estructura, subescalas, cantidad y formato de los ítems y modalidad de aplicación de las escalas. 1.4. Elaborar la Ficha Técnica de las escalas.
2. Desarrollo de las escalas	2.1. Desarrollar las tablas de especificaciones de los ítems de las escalas. 2.2. Elaborar la primera versión de los ítems de las escalas. 2.3. Ensamblar las escalas y armar la versión para su aplicación en papel y lápiz.
3. Recopilación de evidencias de validez del aspecto del contenido y de la imparcialidad de las escalas	3.1. Elaborar el material para la capacitación del Comité de Validación de las Escalas (CVE). 3.2. Integrar y capacitar al CVE para el proceso de validación de las escalas. 3.3. Recopilar evidencias de validez en el aspecto de la imparcialidad de las escalas. 3.4. Recopilar evidencias de validez en el aspecto del contenido de las escalas con jueces expertos. 3.5. Analizar las evidencias de validez del aspecto de la imparcialidad y del contenido de las escalas. 3.6. Ajustar los ítems con base en los resultados del análisis de las evidencias de validez. 3.7. Ensamblar la segunda versión de las escalas para su piloteo en línea.

El presente estudio se desarrolló en tres etapas alineadas al modelo de diseño, desarrollo y validación de los instrumentos de medición. Cada etapa con sus respectivas actividades específicas, condujeron al

cumplimiento de los objetivos de la investigación. En la Etapa I se llevó a cabo la planeación general y el diseño de las escalas. En la Etapa II se desarrollaron las tablas de especificaciones y los ítems de las escalas. Por último, en la Etapa III se recopilaron evidencias de validez de constructo en el aspecto de la imparcialidad y del contenido de las escalas. A continuación, se describen con más detalle las etapas del estudio.

Etapa I. Diseño de escalas para medir de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias en estudiantes de educación superior

Primeramente, se realizó una revisión de la literatura académica sobre la formación científica de los estudiantes atendiendo las sugerencias los Estándares de la AERA, APA y NCME (2018) así como la propuesta para el diseño de instrumentos de Haladyna y Rodríguez (2013) sobre la definición precisa del constructo. La revisión de la literatura se realizó con la finalidad de encontrar las palabras claves más utilizadas en el campo de la educación científica, así como los constructos más apropiados para la realización de este trabajo de investigación.

Se delimitó la búsqueda de literatura científica a la medición o evaluación de los constructos mencionados anteriormente obviando otro tipo de literatura académica como la ensayística o reflexiva. Se agregó a cada constructo operadores booleanos y palabras clave como medición, psicométricos, evaluación y validez con la finalidad de encontrar estudios de diseño, desarrollo y validación de instrumentos de medición. Las bases de datos y revistas científicas que se consideraron para la revisión de la literatura fueron: Scopus, Sage, ERIC, Dialnet, Latindex y Scielo. Se buscó también en revistas indexadas sobre el tema como: International Journal of Science Education, Open Review of Educational Research, Research in Science & Technological Education, Evaluation & Research in Education, Educational Assessment, Applied Measurement in Education y Journal of Career Assessment. En idioma español se consideraron las revistas: Revista Enseñanza de las ciencias, Revista Eureka sobre Enseñanza

y Divulgación de las Ciencias, Psicología Educativa, Sinéctica, Perfiles educativos, Revista de investigación educativa y Psychologia. Avances de la disciplina. Aunque se encontraron y consideraron artículos clásicos, el periodo de búsqueda de la literatura se concentró en la evolución del tema en los últimos 10 años, es decir, del periodo del año 2010 al 2020.

Finalmente, posterior a la revisión de la literatura, se acordó delimitar la medición a los componentes afectivos del aprendizaje de las ciencias como las actitudes hacia la ciencia, la motivación para el aprendizaje de las ciencias y la autoeficacia en ciencia e investigación, dada la creciente atención a los componentes afectivos del aprendizaje por parte de la comunidad de investigadores educativos (Koballa y Glynn, 2007). Se acordaron los temas, teorías y documentos centrales que compusieron el apartado de fundamentos en dos ejes principales: las teorías sustantivas de los constructos de interés y, por el otro lado, los fundamentos teóricos de la medición psicológica y la validez de los instrumentos de medición.

De acuerdo con el modelo de diseño, desarrollo y validación del instrumento de medición propuesto para la presente tesis, en la actividad 1.1 se constituyó el Equipo de Diseño y Desarrollo de las Escalas (EDDE), conformado por dos investigadores con experiencia en el campo de la medición educativa. Una de las primeras actividades posteriores a la revisión de la literatura, fue determinar el propósito, usos e interpretaciones de las escalas. Esta actividad es relevante ya que, de acuerdo con Kane (2009) el propósito de la medición, así como las evidencias de su uso, son el primer elemento de validez en la construcción de instrumentos de medición. Asimismo, las interpretaciones que se le darán a las escalas responden a los Estándares de las pruebas psicológicas y educativas de la AERA, APA y NCME (2018), en donde se reconoce el establecimiento de usos e interpretaciones de las pruebas como una forma de evidencias de validez de constructo.

En la actividad 1.2 se capacitó al EDDE en cuatro ejes principales: el propósito de las escalas, los fundamentos teóricos de los constructos, así como sus dimensiones, en los Estándares para las pruebas

psicológicas y educativas y algunos antecedentes de la medición de estos constructos. Se convocó una reunión con el EDDE en donde se expusieron los marcos contextuales necesarios para comprender el propósito de las escalas, así como los lineamientos de la medición según los Estándares (AERA, APA y NCME, 2018).

En la actividad 1.3. se delimitaron las subescalas, el formato, la modalidad y la estructura de las escalas. De este modo, las escalas para explorar la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias en estudiantes de ES nivel licenciatura se organizan de la siguiente manera: (1) Escala de Motivación en el Aprendizaje de las Ciencias (EMAC), (2) Escala de Actitudes hacia la Ciencia (EAC) y (3) Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC).

En esta actividad se definió también que las escalas tendrían un formato único y dos modalidades: de aplicación en físico (papel y lápiz) y en modalidad en línea en la plataforma *Google Forms*. Cabe aclarar, que para el desarrollo de la tesis y la recopilación de evidencias de validez se utilizó una digitalización de la versión en físico. En la actividad 1.4 se desarrolló la ficha técnica de las escalas, que fungió como el documento rector de los instrumentos de medición.

Etapas II. Desarrollo de escalas para la medición de la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias

Una vez delimitados los referentes de las escalas y desarrollada la ficha técnica, la siguiente etapa fue consistió en el desarrollo de los instrumentos de medición. En la actividad 2.1 se realizó la tabla de contenidos con base a los siguientes referentes: La escala de motivación se fundamenta en el Science Learning Motivation Questionarie (SLMQ-II) de Glynn et al. (2011). La segunda escala de actitudes hacia la ciencia se basa en las subescalas del Test Of Science Related Attitudes (TOSRA) de Fraser (1982) basadas en el concepto de Klopfer (1973), las últimas tres subescalas están basadas en el concepto de Mindware Contaminado por su traducción de Rizeq, Flora y Toplak (2020) para la medición de

actitudes anticientíficas. Por último, la escala de autoeficacia percibida en el aprendizaje del método científico se fundamenta en los indicadores del componente de *Competencia Científica* de la prueba PISA (OCDE, 2013) y en la Autoeficacia Percibida de Bandura (1997).

En la actividad 2.2 se realizaron las primeras versiones de los ítems de las escalas, tomando en cuenta las siguientes directrices: el contexto universitario de los estudiantes, la prudencia de las actividades académicas mencionadas en los ítems y su aplicabilidad en carreras dentro de las ciencias naturales y sociales. Se procuró que las actividades de investigación no fueran atribuidas a ningún sexo o género en específico. Así también, se tomó en cuenta la mención y consideración de materias dentro del eje de metodología de la investigación comunes dentro de la Educación Superior (ES).

Por último, en la actividad 2.3 de esta etapa se ensambló la primera versión de la EMAC, EAC y EAPAMC. Para el ensamblaje de las escalas, se utilizó el diseño del Cuestionario de Violencia Escolar de la Unidad de Evaluación Educativa (UEE), el cual se reorganizó con los apartados necesarios para los propósitos de la medición. Se realizó una revisión ortográfica y de estilo de las instrucciones, los ítems y las respuestas para colocarlas en un primer maquetado en el procesador de texto Word. Posteriormente, se realizó una segunda revisión a profundidad de las escalas, procurando corregir cualquier errata que no se haya atendido en la primera revisión.

Etapas III. Recopilación de evidencias de validez del constructo del contenido y de la imparcialidad de las escalas

Una vez que las escalas se constituyeron en el formato definitivo, la siguiente etapa consistió en la recolección de evidencias de validez de constructo en el aspecto de la imparcialidad y del contenido de las escalas. Por lo tanto, en la primera actividad de la tercera fase 3.1 se elaboró el Manual de procedimientos y se adaptó el Protocolo de validación de contenido de las escalas. Se menciona también que todo el procedimiento de validación fue acompañado con un videotutorial de elaboración propia para

realizar la recopilación de evidencias de validez de manera asincrónica. Para ver el Manual de procedimientos completo, ver **Apéndice 2**. En la actividad 3.2, se conformó el Comité de Validación de las Escalas (CVE). Una vez integrado el CVE, se le envió a cada juez a su correo electrónico una carpeta de Google Drive con los siguientes documentos adjuntos: (1) Manual de procedimientos, (2) Ficha técnica de las escalas, (3) Formato de aplicación y, un (4) Video tutorial con las instrucciones del procedimiento de recopilación de evidencias de validez. Antes de iniciar formalmente el proceso de recopilación de evidencias de validez, en la actividad los jueces leyeron y se familiarizaron con los documentos recibidos. El procedimiento de recopilación de evidencias de validez se llevó a cabo en dos principales actividades asincrónicas: (1) la recopilación de evidencias de validez de constructo en el aspecto de la imparcialidad de las escalas y (2) la recopilación de evidencias de validez de constructo en el aspecto del contenido de las escalas.

En la actividad 3.3 los jueces ingresaron al documento (3) Formato de aplicación de las escalas tal y como lo verán los respondientes. Este documento fue compartido a todos los expertos de manera colaborativa. Es decir, todos los jueces tenían acceso al mismo documento, podían comentar en el mismo formato de PDF y tenían la visualización de los comentarios de los otros jueces participantes en el proceso de validación. Es importante señalar que, por la naturaleza colaborativa y asincrónica del proceso de validación, cada juez accedió a dicho documento en momentos distintos.

De esta manera, los jueces realizaron comentarios de los ítems basándose en las categorías de la Evaluación de Diseño Universal (EDU). Cabe aclarar que se les dio la instrucción de comentar solo aquellos ítems que consideraran que requerían modificaciones sustanciales de acuerdo con las categorías anteriormente mencionadas. Una vez realizado este procedimiento, la actividad 3.4 consistió en atender la recopilación de evidencias de validez de constructo en el aspecto del contenido. En esta actividad, los jueces respondieron el Protocolo de Validación de Contenido de las escalas de manera individual.

Una vez terminado el proceso de recopilación de evidencias de validez de constructo en el aspecto de la imparcialidad y del contenido de las escalas, la actividad 3.5 consistió en ingresar los comentarios de los jueces y las respuestas del protocolo de validación de contenido a dos hojas de cálculo en Excel. Para el análisis de los comentarios de orden cualitativo se ingresaron las observaciones de los jueces en una base de datos. Posterior a un ejercicio interpretativo, se organizó cada comentario en la categoría de la EDU al que correspondiera. Asimismo, se realizó una segunda valorización de los comentarios de los jueces para confirmar que se encontraran en la categoría adecuada. Aquellos comentarios que quedaron en categorías distintas en la primera y segunda revisión requirieron una tercera en donde se estableció la categoría definitiva.

Para el análisis de los datos cuantitativos, se configuraron las ecuaciones aritméticas para calcular el Porcentaje de Acuerdo entre Jueces (PAJ), la V de Aiken (1980), la Razón de Validez de Contenido (RVC) de Lawshe (1975) y la corrección al RVC de Lawshe desarrollado por Tristán-López (2008) desarrollados con más profundidad en el apartado de fundamentos. Una vez analizados los resultados del proceso de recopilación de evidencias de validez de constructo, tanto del aspecto de la imparcialidad como del contenido de las escalas, en la actividad 3.6. se realizaron ajustes pertinentes a las escalas en los aspectos que los jueces mencionaron con más frecuencia. Por último, en la actividad 3.7 se realizó en ensamblaje de la segunda versión de las escalas.

IV. RESULTADOS

El capítulo se organiza en tres grandes subapartados asociados con los resultados obtenidos en las etapas de diseño, desarrollo y validación de las tres escalas para medir la motivación en el aprendizaje de las ciencias (EMAC), las actitudes hacia la ciencia (EAC) y la autoeficacia percibida en el aprendizaje del método científico (EAPAMC) desde un enfoque integrador en estudiantes de ES a nivel de licenciatura. En el primer subapartado, se muestra los resultados de las etapas de diseño y desarrollo de la primera versión de dichas escalas. En el segundo subapartado, se muestran los resultados del cálculo del IVC obtenidos con los métodos de Lawshe (1975), V-Aiken (1985) y Tristán-López (2008). Por su parte, en el tercer subapartado se muestran los resultados de las observaciones del CVE con base en los criterios de la EDU, así como el ajuste de los ítems con base en la integración de evidencias de validez recopiladas y analizadas en el estudio.

4.1. Ficha Técnica, Tablas de contenido y Escalas

El producto derivado de la etapa de diseño es la Ficha Técnica o documento rector de las escalas. Este documento incluye la estructura, el formato, la modalidad, así como los principales referentes teóricos y metodológicos para el diseño de los instrumentos de medición. El propósito de realizar la ficha técnica es contar con una evidencia formal que acompañe y de claridad al proceso de diseño y desarrollo de las escalas, así como la recopilación de evidencias de validez de las mismas. Es importante mencionar también, que un documento que contenga la definición clara del constructo, así como sus referentes teóricos y metodológicos es una de las principales sugerencias de los *Estándares* para el desarrollo de pruebas psicológicas y educativas de la AERA, APA y NCME (2018).

El documento de la ficha técnica de las escalas está compuesto por nueve apartados descritos en la **Tabla 4**, que son:

Tabla 4.

Ficha Técnica de las escalas para medir la motivación, actitudes y autoeficacia percibida en el aprendizaje de las ciencias.

No. Apartado	Contenido
I.	Nombre de las escalas en conjunto
II.	Instrumentos y formato
III.	Usos y propósitos
IV.	Antecedentes del estudio y medición de la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias,
V.	Referentes teóricos y metodológicos de las escalas
VI.	Referentes de la Escala de Motivación en el Aprendizaje de las Ciencias (EMAC)
VII.	Referentes de la Escala de Actitudes hacia la Ciencia (EAC)
VIII.	Referentes de la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC)
IX.	Enfoque de medición.

En la **Tabla 4** se observa una síntesis de la ficha técnica de las escalas, en la columna de la derecha se ubican los nueve apartados que componen las secciones del documento. Asimismo, del lado derecho de cada segmento se encuentra una síntesis de la información que aparece en cada capítulo. Los segmentos uno y dos son aspectos técnicos de las escalas como el nombre, el formato, la modalidad y los instrumentos de medición. El apartado tres responde al propósito de la medición, así como los usos e interpretaciones de las escalas.

Los apartados cuatro y cinco, se componen de los antecedentes de las escalas, es decir, otras propuestas de medición para estos constructos y de los instrumentos, escalas y subescalas que fueron referencia directa para el diseño y desarrollo de los instrumentos de medición, respectivamente. Los apartados seis, siete y ocho se compone de las tablas de contenido de las escalas, así como de las descripciones generales de cada instrumento como las subescalas que lo componen, la cantidad de ítems y el tipo de respuesta y codificación. Por último, en el apartado nueve se encuentra el modelo de medición utilizado para el diseño y desarrollo de los instrumentos de medición.

Los productos derivados de la etapa de desarrollo son la tablas de contenido por escala y la primera versión de la EMAC, EAC y EAPAMC. Las tablas de contenido consisten en la sistematización y operacionalización de los constructos a medir en las escalas desarrolladas. Dichas tablas se componen de tres columnas que se dividen de izquierda a derecha en: subescalas, indicadores y número de ítems que componen la subescala. Asimismo, en la introducción de cada Tabla de contenidos se menciona el tipo y código de respuesta de cada escala. En la **Tabla 5** se muestra una visión general de las tres escalas desarrolladas.

Tabla 5.

Escalas para medir la Motivación, Actitudes y Autoeficacia en el Aprendizaje de las Ciencias en estudiantes de Educación Superior

Escala	Subescalas
Escala de Motivación en el Aprendizaje de las Ciencias (EMAC) 4 subescalas	Motivación intrínseca Motivación Extrínseca Autodeterminación Autoeficacia Implicaciones sociales de la ciencia Normalidad de los científicos Actitudes hacia la investigación Adopción de actitudes científicas
Escala de Actitudes hacia la Ciencia (EAC) 10 subescalas	Interés en las experiencias de aprendizaje de las ciencias Interés en el aprendizaje de las ciencias extraescolares o en tiempo libre Interés en una especialización o trabajo relacionado con las ciencias. Actitudes anticientíficas Creencias conspiranoicas Creencias paranormales
Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC) 3 subescalas	Explicar un fenómeno científicamente Diseñar y evaluar investigación científica Interpretar datos y evidencia científicamente.

De lado izquierdo de la tabla se observa el nombre completo de cada escala y el número de subescalas que la componen. De lado derecho se encuentran los nombres de cada subescala. En la **Tabla 6** se observa el contenido de la Escala de Motivación en el Aprendizaje de las Ciencias (EMAC).

Tabla 6.

Tabla de contenidos de la Escala de Motivación en el Aprendizaje de las Ciencias

Subescalas	Indicadores
Motivación intrínseca (7 ítems)	El estudiante... Se muestra atento en clases asociadas al aprendizaje de contenidos científicos Muestra satisfacción de aprender ciencia, solo por el hecho de aprender. Siente interés por las teorías y explicaciones científicas.
Motivación extrínseca (5 ítems)	Percibe el aprendizaje de las ciencias como un medio para un fin tangible como ingresar a un posgrado u obtener un buen empleo.
Autodeterminación (3 ítems)	Tiene control percibido sobre su aprendizaje de las ciencias.
Autoeficacia (11 ítems)	Tiene la creencia de que puede lograr buenos resultados en materias relacionadas a contenidos específicos de ciencias o procedimentales como metodología de la investigación.

En la columna de la izquierda se encuentran las subescalas y su respectivo número de ítems, mientras que en la columna de la derecha se encuentran los indicadores guía para el desarrollo de los reactivos.

En la **Tabla 7** se observa el contenido de la Escala de Actitudes hacia la Ciencia (EAC).

Tabla 7.

Tabla de contenido de Escala de Actitudes hacia la Ciencia (EAC)

Subescalas	Indicadores
<p>Implicaciones sociales de la ciencia (4 ítems)</p>	<p>El estudiante...</p> <p>Tiene actitudes positivas hacia el progreso científico. Considera que el desarrollo de la ciencia trae consigo beneficios para la sociedad.</p> <p>Tiene actitudes positivas hacia la ciencia y considera que el desarrollo de conocimiento científico es una manera adecuada de solucionar problemas de diversas naturalezas.</p> <p>Tiene una actitud positiva hacia la ciencia, considera que el conocimiento que generan los científicos le impacta directamente de manera positiva.</p> <p>Tiene una actitud negativa hacia el progreso científico y tecnológico. Considera que el desarrollo científico puede traer desventajas y consecuencias negativas para la sociedad.</p> <p>Tiene la apreciación de que los científicos tienen un estilo de vida normal como cualquier otro profesionista.</p> <p>Tiene una apreciación de que el trabajo científico es muy desgastante. Considera que las personas que se dedican a la ciencia no tienen tiempo para descansar.</p> <p>Tiene la apreciación de que las personas que se dedican a la ciencia dedican todo su tiempo al trabajo y poco tiempo a la familia y otros vínculos sociales.</p>
<p>Normalidad de los científicos (5 ítems)</p>	<p>Tiene la apreciación de que las personas que se dedican a la ciencia.</p> <p>Tiene la apreciación de que las personas que se dedican a la ciencia son estereotípicamente excéntricas.</p> <p>Tiene una actitud positiva hacia la aplicación del método científico para entender el mundo</p>
<p>Actitud hacia la investigación (4 ítems)</p>	<p>Tiene una actitud positiva hacia el conocimiento desarrollado por la investigación científica, y considera que puede servirle para mejorar aspectos de su vida diaria.</p> <p>Tiene una buena actitud hacia el método científico y reconoce el experimento como un diseño de investigación y le gustaría aprender sobre metodología científica.</p> <p>Valora la actividad científica y le gustaría aprender cómo se desarrollan las investigaciones científicas.</p> <p>Valora la investigación científica como una forma de obtener información sobre el mundo, así como siente curiosidad por comprender de manera científica fenómenos de diversas naturalezas.</p>
<p>Adopción de actitudes científicas (5 ítems)</p>	<p>Se muestra escéptico de la información que consulta si no muestra la evidencia científica en la que está basada. Con base a este indicador se desarrollan dos reactivos, el 7.15 orientado a la desconfianza hacia la información que no muestra fuentes de evidencia científica. El ítem 7.16 está enfocado en la preferencia del estudiante por tomarse el tiempo de verificar si las fuentes que consulta son confiables.</p> <p>Desarrolla su pensamiento crítico por medio del cuestionamiento y corroboración de información que se le brinda.</p> <p>Tiene una actitud positiva hacia la evidencia científica e inclinación por la objetividad de la ciencia.</p>
<p>Disfrute de las experiencias de aprendizaje de las ciencias (5 ítems)</p>	<p>Siente gusto por materias que tienen como objetivo la formación científica de los estudiantes. Con este indicador se desarrollaron tres ítems.</p> <p>Disfruta las experiencias relacionadas a aprender sobre las fases de una investigación científica.</p>
<p>Interés en el aprendizaje de las ciencias (tiempo libre) (5 ítems)</p>	<p>Disfruta aprendiendo sobre ciencias en su tiempo libre, consume productos de entretenimiento sobre temas científicos diversos.</p>

<p>Interés en una especialización científica, o por opciones laborales relacionadas a la investigación científica. (6 ítems)</p>	<p>Tiene interés en una especialización o estudios de posgrado de tipo científico, o considera la investigación científica como una opción laboral.</p>
<p>Actitudes Anticientíficas (4 ítems)</p>	<p>Siente rechazo y oposición a los métodos objetivos de la ciencia, considera que los métodos científicos tienen poca credibilidad. Suele confiar más en su intuición que en la evidencia científica.</p>
<p>Creencias Conspiranoicas (3 ítems)</p>	<p>Tiene una tendencia a respaldar ideas conspirativas y atribuir la causa de un evento a complots secretos de grupos o fuerzas poderosas, aunque otras explicaciones sean más probables.</p>
<p>Creencias paranormales (5 ítems)</p>	<p>Tiene creencias paranormales como la psicoquinesis, la existencia de hechicería, magia negra, supersticiones, comunicación con personas fallecidas y precognición, como por ejemplo la capacidad de la astrología para predecir el futuro.</p>

De igual manera, en la columna de la izquierda se encuentran las diez subescalas de la EAC, mientras que en la columna de la derecha se observa los indicadores desde donde se construyeron los ítems. En la **Tabla 8** se encuentra la Tabla de contenido de la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC).

Tabla. 8

Tabla de contenido de la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC)

Escala	Subescalas
<p>Explicar un fenómeno científicamente (7 ítems)</p>	<p>El estudiante... Tiene la percepción de que puede recordar y aplicar conocimientos científicos adecuadamente. Tiene la percepción de que puede identificar, utilizar y generar modelos explicativos y representaciones de la realidad. Tiene la percepción de que puede realizar y justificar predicciones apropiadas con conocimiento científico. Tiene la percepción de que puede ofrecer hipótesis explicativas. Tiene la percepción de que puede explicar las posibles implicaciones del conocimiento científico para la sociedad.</p>
<p>Diseñar y evaluar investigación científica (5 ítems)</p>	<p>El estudiante... Tiene la percepción de que puede identificar la pregunta explorada en un estudio científico dado. Tiene la percepción de que puede distinguir preguntas que sean posibles de investigar científicamente. Tiene la percepción de que puede proponer una forma de explorar científicamente una cuestión determinada. Tiene la percepción de que puede evaluar formas de explorar científicamente una cuestión determinada. Tiene la percepción de que puede describir y evaluar una variedad de formas que utilizan los científicos para asegurar la confiabilidad de los datos y la objetividad y generalizabilidad de las explicaciones.</p>
<p>Interpretar datos y evidencia científicamente (5 ítems)</p>	<p>El estudiante... Tiene la percepción de que puede transformar datos de una representación a otra. Tiene la percepción de que puede analizar e interpretar datos y sacar conclusiones apropiadas. Tiene la percepción de que puede identificar los supuestos, la evidencia y el razonamiento en textos relacionados con la ciencia. Tiene la percepción de que puede distinguir entre argumentos que se basan en evidencia científica y teoría y aquellos que se basan en otras consideraciones. Tiene la percepción de que puede evaluar argumentos científicos y evidencia de diferentes fuentes (por ejemplo, periódicos, internet, revistas).</p>

En la columna de la izquierda se observan las subescalas de la EAPAMC y el número de reactivos que componen cada subescala, mientras que en la columna de la derecha se encuentra los indicadores utilizados para el desarrollo de los reactivos. Ahora bien, en las tablas subsecuentes se encuentran las

primeras versiones de las escalas. En la **Tabla 9** se encuentra la Escala de Motivación en el Aprendizaje de las Ciencias (EMAC), en la **Tabla 10** se encuentra la Escala de Actitudes de la Ciencia (EAC) y en la **Tabla 11** se encuentra la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC).

Tabla 9

Escala de Motivación en el Aprendizaje de las Ciencias (EMAC)

Subescala de Motivación Intrínseca	
Afirmaciones	
6.1.	Me resulta interesante aprender el contenido que se enseña en las clases de metodología de la investigación
6.2.	Me siento motivado para concretar las actividades que solicitan en la clase de metodología de la investigación (por ejemplo, una revisión de la literatura sobre un tema de mi interés).
6.3.	Conocer hallazgos científicos es relevante para mi vida tanto dentro como fuera de la escuela.
6.4.	Disfruto leer sobre hallazgos científicos, aunque no sean parte de mis actividades de clase.
6.5.	Es interesante interpretar el mundo en términos científicos.
6.6.	Siento interés por conocer hallazgos científicos en diversas áreas del conocimiento
6.7.	Siento interés por aprender teorías y explicaciones científicas asociadas a mi profesión.
Subescala de Motivación Extrínseca	
6.8.	Adquirir habilidades científicas me será útil para ingresar a un posgrado en un futuro.
6.9.	Adquirir habilidades científicas me permitirá obtener un mejor empleo al salir de la carrera.
6.10.	Aprender a redactar una propuesta de investigación es una habilidad que me servirá al salir de la universidad.
6.11.	Aprender a utilizar programas estadísticos para analizar datos me dará una ventaja profesional en el futuro.
6.12.	Dedicarme a la ciencia me permitirá tener estabilidad económica en el futuro.
Subescala de Autodeterminación	
6.13.	Cuando no entiendo alguna parte del proceso de investigación, busco a alguien que me ayude a resolver mis dudas.
6.14.	Me esfuerzo lo suficiente para comprender los contenidos de la clase de metodología de la investigación.
6.15.	Si es necesario, invierto más tiempo en estudiar los contenidos de la clase de metodología de la investigación.
Subescala de Autoeficacia	
6.16.	Puedo realizar una propuesta de investigación como proyecto final de clase.
6.17.	Puedo ingresar variables y datos en un programa estadístico.
6.18.	Puedo utilizar un programa de análisis de datos para una práctica de clase.
6.19.	Puedo leer un artículo científico y explicar el contenido a la clase.
6.20.	Puedo diferenciar entre información científica y no científica en libros o contenidos de internet.
6.21.	Puedo recolectar datos con cuestionarios.
6.22.	Tengo la habilidad de buscar información científica confiable.
6.23.	Puedo identificar las partes de un artículo.
6.24.	Puedo redactar una propuesta de investigación para ingresar a un posgrado.
6.25.	Puedo explicar con mis propias palabras la metodología de un artículo científico.
6.26.	Puedo encontrar artículos científicos en bases de datos (por ejemplo, ERIC, EBSCO, Wiley, entre otros).

- 6.27. Puedo seleccionar una encuesta o cuestionario apropiado para recolectar datos.
 6.28. Puedo realizar el diseño de un experimento sencillo.
 6.29. Puedo realizar una revisión de la literatura científica sobre un tema.

Nota. El tipo de respuesta es en formato de escala ordinal con cuatro categorías de respuesta: “Totalmente en desacuerdo, en desacuerdo, de acuerdo y totalmente de acuerdo”.

Tabla 10.

Escala de Actitudes hacia la Ciencia (EAC)

Subescala de Implicaciones Sociales de la Ciencia	
Afirmaciones	
7.1.	Pienso que el desarrollo científico trae múltiples beneficios a la sociedad.
7.2.	Pienso que la ciencia es una buena forma de solucionar problemas de tipo natural y social.
7.3.	Pienso que el desarrollo de conocimiento científico contribuye a mejorar mi calidad de vida.
7.4.	Pienso que un rápido progreso científico y tecnológico puede traer más desventajas que ventajas para la sociedad.
Subescala de Normalidad de los Científicos	
7.5.	Las personas que se dedican a la ciencia tienen un estilo de vida normal.
7.6.	Las personas que se dedican a la ciencia no tienen descanso.
7.7.	Para dedicarse a la ciencia, se tiene que sacrificar otras áreas de la vida como la social y la familiar.
7.8.	Las personas que se dedican a la actividad científica suelen ser rígidas e inflexibles.
7.9.	Las personas que se dedican a la ciencia son muy excéntricas.
Subescala de Actitud hacia la Investigación	
7.10.	Es mejor aplicar el método científico para entender el mundo que utilizar el sentido común o la intuición.
7.11.	Si buscara como mejorar mis hábitos de estudio preferiría buscar información científica que consultar otras fuentes de información.
7.12.	Me gusta aprender sobre cómo las personas que se dedican a la ciencia diseñan sus experimentos.
7.13.	Me gusta aprender sobre cómo las personas que se dedican a la ciencia generan y aplican sus ideas de investigación.
7.14.	Siento curiosidad por fenómenos naturales como el cambio climático, o sociales como el comportamiento de las personas.
Subescala de Adopción de Actitudes Científicas	
7.15.	Dudo de la información que consulto si no muestra evidencia científica que la sustente.
7.16.	Aunque requiera de más tiempo, me gusta revisar si las fuentes de información que consulto son confiables.
7.17.	Me gusta cuestionar y corroborar la información que se me brinda en clases.
7.18.	Me gusta dar opiniones objetivas y basadas en datos científicos.
Subescala de interés en las experiencias de aprendizaje de las ciencias	
7.19.	Me gustan las materias en donde puedo aprender cómo se genera el conocimiento científico.
7.20.	Me gustaría que en mi carrera hubiera más materias para aprender cómo se genera el conocimiento científico.
7.21.	Me gusta la clase de metodología de la investigación científica.
7.22.	Me gusta aprender sobre las distintas fases del método científico.
7.23.	Me gustaría aprender sobre el proceso de investigación científica realizando una investigación como proyecto de clase.

Subescala de interés en el aprendizaje de las ciencias extraescolares o en tiempo libre

- 7.24. Me interesan los videos en Youtube que tratan sobre temas científicos diversos.
- 7.25. Cuando tengo tiempo, me gusta ver documentales de temas científicos (por ejemplo, sobre los avances de las neurociencias).
- 7.26. Me gusta ver películas que tratan sobre la vida de científicos famosos o sobre descubrimientos relevantes de la ciencia.
- 7.27. Me gusta leer libros sobre temas científicos ya sea en formato análogo o digital.
- 7.28. Me gusta escuchar podcast (contenido auditivo con formato similar a la radio) sobre temas científicos diversos.

Subescala de Interés en una especialización o trabajo relacionado con las ciencias

- 7.29. Al finalizar la carrera, me gustaría entrar a un posgrado especializado en investigación científica.
- 7.30. En un futuro, me gustaría aportar conocimiento científico a mi campo profesional.
- 7.31. Consideraría dedicarme a la investigación científica en un futuro.
- 7.32. Cuando termine mi carrera me gustaría involucrarme en algún proyecto de investigación científica.
- 7.33. Es importante seguir mi formación científica al terminar mi carrera.
- 7.34. Dedicarme a la investigación es una de mis opciones laborales en el futuro.

Subescala de Actitudes Anticientíficas

- 7.35. No confío en los métodos que usa la ciencia.
- 7.36. Confío más en mi sexto sentido para solucionar problemas que en las evidencias científicas.
- 7.37. Prefiero utilizar mi intuición antes que aplicar procedimientos científicos para resolver problemas de la vida cotidiana.
- 7.38. Confío en los métodos de la ciencia.

Subescala de Creencias Conspiranoicas

- 7.39. Los científicos nos ocultan la cura de enfermedades graves como el cáncer.
- 7.40. Los científicos no aceptan la medicina alternativa porque representa un peligro para los intereses de la industria farmacéutica.
- 7.41. Los científicos trabajan en secreto en cosas que podrían perjudicar a la humanidad.

Subescala de Creencias Paranormales

- 7.42. La posición de los planetas puede influir en la personalidad y el comportamiento de las personas.
- 7.43. Aunque la ciencia no lo acepte, hay formas de comunicarse con personas fallecidas.
- 7.44. Existen personas con la capacidad de atraer sucesos con la mente.
- 7.45. Existen personas con la habilidad para comunicarse con otras personas por medio de la mente.
- 7.46. Creo que puedo atraer situaciones favorables si me concentro lo suficiente.

Nota. El tipo de respuesta es en formato de escala ordinal con cuatro categorías de respuesta: “Totalmente en desacuerdo, en desacuerdo, de acuerdo y totalmente de acuerdo”.

Tabla 11.*Escala de Autoeficacia Percibida en el Aprendizaje de las Ciencias (EAPAMC)*

Subescala de Explicar un fenómeno científicamente	
Actividades	
8.1.	Ofrecer explicaciones científicas para fenómenos naturales y tecnológicos.
8.2.	Recordar y aplicar conocimiento científico.
8.3.	Utilizar modelos y representaciones para explicar la realidad.
8.4.	Explicar una teoría científica
8.5.	Realizar predicciones apropiadas de acuerdo con conocimiento científico previo.
8.6.	Ofrecer hipótesis científicas
8.7.	Explicar apropiadamente las implicaciones del conocimiento científico para la sociedad.
Subescala de Diseñar y evaluar investigación científica	
8.8.	Identificar las preguntas de investigación de un estudio científico.
8.9.	Reconocer cuestiones que pueden ser estudiadas de manera científica.
8.10.	Proponer otras formas de abordar o investigar cuestiones científicas.
8.11.	Proponer formas en las que se podría explorar científicamente alguna cuestión determinada
8.12.	Reconocer las formas en las que los científicos garantizan la objetividad de las investigaciones científicas
Subescala de Interpretar datos y evidencia científica	
8.13.	Transformar datos en representaciones gráficas
8.14.	Analizar datos científicos para sacar conclusiones pertinentes.
8.15.	Identificar los supuestos y el razonamiento con el que se desarrollan los textos científicos.
8.16.	Reconocer la evidencia científica de otras fuentes de información.
8.17.	Distinguir argumentos que se basan en evidencia científica.

Nota. El tipo de respuesta es en formato de escala ordinal con cuatro categorías de respuesta: “Nada en absoluto, Poco, Moderado y Fuerte”.

4.2. Validación del aspecto del contenido por escala

En este subapartado se muestran los resultados del cálculo de distintos índices de validez de contenido por criterio, ítem y escala. Asimismo, se hace una descripción de aquellos ítems que no cumplieron con el estándar establecido por Aiken (1985), Lawshe (1975) y Tristán-López (2008).

4.2.1. Escala de Motivación en el Aprendizaje de las Ciencias (EMAC)

En la **Tabla 12** se muestran los resultados del cálculo del Porcentaje de Acuerdo entre Jueces (PAJ \geq .80). Es decir, el promedio de acuerdo entre jueces por criterio (relevancia, simplicidad, claridad y precisión), por ítem y global por escala.

Tabla 12.

Porcentaje de acuerdo entre jueces especialistas (PAJ) de la EMAC.

Ítem	Relevancia	Precisión	Claridad	Simplicidad	Promedio
6.1	0.92	0.89	0.92	0.86	0.90
6.2	0.86	0.69*	0.69*	0.75*	0.75*
6.3	0.94	0.89	0.83	0.83	0.87
6.4	0.97	0.89	0.92	0.86	0.91
6.5	0.81	0.56*	0.61*	0.86	0.71*
6.6	0.92	0.83	0.75*	0.81	0.83
6.7	1.00	0.92	0.83	0.94	0.92
6.8	0.97	0.72*	0.78*	0.89	0.84
6.9	0.92	0.69*	0.75*	0.89	0.81
6.10	0.94	0.75*	0.81	0.81	0.83
6.11	0.94	0.69*	0.78*	0.83	0.81
6.12	0.92	0.83	0.89	0.89	0.88
6.13	0.94	0.86	0.92	0.94	0.92
6.14	0.89	0.81	0.89	0.97	0.89
6.15	0.89	0.72*	0.78*	0.89	0.82
6.16	0.97	0.89	0.89	0.97	0.93
6.17	0.75*	0.64*	0.78*	0.81	0.74*
6.18	0.83	0.81	0.78*	0.94	0.84
6.19	0.94	0.78*	0.75*	0.86	0.83
6.20	0.94	0.92	0.86	0.97	0.92
6.21	0.75*	0.81	0.67*	0.83	0.76*
6.22	0.86	0.86	0.86	0.94	0.88
6.23	0.94	0.75*	0.78*	0.86	0.83
6.24	0.89	0.86	0.89	0.97	0.90
6.25	0.97	0.83	0.69*	0.86	0.84
6.26	0.97	0.89	0.86	0.97	0.92
6.27	0.92	0.92	0.78*	0.94	0.89
6.28	0.92	0.86	0.78*	0.86	0.85
6.29	1.00	1.00	0.89	0.97	0.97

Nota. *PAJ < 0.80

En cuanto a los resultados del PAJ por criterio, el ítem 6.17 y 6.21 de la subescala de *Autoeficacia* no cumplieron con el estándar ($PAJ \geq .80$) en el criterio de *relevancia*. En el criterio de *precisión* los ítems que obtuvieron un puntaje menor a 0.80 fueron reactivos 6.2 y 6.5 de la subescala de *Motivación intrínseca* con 0.69 y 0.56. Los ítems 6.8, 6.9, 6.10 y 6.11 de la subescala de *Motivación extrínseca* con 0.72, 0.69, 0.75 y 0.69. El ítem 6.15 de la subescala de *autodeterminación* con 0.72 y por último, los ítems 6.17, 6.19 y 6.23 de la subescala de *Autoeficacia* con 0.64, 0.78 y 0.75, respectivamente.

Asimismo, en el criterio de *claridad* 15 ítems de 29 tuvieron un PAJ menor a 0.80. Los ítems 6.2, 6.5, 6.6 de la subescala de *Motivación intrínseca* obtuvieron 0.69, 0.61 y 0.75 respectivamente. En la subescala de *Motivación extrínseca* los ítems 6.8, 6.9 y 6.11 obtuvieron un porcentaje de acuerdo entre jueces de 0.78, 0.75 y 0.78. En la subescala de *Autodeterminación* el ítem 6.15 obtuvo 0.78, y por último, en la subescala de *Autoeficacia* los ítems 6.17, 6.18, 6.19, 6.21, 6.23, 6.25, 6.27 y 6.28 obtuvieron puntajes de 0.78, 0.78, 0.75, 0.67, 0.78, 0.69, 0.78 y 0.78, respectivamente.

En el criterio de *simplicidad* solo el ítem 6.2 de la subescala de *Motivación intrínseca* obtuvo un puntaje menor a 0.80. Por último, en cuanto a los promedios generales que brindan un panorama más amplio de porcentaje de acuerdo entre jueces tomando en cuenta los cuatro criterios de relevancia, precisión, claridad y simplicidad, los resultados fueron los siguientes: cuatro de 29 ítems no cumplieron con el estándar de 0.80. Los ítems 6.2, 6.5, 6.17 y 6.21 obtuvieron porcentajes de 0.75, 0.71, 0.74 y 0.76, respectivamente. Asimismo, estos reactivos pertenecen en orden de aparición a las subescalas de *Motivación intrínseca* (6.2 y 6.5) y *Autoeficacia* (6.17 y 6.21). En cuanto a los resultados del cálculo del IVC, en la **Tabla 13** se muestran los resultados.

Tabla 13*Resultados del cálculo del IVC de la EMAC*

Ítem	V-Aiken	RVC Lawshe	RVC' Tristán-López
6.1	0.89	0.42*	0.71
6.2	0.75*	-0.08*	0.46*
6.3	0.88	0.38*	0.69
6.4	0.91	0.54*	0.77
6.5	0.70*	-0.17*	0.42*
6.6	0.83	0.21*	0.60
6.7	0.92	0.67	0.83
6.8	0.84	0.25*	0.63
6.9	0.81	0.08*	0.54*
6.10	0.83	0.29*	0.65
6.11	0.81	0.38*	0.69
6.12	0.88	0.46*	0.73
6.13	0.92	0.54*	0.77
6.14	0.89	0.54*	0.77
6.15	0.82	0.17*	0.58
6.16	0.93	0.71	0.85
6.17	0.74*	0.04*	0.52*
6.18	0.84	0.38*	0.69
6.19	0.83	0.29*	0.65
6.20	0.92	0.71	0.85
6.21	0.76*	0.17*	0.58
6.22	0.88	0.58	0.79
6.23	0.83	0.38*	0.69
6.24	0.90	0.67	0.83
6.25	0.84	0.38*	0.69
6.26	0.92	0.67	0.83
6.27	0.89	0.54*	0.77
6.28	0.85	0.42*	0.71
6.29	0.87	0.88	0.94

Nota. *V de Aiken < 0.69; RVC de Lawshe <0.56; RVC' <0.58

En cuanto a los resultados del cálculo de la V de Aiken, se encontró que todos los ítems de la EMAC cumplieron con el parámetro ($\geq .69$) establecido por Aiken (1985). Sin embargo, se reportan los ítems con las puntuaciones menores: 6.2 y 6.5 de la subescala de *Motivación intrínseca* con .75, .70 y los ítems 6.17 y 6.21 de la subescala de *Autoeficacia* con .74 y .76, respectivamente.

Asimismo, en los resultados del cálculo del RVC de Lawshe (1975) se encontró que 22 de 29 ítems obtuvieron un puntaje menor a .56, parámetro propuesto por este autor. En la subescala de *Motivación intrínseca*, los ítems del 6.1 al 6.6 obtuvieron puntajes inferiores, siendo el 6.2 y el 6.5 los puntajes más bajos con -0.08 y -0.17, respectivamente. De la subescala de *Motivación extrínseca*, los cinco ítems tuvieron puntuaciones inferiores al criterio, siendo los más bajos los ítems 6.8, 6.9 y 6.10

con 0.25, 0.08 y 0.29. En la subescala de *Autodeterminación*, el ítem con el puntaje más bajo fue el 6.15 con 0.17. Por último, en la subescala de *Autoeficacia*, los ítems con los puntajes más bajos fueron el 6.17, 6.18, 6.19 y 6.21 con 0.04, 0.38, 0.29 y 0.17, respectivamente. En cuanto al cálculo del RVC' de Tristán-López (2008) se encontró que cuatro ítems no cumplieron con el parámetro de 0.58 sugerido por el autor. En la subescala de *Motivación intrínseca* los puntajes más bajos los obtuvieron los ítems 6.2 y 6.5 con 0.46 y 0.42. De la subescala de *Motivación extrínseca* el ítem 6.9 con 0.54 y por último, de la subescala de *Autoeficacia* el ítem 6.17 con 0.52.

4.2.2. Escala de Actitudes hacia la Ciencia (EAC)

En la **Tabla 14** se muestran los resultados del cálculo del Porcentaje de Acuerdo entre Jueces (PAJ \geq .80). Es decir, el promedio de acuerdo entre jueces por criterio (relevancia, simplicidad, claridad y precisión), por ítem y global por escala de la Escala de Actitudes hacia la Ciencia (EAC).

Tabla 14
Porcentaje de Acuerdo entre Jueces (PAJ) de la EAC por ítem

Reactivo	Relevancia	Precisión	Claridad	Simplicidad	Promedio
7.1	1.00	0.97	0.97	0.86	0.95
7.2	0.97	0.81	0.78*	0.78*	0.83
7.3	1.00	0.94	0.92	0.92	0.94
7.4	0.94	0.75*	0.56*	0.64*	0.72*
7.5	0.72*	0.58*	0.58*	0.64*	0.63*
7.6	0.86	0.81	0.86	0.83	0.84
7.7	0.92	0.86	0.86	0.86	0.87
7.8	0.81	0.81	0.81	0.69*	0.78*
7.9	0.72*	0.72*	0.61*	0.72*	0.69*
7.10	0.94	0.86	0.75*	0.81	0.84
7.11	0.75*	0.58*	0.64*	0.72*	0.67*
7.12	0.94	0.75*	0.75*	0.81	0.81
7.13	0.94	0.83	0.81	0.86	0.86
7.14	0.81	0.67*	0.69*	0.78	0.74*
7.15	0.97	0.94	0.92	0.92	0.94
7.16	0.97	0.97	0.92	0.92	0.94
7.17	0.97	0.97	0.94	1.00	0.97
7.18	0.83	0.81	0.81	0.86	0.83
7.19	0.94	0.94	0.97	0.97	0.96
7.20	0.97	0.97	0.97	0.97	0.97
7.21	0.78*	0.83	0.89	0.92	0.85
7.22	0.92	0.89	0.89	0.92	0.90
7.23	0.97	0.89	0.89	0.89	0.91

7.24	0.89	0.56*	0.75*	0.86	0.76*
7.25	0.97	0.67*	0.72*	0.78*	0.78*
7.26	0.92	0.81	0.86	0.86	0.86
7.27	0.97	0.81	0.81	0.86	0.86
7.28	0.94	0.89	0.83	0.78*	0.86
7.29	1.00	0.92	0.89	0.92	0.93
7.30	0.94	0.94	0.92	0.94	0.94
7.31	0.92	0.81	0.86	0.94	0.88
7.32	0.89	0.81	0.83	0.89	0.85
7.33	0.89	0.78*	0.83	0.89	0.85
7.34	0.97	0.92	0.92	0.97	0.94
7.35	0.69*	0.67*	0.64*	0.75*	0.69*
7.36	0.89	0.83	0.83	0.92	0.87
7.37	0.92	0.89	0.83	0.89	0.88
7.38	0.89	0.89	0.89	0.89	0.89
7.39	0.92	0.94	0.94	0.94	0.94
7.40	0.89	0.92	0.94	0.94	0.92
7.41	0.94	0.92	0.94	0.94	0.94
7.42	0.92	0.94	0.89	0.89	0.91
7.43	0.94	1.00	1.00	1.00	0.99
7.44	0.94	0.97	0.94	0.92	0.94
7.45	0.94	0.92	0.89	0.92	0.92
7.46	0.81	0.86	0.86	0.92	0.86

Nota. *PAJ < 0.80

En cuanto al PAJ de la EAC en el criterio de relevancia, en la subescala de *Implicaciones Sociales de la ciencia*, todos los ítems cumplieron con el estándar de .80. Por otro lado, en la subescala de *Normalidad de los científicos* el ítem 7.5 y 7.9 obtuvieron ambos un puntaje de 0.72. En la subescala de *Actitud hacia la investigación*, el ítem 7.11 obtuvo una puntuación de 0.75. En el caso de la subescala de *Adopción de actitudes científicas*, todos los ítems que abarcan del 7.15 al 7.18 obtuvieron puntajes superiores al estándar requerido. En la subescala *Interés en las experiencias de aprendizaje de las ciencias*, el ítem 7.21 obtuvo un porcentaje más bajo del establecido en el parámetro. Asimismo, en la subescala de *Interés en el aprendizaje de las ciencias extraescolares o en tiempo libre*, del ítem 7.24 al 7.28 todos cumplieron con el criterio de .80, al igual que la subescala de *Interés en una especialización o trabajo relacionado con las ciencias* que abarca de los ítems 7.29 al 7.34.

Las últimas tres subescalas de la EAC, comenzando por la de *Actitudes Anticientíficas* del ítem 7.35 al 7.38, únicamente el ítem 7.35 obtuvo una puntuación inferior a la esperada con 0.69. Por último, la subescala de *Creencias conspiranoicas* que abarca los ítems 7.39, 7.40 y 7.41 y la subescala de

Creencias paranormales que abarca del ítem 7.42 al 7.46, obtuvieron un puntaje igual o mayor a 0.80 en el criterio de relevancia.

En el criterio de *precisión*, el ítem 7.4 de la subescala de *Implicaciones sociales de la ciencia* obtuvo 0.75, mientras que en la subescala de *Normalidad de los científicos* los ítems 7.5 y 7.9 obtuvieron 0.58 y 0.72, respectivamente. En la subescala de *Actitudes hacia la investigación*, los ítems 7.11, 7.12 y 7.14 obtuvieron puntajes de 0.58, 0.75 y 0.67, mientras que los ítems de las subescalas de *Adopción de actitudes científicas* e *Interés en las experiencias de aprendizaje de las ciencias* todos los ítems correspondientes obtuvieron puntajes iguales o superiores a 0.80. Sin embargo, en la *subescala de Interés en el aprendizaje de las ciencias extraescolares o en el tiempo libre* los ítems 7.24 y 7.25 obtuvieron puntajes de 0.56 y 0.67, mientras que en la subescala de *Interés en una especialización o trabajo relacionado con las ciencias* el ítem 7.33 obtuvo un puntaje de 0.78. Por otro lado, en la subescala de *Actitudes anticientíficas*, el ítem 7.35 obtuvo 0.67, mientras que las últimas dos subescalas obtuvieron puntajes favorables en el criterio de *precisión*.

En cuanto a los resultados de PAJ en el criterio de *claridad* de la EAC, los ítems 7.2 y 7.4 de la subescala de *Implicaciones sociales de la ciencia*, obtuvieron puntajes de 0.78 y 0.56, respectivamente. Mientras que en la subescala de *Normalidad de los científicos* los ítems que obtuvieron puntuaciones más bajas en el PAJ fueron los ítems 7,5 y 7.9 con 0.58 y 0.61. La subescala de *Actitud hacia la investigación*, los ítems 7.10, 7.11, 7.12 y 7.14, es decir, la mayoría de los ítems de la subescala, obtuvieron puntajes de 0.75, 0.64, 0.75 y 0.69.

Por otro lado, en las subescalas de *Adopción de actitudes científicas* e *Interés en las experiencias de aprendizaje de las ciencias*, todos los ítems correspondientes, es decir, del 7.15 al 7.23 obtuvieron puntajes iguales o superiores a .80 en el criterio de *precisión*. En la subescala de *Interés en el aprendizaje de las ciencias extraescolares o en tiempo libre*, los ítems 7.24 y 7.25 obtuvieron puntajes de 0.75 y 0.72,

respectivamente. Las siguientes cuatro subescalas, con excepción del ítem 7.35 de la subescala de *Actitudes anticientíficas*, obtuvieron puntuaciones iguales o mayores a 0.80 en el criterio de *claridad*.

En el criterio de simplicidad, los ítems 7.2 y 7.4 de la subescala de *Implicaciones sociales de la ciencia* obtuvieron un PAJ de 0.78 y 0.64, mientras que los ítems 7.5, 7.8 y 7.9 de la subescala de *Normalidad de los científicos* obtuvieron puntajes de 0.64, 0.64, 0.72, respectivamente. Asimismo, el ítem 7.11 de la subescala de *Actitudes hacia la investigación* obtuvo un puntaje de acuerdo de 0.72. Las subescalas de *Adopción de actitudes científicas* e *Interés en experiencias de aprendizaje relacionada con las ciencias* obtuvieron puntajes iguales o mayores a 0.80. Mientras que en la subescala de *Interés en actividades relacionadas con la ciencia extraescolares o en tiempo libre*, los ítems 7.25 y 7.28 obtuvieron ambos una puntuación de 0.78. Las subescalas subsecuentes obtuvieron puntajes favorables, con excepción del ítem 7.35 de la subescala *Actitudes anticientíficas* que obtuvo una puntuación de 0.75.

Por último, los promedios generales que brindan un panorama más amplio del porcentaje de acuerdo de los jueces arrojaron lo siguiente: el ítem 7.4 de la subescala de *Implicaciones sociales de la ciencia* obtuvo un puntaje general de 0.72, mientras que los ítems 7.5, 7.8 y 7.9 obtuvieron puntajes de 0.63, 0.78 y 0.69, respectivamente. Los ítems 7.11 y 7.14 de la subescala de *Actitud hacia la investigación* obtuvieron puntajes generales de 0.67 y 0.74. Asimismo, los ítems 7.24 y 7.25 de la subescala de *Interés en experiencias de aprendizaje relacionadas con la ciencia* obtuvieron puntajes de 0.76 y 0.78. Por último, el ítem 7.35 de la subescala de *Actitudes anticientíficas* obtuvo un puntaje general de 0.69.

Tabla 15.*Resultados del cálculo del IVC de la EAC por ítem*

Ítem	V-Aiken	RVC Lawshe	RVC' Tristán-López
7.1	0.95	0.75	0.88
7.2	0.83	0.25*	0.63
7.3	0.94	0.67	0.83
7.4	0.72	-0.21*	0.40
7.5	0.63*	-0.29*	0.35*
7.6	0.84	0.17*	0.58*
7.7	0.88	0.38*	0.69
7.8	0.78	0.04*	0.52*
7.9	0.69*	-0.13*	0.44
7.10	0.84	0.38*	0.69
7.11	0.67*	-0.13*	0.44
7.12	0.81	0.17*	0.58*
7.13	0.86	0.33*	0.67
7.14	0.74	0.13*	0.56*
7.15	0.94	0.83	0.92
7.16	0.94	0.75	0.88
7.17	0.97	0.88	0.94
7.18	0.83	0.25*	0.63
7.19	0.96	0.83	0.92
7.20	0.97	0.83	0.92
7.21	0.85	0.42*	0.71
7.22	0.90	0.54*	0.77
7.23	0.91	0.67	0.83
7.24	0.76	0.04*	0.52*
7.25	0.78	0.17*	0.58*
7.26	0.86	0.50*	0.75
7.27	0.86	0.33*	0.67
7.28	0.86	0.42*	0.71
7.29	0.93	0.71	0.85
7.30	0.94	0.75	0.88
7.31	0.88	0.46*	0.73
7.32	0.85	0.29*	0.65
7.33	0.85	0.46*	0.73
7.34	0.94	0.75	0.88
7.35	0.69*	-0.17*	0.42*
7.36	0.87	0.29*	0.65
7.37	0.88	0.63	0.81
7.38	0.89	0.67	0.83
7.39	0.94	0.63	0.81
7.40	0.92	0.54*	0.77
7.41	0.94	0.67	0.83
7.42	0.91	0.67	0.83
7.43	0.99	0.92	0.96
7.44	0.94	0.75	0.88
7.45	0.92	0.67	0.83
7.46	0.86	0.54*	0.77

Nota. *V de Aiken < 0.69; RVC de Lawshe <0.56; RVC' <0.58

En cuanto a los resultados del cálculo de la V de Aiken, se encontró que los ítems 7.5 y 7.11 de las subescalas de *Normalidad de los científicos* y *Actitud hacia la investigación* respectivamente, fueron los únicos que no obtuvieron resultados favorables de acuerdo con el estándar (≥ 0.69) propuesto por

Aiken (1985). Sin embargo, se encontró que el ítem 7.9 de la subescala de *Normalidad de los científicos* y el ítem 7.35 de la subescala de *Actitudes anticientíficas* obtuvieron puntuaciones límite de 0.69.

Asimismo, en los resultados del cálculo del RVC de Lawshe (1975) se encontró que 27 de 46 ítems obtuvieron un puntaje menor a 0.56, parámetro sugerido por el autor. Se reportan los resultados de igual manera por subescala comenzando por la subescala de *Implicaciones sociales de la ciencia* en donde el ítem 7.2 y 7.4 obtuvieron puntuaciones de 0.25 y -0.21. En la subescala de *Normalidad de los científicos*, todos los ítems pertenecientes (7.5 a 7.9) obtuvieron puntuaciones bajas con -0.29, 0.17, 0.38, 0.04 y -0.13, respectivamente. Un caso similar sucedió con la subescala de *Actitud hacia la investigación*, en donde todos sus ítems (7.10 a 7.14) obtuvieron puntuaciones menores al parámetro establecido con 0.38, -0.13, 0.17, 0.33 y 0.13. Por otro lado, en la subescala de *Adopción de actitudes científicas*, el único ítem con puntuación baja fue el 7.18 con 0.25. En la subescala de *Interés en las experiencias de aprendizaje relacionadas con la ciencia* los ítems 7.21 y 7.22 obtuvieron puntuaciones de 0.42 y 0.54. Por otro lado, en la subescala de *Interés en actividades relacionadas con la ciencia extraescolares o en tiempo libre*, todos los ítems (7.24 a 7.28) resultaron bajos a comparación del estándar con puntuaciones de 0.04, 0.17, 0.50, 0.33 y 0.42.

En cuanto a la subescala de *Interés por una especialización o trabajo relacionado a las ciencias*, los ítems 7.31, 7.32 y 7.33, resultaron bajos con puntuaciones de 0.46, 0.29 y 0.46. Posteriormente, en la subescala de *Actitudes anticientíficas* los ítems 7.35 y 7.36 obtuvieron puntuaciones de -0.17 y 0.29, respectivamente. Por último, en la subescala de *Creencias conspiranoicas* resalta el ítem 7.40 con una puntuación de 0.54 y en la subescala de *Creencias paranormales*, el ítem 7.46 de igual manera con 0.54.

En cuanto al cálculo del RVC' de Tristán-López (2008) se encontró que cuatro ítems no cumplieron con el parámetro de 0.58 propuesto por dicho autor. De igual manera se reportan aquellos ítems con puntuación límite. De la subescala de *Normalidad de los científicos*, los ítems 7.5, 7.6 y 7.8

obtuvieron puntuaciones de 0.35, 0.58 y 0.52, respectivamente. El ítem 7.12 y 7.14 de la subescala de *Actitud hacia la investigación* tuvieron índices de 0.58 y 0.56, mientras que los ítems 7.24 y 7.25 de la subescala de *Interés en el aprendizaje de las ciencias extraescolares o en tiempo libre* obtuvieron puntuaciones de 0.52 y 0.58. Por último, el ítem 7.35 de la subescala de *Actitud anticientífica* obtuvo un valor de 0.42.

4.2.3. Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC)

En la **Tabla 16** se muestran los resultados del cálculo del Porcentaje de Acuerdo entre Jueces (PAJ \geq .80). Es decir, el promedio de acuerdo entre jueces por criterio (relevancia, simplicidad, claridad y precisión), por ítem y global por escala.

Tabla 16
Porcentaje de acuerdo entre jueces especialistas (PAJ) de la EAPAMC

Ítem	Relevancia	Precisión	Claridad	Simplicidad	Promedio de reactivo
8.1	0.97	0.81	0.75*	0.89	0.85
8.2	0.97	0.83	0.83	0.86	0.87
8.3	0.97	0.92	0.92	0.92	0.93
8.4	0.97	0.89	0.94	0.92	0.93
8.5	1.00	0.92	0.83	0.86	0.90
8.6	0.97	0.92	0.92	0.92	0.93
8.7	0.97	0.92	0.89	0.83	0.90
8.8	1.00	0.94	0.94	1.00	0.97
8.9	0.89	0.89	0.86	0.92	0.89
8.10	0.89	0.89	0.86	0.89	0.88
8.11	0.92	0.83	0.83	0.89	0.87
8.12	1.00	0.94	0.86	0.89	0.92
8.13	0.94	0.89	0.92	1.00	0.94
8.14	1.00	1.00	1.00	1.00	1.00
8.15	0.97	0.97	0.89	0.92	0.94
8.16	1.00	0.94	0.92	0.94	0.95
8.17	1.00	0.94	0.94	0.97	0.97

Nota. *PAJ < 0.80

En el caso de la EAPAMC, solo el ítem 8.1 de la subescala de *Explicar un fenómeno científicamente* obtuvo una puntuación inferior a 0.80 en el criterio de *claridad*. Los ítems restantes obtuvieron puntajes favorables incluso con acuerdos cerrados con puntuación general de 1.00 como es el caso del ítem 8.14.

Tabla 17.

Resultados del cálculo del IVC de la EAPAMC por ítem

Ítem	V-Aiken	RVC Lawshe	RVC' Tristán-López
8.1	0.85	0.46*	0.73
8.2	0.88	0.50*	0.75
8.3	0.93	0.67	0.83
8.4	0.93	0.75	0.88
8.5	0.90	0.63	0.81
8.6	0.93	0.71	0.85
8.7	0.90	0.54*	0.77
8.8	0.97	0.83	0.92
8.9	0.89	0.67	0.83
8.10	0.88	0.42*	0.71
8.11	0.87	0.54*	0.77
8.12	0.92	0.63	0.81
8.13	0.94	0.75	0.88
8.14	1.00	1.00	1.00
8.15	0.94	0.67	0.83
8.16	0.95	0.71	0.85
8.17	0.97	0.79	0.90

Nota. *V de Aiken < 0.69; RVC de Lawshe <0.56; RVC' <0.58.

En el caso de la EAPAMC, todos los ítems cumplieron el estándar de ≥ 0.69 propuesto por Aiken (1985). Del cálculo de la V de Aiken se obtuvieron resultados favorables, e incluso acuerdos cerrados como el ítem 8.14. En los resultados del RVC de Lawshe (1975) el ítem 8.1, 8.2 y 8.7 de la subescala de *Explicar un fenómeno científicamente* obtuvieron puntajes de .46 y .50, respectivamente, puntuaciones inferiores al estándar de ≥ 0.56 propuesto por dicho autor. Asimismo, los ítems 8.10 y 8.11 de la subescala de *Diseñar y evaluar investigación científica*, resultaron con puntuaciones de 0.42 y 0.54. Por último, en el cálculo del RVC' de Tristán-López (2008), todos los ítems de la EAPAMC obtuvieron puntajes favorables superiores al estándar de ≥ 0.58 .

4.3. Análisis de la imparcialidad por escala

En cuanto a los resultados de la atención del estándar 3.4 de los Estándares para las pruebas psicológica y educativas de la APA, AERA y NCME (2018), los comentarios de los jueces en los ítems de las tres escalas se enfocaron principalmente en los criterios de *comprensibilidad*, *definición precisa del constructo* e *inclusión poblacional*. En la sección de datos de identificación, los jueces comentaron la posibilidad de agregar más opciones de respuesta al apartado de “sexo” y a la cantidad de semestres de licenciatura con la finalidad de contemplar carreras que duran más de cuatro años.

4.3.1. Escala de Motivación en el Aprendizaje de las Ciencias

En la **Tabla 18** se muestra una tabla resumen de las observaciones del CVE por criterio de la EDU y por ítem.

Tabla 18

Numero de comentarios por ítem y por criterio de la EDU en la Escala de Motivación en el Aprendizaje de las Ciencias (EMAC)

ÍTEM	DEF	INP	AI	AFC	PIS	COMP	ML	TOTAL
6.1		1				4		5
6.2		2				4		6
6.3						3		3
6.4	1					1		2
6.5	1	1				1		3
6.6						3		3
6.7						2		2
6.8						4		4
6.9						3		3
6.10		2				1		2
6.11	1	1				2		4

6.12	2				2
6.13		1			1
6.15			1	1	2
6.16				3	3
6.17				2	2
6.18				1	1
6.19				2	2
6.20				1	1
6.21		1		1	2
6.22	1			1	2
6.23				1	1
6.24				2	2
6.25				2	2
6.26				6	6
6.29			1	1	2

Nota: Definición precisa del constructo (DEF); Inclusión Poblacional (INP); Accesibilidad e Imparcialidad (AI); Acomodo flexible de los contenidos (AFC); Procedimientos e instrucciones simples (PIS); Comprensibilidad (COMP) y Máxima Legibilidad (ML).

Se observa en la **Tabla 18**, el concentrado de los ítems que recibieron comentarios por parte de los jueces, se especifica también cuántos comentarios recibió cada ítem y a qué categoría pertenece dicho comentario. De acuerdo con los datos de la tabla, se observa una participación relativamente homogénea del panel de expertos. La mayoría ofrece sus opiniones respecto de los ítems que conforman la escala, sin que existan preferencias excesivamente claras. Sin embargo, en algunos bloques (del 12 al 15 y del 21 al 25) un número menor de jueces decidió emitir valorizaciones de los ítems.

Dados los resultados de la recopilación de evidencias de validez de constructo en el aspecto del contenido, los ítems 6.2, 6.5 y 6.17 mostraron los índices más bajos. Asimismo, se observó que el ítem 6.2 de la subescala de *Motivación intrínseca*, “*Me siento motivado para concretar las actividades que*

solicitan en la clase de metodología de la investigación (por ejemplo, una revisión de la literatura sobre un tema de mi interés)” atrajo observaciones del panel de expertos. Algunas de las observaciones fueron hacia la ambigüedad de la palabra “concretar”. Se sugiere modificarla por la palabra “realizar”. También se debatió la pertinencia del ítem hacia estudiantes de nivel superior que, al momento de responder la escala, no estén cursando materias relacionadas con Metodología de la investigación científica. Asimismo, el ítem 6.5 de la subescala de *Motivación Intrínseca*, recibió una propuesta de mejora a “*considero importante interpretar el mundo en términos científicos*” para hacerlo más comprensible y menos redundante en comparación con el ítem 6.4.

En la subescala de *Autoeficacia*, los comentarios de los jueces abordaron los ítems 6.16 al 6.26 y el 6.29 en torno a los criterios de comprensibilidad, definición precisa del constructo, inclusión poblacional y máxima legibilidad. Un ejemplo de las propuestas de modificación a la subescala es agregar el ítem “Puedo identificar claramente las variables comprendidas en los objetivos de investigación”.

4.3.2. Escala de Actitud hacia la Ciencia (EAC)

En la **Tabla 19** se muestra una tabla resumen de las observaciones del CVE por criterio de la EDU y por ítem.

Tabla 19

Número de comentarios por ítem y por criterio de la EDU en la Escala de Actitudes hacia la Ciencia (EAC)

ÍTEM	DEF.	IP	AI	AFC	PIS	COMP	ML	TOTAL
7.1						3		3
7.2						2		2
7.3						1		1
7.4						3		3
7.5	5							5
7.6								0
7.7						2		2
7.8						1		1
7.9	1					2		3
7.10	1					1		2
7.11	2					3		5
7.12	1	1				1		3
7.13	1						1	2
7.14	2	1				2		5
7.15						1		1
7.16						1		1
7.17		1				1		2
7.18	1					2		3
7.19		1						1
7.20		1						1
7.21		1		1				2
7.22	1							1
7.23		1				1		2
7.24		4						4
7.25		2				3		5
7.26						1		1
7.27						1		1
7.28		1				2	1	4
7.29	1					1		2
7.30								0
7.31	2					1		3
7.32	1							1
7.33						1		1
7.34								0
7.35				1		2		3
7.36	2					1		3
7.37							1	1
7.38	3							3
7.39								0
7.40	1							1
7.41	1							1
7.42						1		1
7.43								0
7.44					1	2		3
7.45						1		1
7.46	1			2	2	1		6
7.47								0
7.48								0
7.49								0

Nota: Definición precisa del constructo (DEF); Inclusión Poblacional (INP); Accesibilidad e Imparcialidad (AI); Acomodo flexible de los contenidos (AFC); Procedimientos e instrucciones simples (PIS); Comprensibilidad (COMP) y Máxima Legibilidad (ML).

En cuanto a los resultados de la recopilación de evidencias de validez de constructo en el aspecto de la imparcialidad de las escalas, los comentarios de los jueces en los ítems de la Escala de Actitudes hacia la Ciencia (EAC) se enfocaron principalmente en los criterios de comprensibilidad, definición precisa de constructo e inclusión poblacional.

En la tabla se observa que, en el caso de la EAC los comentarios de los jueces parecen formar dos segmentos: el primero (Jueces del 1 al 5) comentó en casi todos los ítems de la escala, salvo por algunos reactivos; el segundo segmento (Jueces 11 y 12) concentró sus observaciones en aquellos ítems sobre los que el primer segmento no emitió opiniones. Esta dinámica generó que los comentarios del panel abarquen prácticamente todos los reactivos de la escala.

De la subescala de *Implicaciones sociales de la ciencia*, los ítems 7.1, 7.2, 7.3 y 7.4 recibieron comentarios por parte de los jueces en torno al criterio de comprensibilidad. En el caso de todos los ítems de esta subescala, el tópico general que se abordó así como las sugerencias de modificación, radicaron principalmente en el retirar la frase “Pienso que” de todos los ítems. En el caso particular del 7.1 “Pienso que el desarrollo científico trae múltiples beneficios a la sociedad”, se propone acotarlo a “El desarrollo científico trae múltiples beneficios a la sociedad”.

En el caso de la subescala de *Normalidad de los científicos* de la EAC, los comentarios de los jueces se concentraron en los ítems 7.5, 7.7, 7.8 y 7.9 en torno a los criterios de definición precisa del constructo y comprensibilidad. De manera general, los jueces consideraron la necesidad de clarificar o modificar la palabra “normal” del ítem 7.7 “Las personas que se dedican a la ciencia tienen un estilo de vida normal”, de igual manera con el ítem “Las personas que se dedican a la ciencia son muy excéntricas” se propone modificar la palabra “excéntricas” debido a la variedad de interpretaciones que los respondientes podrían evocar al respecto de este término.

Posteriormente, en la subescala de *Actitud hacia la investigación*, el panel de jueces realizó observaciones de los ítems 7.10, al 7.14 en cuanto a los criterios de comprensibilidad, definición precisa

del constructo, máxima legibilidad e inclusión poblacional. Un ejemplo de propuesta por parte de los jueces es el ítem 7.11 “Si buscara como mejorar mis hábitos de estudio preferiría buscar información científica que consultar otras fuentes de información”, de esta manera dos de los cambios que se sugiere para modificarlo es “Las mejores fuentes de información son las que provienen de la investigación científica” o “Para mejorar mis hábitos de estudio prefiero buscar...”. Otras observaciones en esta subescala radican en considerar la deseabilidad social de los estudiantes al momento de responder la escala.

De la subescala de *Adopción de actitudes científicas*, los comentarios se centraron en los ítems 7.15, 7.16, 7.17 y 7.18 en torno a los criterios de comprensibilidad e inclusión poblacional. En esta subescala se propusieron algunas modificaciones en los reactivos. Por mencionar algunos ejemplos, se propone que el ítem 7.15 “Dudo de la información que consulto si no muestra evidencia científica que la sustente” se modifique por “Desconfío de la información que carece de respaldo científico” con la finalidad de que el ítem sea más claro y directo. Otro ejemplo de propuesta de modificación es el ítem 7.17 “Me gusta cuestionar y corroborar la información que se me brinda en clases” por “Cuestiono y corroboro la información que se me brinda en clases”.

En cuanto a la subescala de *Interés en las experiencias de aprendizaje de las ciencias*, los ítems que tuvieron comentarios por parte de los jueces fueron del 7.19, al 7.23 en los criterios de inclusión poblacional, definición precisa de constructo y comprensibilidad. Dichos comentarios se enfocan sobre todo en los ítems que contienen referencia al curso de Metodología de la investigación, por lo tanto, se sugiere modificarlo o delimitar las carreras a las que se les aplicará la escala. Esto último con la finalidad de que cada respondiente tenga o haya tenido experiencia con este tipo de materias.

De la subescala de *Interés en el aprendizaje de las ciencias extraescolares o en tiempo libre*, el panel de jueces comentó los ítems 7.24, 7.25, 7.26, 7.27 y 7.28 en torno a los criterios de inclusión poblacional, comprensibilidad y máxima legibilidad. Las observaciones dadas en el ítem 7.24 “Me

interesan los videos en Youtube que tratan sobre temas científicos diversos” se enfocaron sobre todo en no ser tan específico al momento de mencionar una plataforma o red social. Se sugiere utilizar términos más generales como “en internet” o en “redes sociales”. Por mencionar otro ejemplo, se propone una modificación para el ítem 7.27 “Me gusta leer libros sobre temas científicos ya sea en formato análogo o digital” a “Me gusta leer sobre temas científicos, ya sea en libros, revistas, o en fuentes digitales” con la finalidad de que el ítem sea mucho más comprensible para la población objetivo.

En cuanto a la subescala de *Interés en una especialización o trabajo relacionado con las ciencias*, el panel de jueces dio observaciones a los ítems 7.29 al 7.34 sobre todo en los criterios de definición precisa del constructo y comprensibilidad. En esta subescala los comentarios fueron más dirigidos hacia la necesidad de esclarecer la diferencia entre lo que mide cada ítem con la finalidad de evitar que suene redundante. Posteriormente, en la subescala de *Actitudes anticientíficas*, los jueces dieron observaciones en los ítems 7.35, 7.36, 7.37 y 7.38 en los criterios de definición precisa del constructo y comprensibilidad. En una de las discusiones sobre esta subescala, los jueces sugieren agregar además de las actitudes mencionadas, otras actitudes anticientíficas más complejas. Asimismo, una propuesta de modificación para el ítem 7.37 “Prefiero utilizar mi intuición antes que aplicar procedimientos científicos para resolver problemas de la vida cotidiana” por “Es imposible conocer la realidad con los métodos rígidos de la ciencia”.

De la subescala de *Creencias conspiranoicas*, los comentarios de los jueces se enfocaron en los ítems 7.40 y 7.41 en torno al criterio de definición precisa del constructo. Los comentarios al respecto reflejaron que los ítems de esta subescala podrían contener afirmaciones parcialmente verdaderas. Por último, en la subescala de *Creencias paranormales*, las observaciones del panel de expertos se concentraron en los ítems 7.42, 7.44, 7.45 y 7.46 en los criterios de procedimientos e instrucciones simples y comprensibilidad. Se sugiere en cuanto al primer criterio, reiterar los encabezados de las opciones de respuesta. Un ejemplo de las modificaciones que proponen los jueces en esta subescala es el

ítem 7.45 “Existen personas con la habilidad para comunicarse con otras personas por medio de la mente” a “Existen personas que se pueden comunicar utilizando solo su mente” con la finalidad de hacer el ítem más claro y comprensible.

4.3.3. Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC)

En la **Tabla 20** se muestra una tabla resumen de las observaciones del CVE por criterio de la EDU y por ítem.

Tabla 20.

Número de comentarios por ítem y por criterio de la EDU en la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC).

ÍTEM	DEF	IP	AI	AFC	PIS	COMP	ML	TOTAL
8.1						1		1
8.2							1	1
8.3								0
8.4						1		1
8.5								0
8.6								0
8.7	1					3		4
8.8								0
8.9								0
8.10								0
8.11								0
8.12						2		2
8.13						1		1
8.14								0
8.15						2	1	3
8.16	1							1
8.17	1	1						2

Nota: Definición precisa del constructo (DEF); Inclusión Poblacional (INP); Accesibilidad e Imparcialidad (AI); Acomodo flexible de los contenidos (AFC); Procedimientos e instrucciones simples (PIS); Comprensibilidad (COMP) y Máxima Legibilidad (ML).

En cuanto a la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC) los comentarios de los jueces se concentraron en los criterios de definición precisa, comprensibilidad y máxima legibilidad. La EAPAMC es en general la escala menos comentada; particularmente la subescala de *Autoeficacia para Diseñar y Evaluar Investigación* atrajo muy pocas observaciones. Asimismo, la EAPAMC es la escala que obtuvo mejores resultados en la recopilación de evidencias de validez de constructo en el aspecto del contenido.

De la subescala *Explicar un fenómeno científicamente*, los ítems 8.1, 8.2, 8.4 y 8.7 tuvieron observaciones por parte de los jueces en torno a los criterios de comprensibilidad, definición precisa del constructo y máxima legibilidad. Por mencionar algunos ejemplos, en el ítem 8.1 “Ofrecer explicaciones científicas para fenómenos naturales y tecnológicos”, la sugerencia fue clarificar el concepto de “fenómeno tecnológico” ya que podría resultar confuso para los respondientes.

Por otro lado, en el ítem 8.7 “Explicar apropiadamente las implicaciones del conocimiento científico para la sociedad” las sugerencias tienen que ver con los criterios de definición precisa del constructo y comprensibilidad. En el caso de la definición precisa del constructo la observación radica en que el ítem y el constructo no pertenecen al tipo de escala desarrollada, sin embargo, por retomar de manera íntegra los indicadores de la prueba PISA se decidió conservar el ítem. En cuanto a la comprensibilidad, se sugirió retirar el adverbio “apropiadamente” con la finalidad de hacer el ítem más corto y comprensible.

De la subescala *Diseñar y evaluar investigación científica*, el ítem 8.12 “Reconocer las formas en las que los científicos garantizan la objetividad de las investigaciones científicas” las observaciones fueron en torno al criterio de comprensibilidad. En este ítem se sugiere evitar la redundancia de los términos “científicos” y “científicas”. Por otro lado, de la subescala *Interpretar datos y evidencia científica*, en los ítems 8.15, 8.16 y 8.17 los comentarios emitidos por parte de los jueces fueron en torno a los criterios de máxima legibilidad, comprensibilidad e inclusión poblacional. Un ejemplo de ello es el ítem 8.16 “Reconocer la evidencia científica de otras fuentes de información”, en este reactivo las sugerencias fueron con respecto al criterio de definición precisa del constructo. Ejemplos de estos comentarios son los siguientes: “8.16 y 8.17 parecen estar al revés, según los indicadores”, “el 8.16 debería enfocarse en distinguir entre argumentos”. Por último, el ítem 8.17 “Distinguir argumentos que se basan en evidencia científica” recibió comentarios referentes a la inclusión poblacional del ítem en referencia a la población objetivo de la escala. Se sugirió, por ejemplo, extender la aplicación de la escala a estudiantes de Educación Media Superior.

Una vez integrados los comentarios de los jueces y los resultados del cálculo del IVC por escala, se observó que algunos ítems atrajeron más la atención de los jueces, en cuanto a propuestas de modificación y bajas puntuaciones en el *Protocolo de validación de contenido de las escalas*. Por este motivo, se presenta en la **Tabla 21** y **22** los ítems que, de acuerdo con el proceso de recopilación de evidencias de validez requieren de cambios mayores.

Tabla 21.
Ítems que requieren modificaciones mayores de la EMAC

Subescala	Ítem	Propuesta de modificación	Categoría
Motivación Intrínseca	6.2. Me siento motivado para concretar las actividades que solicitan en la clase de metodología de la investigación (por ejemplo, una revisión de la literatura sobre un tema de mi interés).	6.2. Me siento motivado a realizar las actividades que solicitan en la clase de Metodología de la investigación (por ejemplo, una revisión de la literatura sobre un tema de mi interés).	Comprensibilidad
	6.5. Es interesante interpretar el mundo en términos científicos.	6.5. Considero importante conocer el mundo desde una perspectiva científica.	Comprensibilidad
Autoeficacia	6.17. Puedo ingresar variables y datos en un programa estadístico.	6.17. Puedo identificar las variables de un estudio en los objetivos de investigación.	Comprensibilidad

Tabla 22.
Ítems que requieren modificaciones mayores de la EAC

Subescala	Ítem	Propuesta de modificación	Categoría
Implicaciones sociales de la ciencia	7.4. Pienso que un rápido progreso científico y tecnológico puede traer más desventajas que ventajas para la sociedad.	7.4. El progreso científico y tecnológico acelerado puede traer más ventajas que desventajas para la sociedad.	Comprensibilidad
Normalidad de los científicos	7.5. Las personas que se dedican a la ciencia tienen un estilo de vida normal.	7.5. Las personas que se dedican a la ciencia tienen un estilo de vida similar al de cualquier otro profesionalista.	Definición precisa del constructo
	7.9. Las personas que se dedican a la ciencia son muy excéntricas.	7.9. Las personas que se dedican a la ciencia son muy peculiares.	Comprensibilidad
Actitud hacia la investigación	7.11. Si buscara como mejorar mis hábitos de estudio preferiría buscar información científica que consultar otras fuentes de información.	7.11. Para mejorar mis hábitos de estudio prefiero buscar información científica que consultar otro tipo de fuentes.	Comprensibilidad
	7.12. Me gusta aprender sobre cómo las personas que se dedican a la ciencia diseñan sus experimentos.	7.12. Me interesa aprender cómo diseñan sus estudios las personas que se dedican a la ciencia.	Definición precisa del constructo
	7.14. Siento curiosidad por fenómenos naturales como el	7.14. Siento curiosidad por comprender las explicaciones científicas de fenómenos naturales	Inclusión poblacional

	cambio climático, o sociales como el comportamiento de las personas.	como el cambio climático, o sociales como el comportamiento de las personas.	
Interés en las experiencias de aprendizaje de las ciencias extraescolares o en tiempo libre	7.24. Me interesan los videos en Youtube que tratan sobre temas científicos diversos.	7.24. Me interesan los videos en redes sociales que tratan sobre temas científicos diversos.	Inclusión poblacional
	7.25. Cuando tengo tiempo, me gusta ver documentales de temas científicos (por ejemplo, sobre los avances de las neurociencias).	7.25. Me gusta ver documentales de temas científicos.	Inclusión poblacional
Actitudes anticientíficas	7.35. No confío en los métodos que usa la ciencia.	7.35. No confío en los métodos que usa la ciencia.	Comprensibilidad

Las modificaciones de los ítems, resultado de la recopilación de evidencias de validez de constructo en el aspecto de la imparcialidad y del contenido de los instrumentos de medición se concentraron principalmente en las escalas EMAC y EAC. De la subescala de *Motivación intrínseca*, el ítem 6.2 “Me siento motivado para concretar las actividades que solicitan en la clase de metodología de la investigación (por ejemplo, una revisión de la literatura sobre un tema de mi interés)”, se modificó a “Me siento motivado a realizar las actividades que solicitan en la clase de Metodología de la investigación (por ejemplo, una revisión de la literatura sobre un tema de mi interés)”, atendiendo la sugerencia de modificar el verbo “concretar” por “realizar”, así como cuestiones de gramática.

El ítem 6.5 de la misma subescala “Es interesante interpretar el mundo en términos científicos”, se modificó a “Considero importante conocer el mundo desde una perspectiva científica”, atendiendo las observaciones en cuanto a mejorar la comprensibilidad y simplicidad del ítem. Asimismo, el ítem 6.17 de la subescala de *Autoeficacia* “Puedo ingresar variables y datos en un programa estadístico” a “Puedo identificar las variables de un estudio en los objetivos de investigación”.

De la escala EAC, se realizaron modificaciones en varias de sus subescalas, comenzando por el ítem 7.4 de la subescala de *Implicaciones sociales de la ciencia* “Pienso que un rápido progreso científico y tecnológico puede traer más desventajas que ventajas para la sociedad” a “El progreso científico y

tecnológico acelerado puede traer más ventajas que desventajas para la sociedad”, atendiendo las observaciones de retirar la primera parte del ítem “pienso que”, así como las sugerencias de comenzar por las ventajas en lugar de las desventajas.

Una de las subescalas que atrajo más la atención de los jueces fue la de *Normalidad de los científicos*. De esta subescala se realizaron dos modificaciones, el ítem 7.5 “Las personas que se dedican a la ciencia tienen un estilo de vida normal”, se corrigió a “Las personas que se dedican a la ciencia tienen un estilo de vida similar al de cualquier otro profesionalista”, con la finalidad de evitar utilizar la palabra “normal” que fue la que suscitó más debate en el comité de jueces. De la misma subescala, el ítem 7.9 “Las personas que se dedican a la ciencia son muy excéntricas”, se modificó a “Las personas que se dedican a la ciencia son muy peculiares”, esto con el objetivo de no utilizar la palabra “excéntrico” que, de acuerdo con el panel de jueces podía generar confusión en los respondientes por la polisemia de la palabra.

De la subescala de *Actitud hacia la investigación*, se modificó el ítem 7.11 “Si buscara como mejorar mis hábitos de estudio preferiría buscar información científica que consultar otras fuentes de información” a “Para mejorar mis hábitos de estudio prefiero buscar información científica que consultar otro tipo de fuentes”, con la finalidad de evitar conjugaciones de confusas en los verbos. Por otro lado, el ítem 7.12 “Me gusta aprender sobre cómo las personas que se dedican a la ciencia diseñan sus experimentos” a “Me interesa aprender cómo diseñan sus estudios las personas que se dedican a la ciencia”. El ítem 7.14 “Siento curiosidad por fenómenos naturales como el cambio climático, o sociales como el comportamiento de las personas” a “Siento curiosidad por comprender las explicaciones científicas de fenómenos naturales como el cambio climático, o sociales como el comportamiento de las personas”.

De la subescala de *Interés en actividades relacionadas con la ciencia extraescolares o de tiempo libre*, se modificó el ítem 7.24 “Me interesan los videos en Youtube que tratan sobre temas científicos

diversos” a “Me interesan los videos en redes sociales que tratan sobre temas científicos diversos”, así como el ítem 7.25 “Cuando tengo tiempo, me gusta ver documentales de temas científicos (por ejemplo, sobre los avances de las neurociencias)” a “Me gusta ver documentales de temas científicos”.

Por último, cabe aclarar que, aunque el ítem 7.35 de la subescala de *Actitudes anticientíficas* “No confío en los métodos que usa la ciencia” recibió varios comentarios, así como bajos puntajes en los índices. Una de las sugerencias para corregir el ítem era redactarlo en positivo, sin embargo, al ser una actitud anticientífica es motivo para utilizar la base del ítem “no confío” por lo que se decidió conservarlo tal y como está debido a la naturaleza del constructo.

Es relevante mencionar que las observaciones emitidas por el CVE se tomaron como guía para modificar los ítems con menor puntuación. Se atendieron las observaciones que apuntaran a mejorar la comprensibilidad y accesibilidad de los ítems, puesto que las escalas puntuaron favorablemente en cuestiones de relevancia. Para las modificaciones, se seleccionaron comentarios que se repitieran y que tuvieran propuestas claras y explícitas para mejorar el reactivo. Se repitió el mismo procedimiento en todos los ítems que tuvieron puntuaciones bajas, con excepción de aquellos que no tenían ningún comentario o propuesta de mejora o con aquellos que por fidelidad al constructo se decidió conservar tal y como estaba. Dicho esto, en la **Tabla 23** y **24** se muestran la EMAC y la EAC en sus versiones corregidas. Debido a que los ítems de la EAPAMC obtuvieron puntajes favorables, se decidió no realizar ningún cambio en los ítems hasta observar su comportamiento en la aplicación piloto.

Tabla 23.*Escala de Motivación en el Aprendizaje de las Ciencias (EMAC-II)*

Subescala de Motivación Intrínseca	
Afirmaciones	
6.1.	Me resulta interesante aprender el contenido que se enseña en las clases de Metodología de la investigación
6.2.	Me siento motivado a realizar las actividades que solicitan en la clase de Metodología de la investigación (por ejemplo, una revisión de la literatura sobre un tema de mi interés).
6.3.	Conocer hallazgos científicos es relevante para mi vida tanto dentro como fuera de la escuela.
6.4.	Disfruto leer sobre hallazgos científicos, aunque no sean parte de mis actividades de clase.
6.5.	Considero importante conocer el mundo desde una perspectiva científica.
6.6.	Siento interés por conocer hallazgos científicos en diversas áreas del conocimiento
6.7.	Siento interés por aprender teorías y explicaciones científicas asociadas a mi profesión.
Subescala de Motivación Extrínseca	
6.8.	Adquirir habilidades científicas me será útil para ingresar a un posgrado en un futuro.
6.9.	Adquirir habilidades científicas me permitirá obtener un mejor empleo al salir de la carrera.
6.10.	Aprender a redactar una propuesta de investigación es una habilidad que me servirá al salir de la universidad.
6.11.	Aprender a utilizar programas estadísticos para analizar datos me dará una ventaja profesional en el futuro.
6.12.	Dedicarme a la ciencia me permitirá tener estabilidad económica en el futuro.
Subescala de Autodeterminación	
6.13.	Cuando no entiendo alguna parte del proceso de investigación, busco a alguien que me ayude a resolver mis dudas.
6.14.	Me esfuerzo lo suficiente para comprender los contenidos de la clase de metodología de la investigación.
6.15.	Si es necesario, invierto más tiempo en estudiar los contenidos de la clase de metodología de la investigación.
Subescala de Autoeficacia	
6.16.	Puedo realizar una propuesta de investigación como proyecto final de clase.
6.17.	Puedo identificar las variables de un estudio en los objetivos de investigación.
6.18.	Puedo utilizar un programa de análisis de datos para una práctica de clase.
6.19.	Puedo leer un artículo científico y explicar el contenido a la clase.
6.20.	Puedo diferenciar entre información científica y no científica en libros o contenidos de internet.
6.21.	Puedo recolectar datos con cuestionarios.
6.22.	Tengo la habilidad de buscar información científica confiable.
6.23.	Puedo identificar las partes de un artículo.
6.24.	Puedo redactar una propuesta de investigación para ingresar a un posgrado.
6.25.	Puedo explicar con mis propias palabras la metodología de un artículo científico.
6.26.	Puedo encontrar artículos científicos en bases de datos (por ejemplo, ERIC, EBSCO, Wiley, entre otros).
6.27.	Puedo seleccionar una encuesta o cuestionario apropiado para recolectar datos.
6.28.	Puedo realizar el diseño de un experimento sencillo.
6.29.	Puedo realizar una revisión de la literatura científica sobre un tema.

Nota. El tipo de respuesta es en formato de escala ordinal con cuatro categorías de respuesta: “Totalmente en desacuerdo, en desacuerdo, de acuerdo y totalmente de acuerdo”.

Tabla 24.*Escala de Actitudes hacia la Ciencia (EAC-II)*

Subescala de Implicaciones Sociales de la Ciencia	
Afirmaciones	
7.1.	Pienso que el desarrollo científico trae múltiples beneficios a la sociedad.
7.2.	Pienso que la ciencia es una buena forma de solucionar problemas de tipo natural y social.
7.3.	Pienso que el desarrollo de conocimiento científico contribuye a mejorar mi calidad de vida.
7.4.	El progreso científico y tecnológico acelerado puede traer más ventajas que desventajas para la sociedad.
Subescala de Normalidad de los Científicos	
7.5.	Las personas que se dedican a la ciencia tienen un estilo de vida similar al de cualquier otro profesionista.
7.6.	Las personas que se dedican a la ciencia no tienen descanso.
7.7.	Para dedicarse a la ciencia, se tiene que sacrificar otras áreas de la vida como la social y la familiar.
7.8.	Las personas que se dedican a la actividad científica suelen ser rígidas e inflexibles.
7.9.	Las personas que se dedican a la ciencia son muy peculiares.
Subescala de Actitud hacia la Investigación	
7.10.	Es mejor aplicar el método científico para entender el mundo que utilizar el sentido común o la intuición.
7.11.	Para mejorar mis hábitos de estudio prefiero buscar información científica que consultar otro tipo de fuentes.
7.12.	Me interesa aprender cómo diseñan sus estudios las personas que se dedican a la ciencia.
7.13.	Me gusta aprender sobre cómo las personas que se dedican a la ciencia generan y aplican sus ideas de investigación.
7.14.	Siento curiosidad por comprender las explicaciones científicas de fenómenos naturales como el cambio climático, o sociales como el comportamiento de las personas.
Subescala de Adopción de Actitudes Científicas	
7.15.	Dudo de la información que consulto si no muestra evidencia científica que la sustente.
7.16.	Aunque requiera de más tiempo, me gusta revisar si las fuentes de información que consulto son confiables.
7.17.	Me gusta cuestionar y corroborar la información que se me brinda en clases.
7.18.	Me gusta dar opiniones objetivas y basadas en datos científicos.
Subescala de interés en las experiencias de aprendizaje de las ciencias	
7.19.	Me gustan las materias en donde puedo aprender cómo se genera el conocimiento científico.
7.20.	Me gustaría que en mi carrera hubiera más materias para aprender cómo se genera el conocimiento científico.
7.21.	Me gusta la clase de metodología de la investigación científica.
7.22.	Me gusta aprender sobre las distintas fases del método científico.
7.23.	Me gustaría aprender sobre el proceso de investigación científica realizando una investigación como proyecto de clase.
Subescala de interés en el aprendizaje de las ciencias extraescolares o en tiempo libre	
7.24.	Me interesan los videos en redes sociales que tratan sobre temas científicos diversos.
7.25.	Me gusta ver documentales de temas científicos.
7.26.	Me gusta ver películas que tratan sobre la vida de científicos famosos o sobre descubrimientos relevantes de la ciencia.
7.27.	Me gusta leer libros sobre temas científicos ya sea en formato análogo o digital.
7.28.	Me gusta escuchar podcast (contenido auditivo con formato similar a la radio) sobre temas científicos diversos.
Subescala de Interés en una especialización o trabajo relacionado con las ciencias	
7.29.	Al finalizar la carrera, me gustaría entrar a un posgrado especializado en investigación científica.
7.30.	En un futuro, me gustaría aportar conocimiento científico a mi campo profesional.
7.31.	Consideraría dedicarme a la investigación científica en un futuro.
7.32.	Cuando termine mi carrera me gustaría involucrarme en algún proyecto de investigación científica.

- 7.33. Es importante seguir mi formación científica al terminar mi carrera.
- 7.34. Dedicarme a la investigación es una de mis opciones laborales en el futuro.

Subescala de Actitudes Anticientíficas

- 7.35. No confío en los métodos que usa la ciencia.
- 7.36. Confío más en mi sexto sentido para solucionar problemas que en las evidencias científicas.
- 7.37. Prefiero utilizar mi intuición antes que aplicar procedimientos científicos para resolver problemas de la vida cotidiana.
- 7.38. Confío en los métodos de la ciencia.

Subescala de Creencias Conspiranoicas

- 7.39. Los científicos nos ocultan la cura de enfermedades graves como el cáncer.
- 7.40. Los científicos no aceptan la medicina alternativa porque representa un peligro para los intereses de la industria farmacéutica.
- 7.41. Los científicos trabajan en secreto en cosas que podrían perjudicar a la humanidad.

Subescala de Creencias Paranormales

- 7.42. La posición de los planetas puede influir en la personalidad y el comportamiento de las personas.
- 7.43. Aunque la ciencia no lo acepte, hay formas de comunicarse con personas fallecidas.
- 7.44. Existen personas con la capacidad de atraer sucesos con la mente.
- 7.45. Existen personas con la habilidad para comunicarse con otras personas por medio de la mente.
- 7.46. Creo que puedo atraer situaciones favorables si me concentro lo suficiente.

Nota. El tipo de respuesta es en formato de escala ordinal con cuatro categorías de respuesta: “Nada en absoluto, Poco, Moderado y Fuerte”.

V. DISCUSIÓN Y CONCLUSIONES

En este capítulo se presenta una síntesis del logro de los objetivos de investigación y de las aportaciones a la luz de los hallazgos obtenidos, así como su discusión en contraste con los resultados de otras investigaciones en el campo de la psicometría, y la evaluación educativa. Aunado a ello, en el segundo subapartado de este capítulo se presentan las limitaciones del trabajo de investigación. Subsecuentemente, se desarrollan las recomendaciones del estudio, en especial las recomendaciones metodológicas y prácticas para investigaciones futuras.

5.1. Discusión de los resultados

El primer objetivo específico consistió en la revisión de la literatura sobre los antecedentes teóricos y metodológicos de los constructos a medir. El logro de este objetivo permitió desarrollar el apartado de antecedentes y fundamentos de la presente tesis, así como la Ficha Técnica o documento rector de las escalas. Asimismo, la revisión de la literatura y la definición clara del constructo formó parte de los requisitos indispensables del plan de diseño de instrumentos de acuerdo con las directrices de los Estándares para el desarrollo de pruebas educativas y psicológicas de la AERA, APA y NCME (2018).

El segundo objetivo específico consistió en el diseño y desarrollo de tres instrumentos de medición. Para el logro de dicho objetivo se conformó el Equipo de Diseño y Desarrollo de las Escalas (EDDE) para el desarrollo de las tablas de contenido y diseño de los ítems de las escalas. Los productos derivados del logro de este objetivo son el diseño y desarrollo de tres escalas: (1) la Escala de Motivación en el Aprendizaje de las Ciencias (EMAC), (2) la Escala de Actitud hacia la Ciencia (EAC) y (3) la Escala de Autoeficacia Percibida en el Método Científico (EAPAMC) para estudiantes de ES nivel licenciatura. La finalidad del diseño, desarrollo y validación de estas escalas fue contar con instrumentos de medición

que permitan explorar el dominio afectivo del aprendizaje de las ciencias con un enfoque integrador. Asimismo, estas escalas pueden ser aplicadas a estudiantes de licenciatura de carreras diversas dentro de las áreas de ciencias naturales y exactas, así como de las ciencias sociales, educativas y del comportamiento.

Otro producto relevante derivado del logro de este objetivo es la Ficha Técnica de las escalas o el documento rector de los instrumentos de medición. En este documento se atienden las sugerencias y directrices de los *Estándares para el desarrollo de pruebas psicológicas y educativas* de la AERA, APA y NCME (2018). De esta manera, tal y como se sugiere en los *Estándares* y en la realización de pruebas a gran escala como PISA y TALIS, en el diseño de estas escalas se realizó una revisión de la literatura para localizar los principales referentes de los constructos a medir, así como los modelos de medición e instrumentos consolidados. Dichos instrumentos fungieron como base para el diseño y desarrollo de las tablas de contenido, de la misma manera con los ítems de las escalas.

En el desarrollo de tablas de contenido, se atendió la sugerencia de los *Estándares* de asegurar la *accesibilidad de los ítems* con ayuda del *Modelo de Diseño Universal*. Por ello, se hizo uso de la Evaluación de Diseño Universal (EDU) por su traducción (Pérez-Morán, 2014) para el diseño y desarrollo de los ítems de las tres escalas. Asimismo, con la finalidad de atender la accesibilidad de los ítems, se consideraron aspectos contextuales de la población meta de la medición.

Como se mencionó en apartados anteriores, uno de los puntos observados en instrumentos recogidos en la revisión de la literatura es el uso común de palabras que hacen referencia a ciencias específicas. Un ejemplo de esto es la referencia a actividades como el uso de microscopio o de la tabla periódica de elementos en instrumentos de interés vocacional hacia la investigación científica. Dicho esto, el desarrollo de tres escalas con el uso de MUD, pretende que estos instrumentos sean aplicables a

estudiantes de distintos tipos de ciencias, sin importar el objeto específico de estudio de cada ciencia ya que parte desde un enfoque integral en donde se reconoce que las ciencias utilizan estrategias investigativas para la generación de conocimiento. Otro ejemplo de la aplicación del MUD es la relación de los reactivos con materias relacionadas a la Metodología de la Investigación en ES, así como actividades comunes en el aprendizaje de estos contenidos como la realización de un proyecto de investigación o una revisión de literatura científica. En contraste, se puede mencionar el TOSRA (Fraser, 1982) o el SDS de Holland et al. (2005), en el primero los ítems con relación al gusto por la investigación científica hacen referencia directa a la realización de experimentos, en el segundo los ítems relacionados a la personalidad del investigador hacen referencia al gusto por contenidos específicos de química, biología o astronomía.

Las tres escalas producto de este trabajo de tesis son: la Escala de Motivación en el Aprendizaje de las Ciencias (EMAC) con 29k divididos en 4 dimensiones: *motivación intrínseca*, *motivación extrínseca*, *autodeterminación* y *autoeficacia*. La Escala de Actitudes hacia la Ciencia (EAC) que consta de 46k divididos en 10 subescalas: implicaciones sociales de la ciencia, normalidad de los científicos, actitudes hacia la investigación, adopción de actitudes científicas, disfrute de las experiencias de aprendizaje de las ciencias, interés en el aprendizaje de las ciencias en tiempo de ocio, interés en una especialización en ciencias u opciones laborales relacionadas, actitudes anticientíficas, creencias conspiranoicas y creencias paranormales. Por último, la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC) con 17k integrados en tres subescalas: explicar un fenómeno científicamente, diseñar y evaluar investigación científica e interpretar datos y evidencia científicamente.

Como síntesis del diseño y desarrollo de estas escalas, se puede mencionar que las escalas tienen como referentes el SLMQ-II de Glynn et al. (2011), la definición de actitud hacia la ciencia de Klopfer

(1973) y PISA (2015), el TOSRA (Fraser, 1982), el concepto *Contaminated Mindware* de Rizeq et al. (2020) y la *Competencia Científica* de la prueba PISA (2015). El principal aporte del diseño y desarrollo de las escalas fue el introducir el constructo de *Contaminated Mindware*, compuesto por actitudes anticientíficas, creencias paranormales y creencias conspiranoicas para la recopilación de evidencias de validez divergentes. Con el diseño y desarrollo de estos instrumentos, se brinda al campo de la investigación educativa herramientas de medición para conocer el componente afectivo del aprendizaje de las ciencias en estudiantes de ES, en español y para población mexicana. Asimismo, por su formato, las escalas pueden utilizarse en evaluaciones en pequeña y gran escala.

El tercer objetivo específico de la presente tesis fue la recopilación de evidencias de validez de constructo en el aspecto del contenido de la EMAC, EAC y EAPAMC. Este objetivo se logró de manera satisfactoria con la realización de un reporte de resultados que contiene el cálculo de Índices de Validez de Contenido en relación con los criterios de relevancia, claridad, simplicidad y precisión de cada ítem y un puntaje global por escala. Este tipo de evidencia del contenido de los instrumentos de medición da más solidez a las observaciones de los jueces, ya que permite visualizar un porcentaje de acuerdo de la intersubjetividad de los jueces.

Este reporte es accesible a cualquier persona que requiera información sobre la calidad de cada ítem. Esto es relevante, ya que no es un procedimiento que realice de manera frecuente en los estudios psicométricos relacionados a los constructos centrales de la presente tesis. Lo anterior a diferencia del diseño, desarrollo y validación de instrumentos de medición para el área médica y de la salud, como se menciona en el trabajo de Polit y Beck (2006). En esta vía, es relevante mencionar que, debido a la ausencia de estudios psicométricos con reportes de IVC en escalas de constructos similares o de los mismos constructos, no es posible hacer una comparación directa de las puntuaciones. Hasta la fecha, el cálculo del IVC ha sido más frecuente en instrumentos desarrollados en el campo de la enfermería, por

lo que es necesario aumentar el uso de estos índices para la recopilación de evidencias de validez de contenido (Polit y Beck, 2006). En constructos más relacionado a los de interés de la presente tesis, se puede mencionar el diseño de un instrumento para evaluar la competencia matemática (García Perales, 2018) en donde se realizó un reporte psicométrico del cálculo del IVC por cada ítem similar a este trabajo de investigación.

No obstante, no se encontró una situación similar en el diseño de escalas del dominio afectivo en el aprendizaje de las ciencias ya que, en algunos casos, se realizó un jueceo para la recopilación de evidencias de validez de contenido, pero no necesariamente el cálculo del IVC. Algunos ejemplos de esta situación es el propio SMLQ-II (Glynn et al., 2011), la Escala de Actitud hacia la Ciencia de Molina et al. (2013), la Escala de Actitudes hacia la Investigación-EACIN (Aldana et al., 2020), el School Science Attitude Questionnaire (SSAQ) de Aguilera y Perales-Palacios (2019). Cabe mencionar que los instrumentos anteriormente mencionados recopilaron otros tipos de evidencias de validez igualmente importantes como el análisis exploratorio, confirmatorio y análisis de confiabilidad.

Sin embargo, la recopilación de evidencias de validez en el aspecto del contenido, así como de la imparcialidad, son de suma relevancia de acuerdo con los Estándares para el desarrollo de pruebas psicológicas y educativas. En esta vía, en el trabajo de Polit y Beck (2006) y Toma y Lederman (2020) se señala que la ausencia de reportes psicométricos de validez de contenido tiene como consecuencia que se desconozca la calidad y la exhaustividad de los ítems. En este punto es importante mencionar que las escalas se encuentran en una primera fase del proceso de validación, de acuerdo con los estándares para las pruebas educativas y psicológicas de la AERA, APA y NCME (2018). Es decir, los resultados del cálculo del IVC revelan información relevante y detallada de la calidad de los ítems de acuerdo con su contenido, por lo que esta información deberá integrarse a otras evidencias de validez para el correcto uso e interpretación de los resultados de la aplicación de las escalas. Cabe mencionar también que, debido

al énfasis en la recolección de evidencias de validez en torno al contenido y la imparcialidad, se decidió en etapas tempranas del trabajo de tesis enfocarse en la recolección de las evidencias anteriormente mencionadas debido a la ausencia de reportes de IVC detallados en instrumentos de medición relacionados a los constructos de interés.

En cuanto a los resultados de recopilación de evidencias de validez, hay suficiente evidencia para concluir que las tres escalas son relevantes para la medición de la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias en estudiantes de ES nivel licenciatura. Las tres escalas obtuvieron puntajes globales adecuados en el cálculo del Porcentaje de Acuerdo entre Jueces (PAJ), en la V-Aiken (1985), en el RVC de Lawshe clásico (1975) y en la corrección de Tristán-López (2008). En cuanto al cálculo del RVC de Lawshe tomando en cuenta todos los criterios, además de la relevancia, se observaron bajas puntuaciones en los ítems, por lo que se recomienda esperar el comportamiento de la escala en su aplicación en campo.

El uso de distintos índices de validez permitió la comparación entre resultados por cada ítem de cada escala. De acuerdo con Polit y Beck (2006) los investigadores utilizan estos reportes de índice de validez de contenido como una guía para la revisión, corrección o eliminación de ítems. En el presente trabajo, la comparación entre los resultados de los índices permitió recopilar las evidencias indispensables para hacer las correcciones necesarias en ítems que requerían modificaciones. Esto es una ventaja ya que este tipo de reportes, así como de decisiones con respecto a los reactivos, no suelen reportarse en los estudios de diseño y desarrollo de escalas, siendo más frecuente la transparencia de los resultados por ítem en estudios metodológicos que tienen la finalidad de recopilar evidencias de validez de contenido (Polit y Beck, 2006).

Asimismo, en la obtención de evidencias de validez en cuanto a la imparcialidad de las escalas, el panel de expertos emitió observaciones sobre todo en los criterios de *comprensibilidad e inclusión poblacional*. Los señalamientos de los jueces abarcaron un rango amplio de observaciones desde las ortográficas, gramaticales y de redacción, así como la prudencia de las instrucciones, la precisión del constructo y las consideraciones epistemológicas de las escalas vistas desde la perspectiva de la filosofía de la ciencia. En relación con la integración de evidencias de validez, se realizaron modificaciones que se consideraron pertinentes en los ítems de acuerdo con las puntuaciones de los índices de validez, así como de las observaciones de los jueces.

Es relevante mencionar que una de las aportaciones más relevantes del trabajo de tesis fue la metodología utilizada para la recopilación de evidencias de validez. El procedimiento propuesto consistió en extraer dos tipos de evidencias de validez de constructo: del contenido y de la imparcialidad en una misma intervención asincrónica y colaborativa con el panel de expertos. De esta manera, los jueces desarrollaron su colaboración en un documento compartido que permitió el debate y el enriquecimiento de la recopilación de evidencias de validez de las escalas.

Otro aspecto relevante de este tipo de metodología fue la modalidad virtual y asincrónica, ya que este procedimiento permitió tener expertos con horarios diversos y en diferentes ubicaciones geográficas. Esta modalidad brindó la oportunidad de que los expertos pudieran gestionar el tiempo invertido en el procedimiento de recopilación de evidencias, desde la lectura del Manual de procedimientos hasta responder el Protocolo de recopilación de evidencias de validez de contenido. Asimismo, las ventajas de este tipo de metodología es que se logró contar con jueces expertos en diseño, desarrollo y validación de instrumentos, evaluación educativa, así como docencia en metodología de la investigación en nivel superior. En resumen, todas las etapas del procedimiento se realizaron a través de medios digitales con el objetivo de adaptarse a las nuevas tendencias del trabajo remoto. Esta forma de trabajo, así como los

modelos híbridos, además de tener múltiples beneficios como la flexibilidad laboral, es una tendencia que se mantendrá de manera duradera (OCDE, 2021).

Por último, es importante aclarar que, aunque las escalas obtuvieron resultados favorables en la recopilación de evidencias de validez con respecto a su contenido, al término de este trabajo de tesis la EMAC, EAC y EAPAMC están en una fase inicial de desarrollo. Debe señalarse también que el proceso de validación de los instrumentos de medición implicar reunir la suficiente evidencia para sostener los usos e interpretaciones propuestas por los desarrolladores de las escalas (AERA, APA y NCME, 2018), por lo que no se puede generalizar con respecto a sus resultados en una etapa tan temprana de desarrollo. Asimismo, y como se verá en el apartado de limitaciones, una vez concluido este trabajo de investigación se deberá realizar una aplicación piloto de los instrumentos para continuar con el proceso de recopilación de evidencias de validez con respecto a la estructura interna, el proceso de respuesta y de relación con otras variables.

5.2. Limitaciones del estudio

En cuanto a las limitaciones del estudio se puede mencionar que en este trabajo se realizó una revisión de la literatura en torno a la temática del aprendizaje de las ciencias en estudiantes de ES con énfasis en el dominio afectivo del aprendizaje de las ciencias, dejando de lado el dominio cognitivo. En este tenor, es relevante mencionar que se abordaron solamente algunos constructos del dominio afectivo del aprendizaje de las ciencias, así como modelos teóricos y de medición desde una perspectiva psicológica. Se utilizaron las subescalas propuestas por otros referentes de la medición de estos constructos como el SLMQ-II (Glynn et al. 2007, 2011), el TOSRA (1982) y la prueba PISA (OCDE, 2015), así como los modelos de medición en escalamiento propuestos por dichos referentes.

Es importante señalar que en este estudio se recopilieron únicamente las evidencias de validez con respecto al contenido de las escalas, así como de la imparcialidad de las mismas, por lo que es necesario seguir los *Estándares* para un correcto uso e interpretación de resultados de los instrumentos de medición. En cuanto a la recopilación de evidencias del contenido, se pueden mencionar las siguientes limitaciones: por un lado, se observó que, aunque se pudo reunir un comité robusto de expertos en evaluación educativa y en el diseño, desarrollo y validación de instrumentos de medición, no se tuvo acceso o acercamiento a investigadores expertos en los constructos de interés.

Por otro lado, aunque el formato virtual del proceso de recopilación de evidencias de validez tuvo múltiples ventajas, también tuvo limitaciones con respecto al tiempo de capacitación y calibración de los jueces. En este tenor, una de las probables limitaciones fue una comprensión diversa de los criterios para el cálculo del IVC y de las categorías de la EDU, incluso de las propias instrucciones del proceso de validación. Por mencionar un ejemplo, aunque se les dio la instrucción a los jueces de comentar un ítem en caso de observar detalles en referencia a la EDU y la relevancia, claridad, simplicidad y precisión de los ítems, hubo dos jueces que no emitieron comentarios en ninguna de las escalas. Otra limitación acerca de la aplicación de la metodología es que algunos de los jueces no fueron congruentes en sus opiniones en la fase cualitativa del procedimiento y las puntuaciones del *Protocolo de validación de contenido*. Es decir, un ítem pudo resultar con una baja puntuación en los cálculos del IVC, pero no tener ningún comentario con alguna sugerencia de mejora y viceversa, por ejemplo, un ítem tuvo muchos comentarios, pero obtuvo puntuaciones adecuadas en el PAJ o en el IVC.

Otra limitación con respecto al procedimiento es que se produjo un efecto de *eco* en los comentarios de los jueces. Un ejemplo de esto es que un ítem puede llegar a tener más de seis o siete comentarios con la misma observación, mientras que otros ítems no recibieron ningún comentario. En torno a este fenómeno, es probable que sea prudente adoptar un menor número de jueces (no mayor de diez) como

sugiere Lynn (1986 en Polit y Beck, 2006) para evitar estas situaciones. En la misma vía, algunos jueces comentaban estar de acuerdo con las observaciones de algún otro juez sin agregar alguna otra aportación. Debido a esto, es recomendable reevaluar las fortalezas y debilidades del trabajo colaborativo. Una posible solución a esta situación podría ser el volver anónima la participación de los jueces al mismo tiempo que continúa siendo colaborativa.

Otra limitación observada fue la comprensión del procedimiento con respecto al *Manual de procedimientos*, por ejemplo, la mayoría de los jueces acataron las instrucciones de manera precisa, mientras otros hicieron ligeras variaciones. Aunque estas variaciones no afectaron el procedimiento de recopilación de evidencias de validez en sí mismo, pero si al momento de hacer el análisis de los datos, sobre todo los comentarios de naturaleza cualitativa, lo cual demoró la integración de la información. Una hipótesis de este fenómeno podría ser que algunos jueces consideraron más rápido o práctico el realizar las observaciones de manera individual.

5.3. Recomendaciones

Dada la discusión y las limitaciones del estudio se proponen una serie de recomendaciones para investigaciones futuras en el diseño, desarrollo y validación de escalas para medir la motivación, las actitudes y la autoeficacia en el aprendizaje de las ciencias desde un enfoque integrador en estudiantes de ES a nivel de licenciatura. En especial, se recomienda continuar con la recopilación de evidencias de validez en concordancia con las directrices de los *Estándares* para el desarrollo de instrumentos de medición sugeridas por la AERA, APA y NCME (2018). Aunado a ello, se proponen una serie de recomendaciones metodológicas y prácticas para investigaciones futuras:

- Ampliar la revisión de la literatura para la elaboración de antecedentes y fundamentos teóricos de la medición de la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias tomando en cuenta otros términos clave, periodos de tiempo, bases de datos, revistas y repositorios.
- Contemplar la utilización de constructos similares que surjan de la revisión de marcos teóricos distintos, como por ejemplo sociológico, antropológico y económico, y no estrictamente psicológico.
- Ampliar la revisión de antecedentes y fundamentos teóricos de otros constructos relevantes en el aprendizaje de las ciencias que pudieron haberse obviado en la búsqueda para la realización de la presente tesis, tanto de dominio afectivo como cognitivo. Por ejemplo, las habilidades de pensamiento científico o el pensamiento crítico, entre otros.
- Darle seguimiento a la recopilación de evidencias de validez de constructo de acuerdo con las recomendaciones de los Estándares de la AERA, APA y NCME (2018).
- Aplicar el instrumento a una muestra amplia y diversa de estudiantes universitarios, con la finalidad de obtener evidencias de validez en el aspecto de proceso de respuesta al ítem de las escalas, de estructura interna y confiabilidad.
- Aplicar el instrumento en diferentes regiones del país y en la medida de lo posible, realizar estudios comparativos con países hispanoparlantes.
- Adaptar el instrumento para población más joven, en otros niveles educativos como la educación básica, y educación media superior.
- Adaptar el instrumento al idioma inglés, para su uso y estandarización en población angloparlante.
- Adaptar el instrumento a formato físico para su aplicación en microescala, como grupos de clase.
- Incluir en las instrucciones del *Protocolo de validación de contenido* en su versión en línea, que los jueces deberán realizar un comentario de manera obligatoria de seleccionar puntuaciones bajas como 1 o 2 en los criterios para el cálculo del IVC.

- Evaluar la capacitación de los jueces, con la finalidad de conocer si el procedimiento de recopilación de evidencias de validez fue comprendido en su totalidad. En este punto es importante señalar la necesidad de reforzar estas indicaciones si se optará por una capacitación en línea que, aunque es un formato práctico y que permite tener jueces en diferentes zonas geográficas, puede tener como consecuencia un entendimiento parcial de las instrucciones del procedimiento.
- De ser posible, diseñar una rúbrica similar al *Protocolo de validación de contenido* para los criterios de la EDU, además de los comentarios abiertos en los ítems de las escalas. Esto con la finalidad de tener más información sistematizada y directamente relacionada con un criterio de la EDU.

REFERENCIAS

- Abd-El-Khalick, F., Summers, R., Said, Z., Wang, S. & Culbertson, M. (2015). Development and Large-Scale Validation of an Instrument to Assess Arabic Speaking Students' Attitudes Toward Science. *International Journal of Science Education*, 37 (16), 2637-2663, DOI: 10.1080/09500693.2015.1098789.
- Aiken, L. R. (1980). Content Validity and Reliability of Single Items or Questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959. doi:10.1177/001316448004000419.
- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131–142. doi:10.1177/0013164485451012
- Aiken, L. R. & Aiken, D. R. (1969). Recent research on attitudes concerning science. *Science Education*, 53 (4), 295-305.
- Aguilera, D., & Perales Palacios, F. J. (2019). Actitud hacia la Ciencia: desarrollo y validación estructural del School Science Attitude Questionnaire (SSAQ). *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias.*, 16(3), 1–20. doi:10.25267/rev_eureka_ensen_di
- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84(5), 888–918. <https://doi.org/10.1037/0033-2909.84.5.888>
- Albarracín, D., Johnson, B.T. & Zanna, M.P. (2005). *The handbook of attitudes* (1st ed). Psychology Press.
- Aldana, G.M., Babativa, D.A, Caraballo, G.J, & Rey, C.A. (2020). Escala de actitudes hacia la investigación (EACIN): evaluación de sus propiedades psicométricas en una muestra colombiana. *Rev. CES Psico*, 13(1), 89-103.
- American Educational Research Association, American Psychological Association y National Council on Measurement Education (2018). *Estándares para el desarrollo de pruebas psicológicas y educativas*. American Educational Research Association. AERA, APA, NCME.
- Anderson, Ch. W. (2009). Perspectives on Science Learning en S. K. Abell y N.G. Lederman (Ed.), *Handbook of Research on Science Education* (2nd ed., pp. 3–30). Routledge.
- Ardura, D., & Pérez-Bitrián, A. (2018). The effect of motivation on the choice of chemistry in secondary schools: adaptation and validation of the Science Motivation Questionnaire II to Spanish students. *Chemistry Education Research and Practice*, 19(3), 905–918. doi:10.1039/c8rp00098k

- Bandura, A. (1977). *Social learning theory*. Prentice-Hall.
- Bandura, A. (1986). Fearful expectations and avoidant actions as coefficients of perceived self-inefficacy. *American Psychologist*, *41*(12), 1389–1391. <https://doi.org/10.1037/0003-066X.41.12.1389>.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.
- Bandura, A. (2010). *Self-Efficacy*. *The Corsini Encyclopedia of Psychology*. doi:10.1002/9780470479216.corpsy0836
- Benjamin, T. E., Marks, B., Demetrikopoulos, M. K., Rose, J., Pollard, E., Thomas, A., & Muldrow, L. L. (2015). Development and Validation of Scientific Literacy Scale for College Preparedness in STEM with Freshmen from Diverse Institutions. *International Journal of Science and Mathematics Education*, *15*(4), 607–623. doi:10.1007/s10763-015-9710-x.
- Bieschke, K. J., Bishop, R. M., & Garcia, V. L. (1996). The Utility of the Research Self-Efficacy Scale. *Journal of Career Assessment*, *4*(1), 59–75. doi:10.1177/106907279600400104
- Blalock, Ch. L., Lichtenstein, M. J., Owen, S., Pruski, L., Marshall, C. & Toepperwein, M. (2008). In Pursuit of Validity: A comprehensive review of science attitude instruments 1935–2005. *International Journal of Science Education*, (30) 7, 961-977. doi: 10.1080/09500690701344578
- Blanco-Blanco, A., Casas Moreno, Y. & Mafokozi Ndabishibije, J. (2016). Adaptación y propiedades psicométricas de escalas sociocognitivas. Una aplicación en el ámbito vocacional científico-matemático. *REOP*, (27) 1, 8-28.
- Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V. & Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, *59* (11), 977–984. <https://doi.org/10.1525/bio.2009.59.11.9>
- Bonney, R., Phillips, T. B., Ballard, H. L., & Enck, J. W. (2015). Can citizen science enhance public understanding of science? *Public Understanding of Science*, *25*(1), 2–16. doi:10.1177/0963662515607406
- Breakwell, G.M. & Robertson, T. (2001). The gender gap in science attitudes, parental and peer influences: changes between 1987–88 and 1997–98. *Public Understanding of Science*, *10*, 71-82. doi.org/10.1088/0963-6625/10/1/305
- Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, *43*(5), 485–499. doi:10.1002/tea.20131

- Buccheri, G., Gürber, N. A., & Brühwiler, C. (2011). The Impact of Gender on Interest in Science Topics and the Choice of Scientific and Technical Vocations. *International Journal of Science Education*, 33(1), 159–178. doi:10.1080/09500693.2010.518643
- Bunge, M. (2015). *Epistemología*. Siglo veintiuno editores.
- Cardona Mora, J. N., Bribiescas Silva, F., Romero López, R. & Corona Cortez, R. E. (2016). Design, Adaptation and Content Validity Process of a Questionnaire: A Case Study. *International Journal of Management*, 7(7), pp. 204–216.
- Chen, J. A., & Usher, E. L. (2013). Profiles of the sources of science self-efficacy. *Learning and Individual Differences*, 24, 11–21. doi:10.1016/j.lindif.2012.11.002
- Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The Role of Efficacy and Identity in Science Career Commitment Among Underrepresented Minority Students. *Journal of Social Issues*, 67(3), 469–491. doi:10.1111/j.1540-4560.2011.01710.x
- Contreras, L.A. (2009). Manual para el desarrollo y elaboración exámenes criteriosales de gran escala alineados con el currículum. Ensenada: UEE.
- Criollo, M., Romero, M. & Fontaines-Ruiz, T. (2017). Autoeficacia para el aprendizaje de la investigación en estudiantes universitarios. *Psicología Educativa*, 23(1), 1-10.
- Crotwell Timmerman, B. E., Strickland, D.C., Johnson, R. L. & Payne, J.R. (2011). Development of a ‘universal’ rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, (36) 5, 509-547. doi: 10.1080/02602930903540991
- Cuevas Romo, A., Hernández, R., Leal, B. E., & Mendoza, C. P. (2016). Enseñanza- Aprendizaje de la investigación en educación básica en México. *Revista Electrónica de Investigación Educativa*, 18 (3), 187-200. <http://dx.doi.org/10.4067/S0718-50062015000400011>
- Davidescu, D. C., Dafinei, M., Dafinei, A., & Antohe, S. (2011). Scientific abilities and their assessment applying SMELT Physical Study of the MatErial erupTing from mud volcanoes. *Procedia - Social and Behavioral Sciences*, 15, 1140–1152. doi:10.1016/j.sbspro.2011.03.253
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers.
- Embretson, S., & Gorin, J. (2001). Improving Construct Validity With Cognitive Psychology Principles. *Journal of Educational Measurement*, 38(4), 343–368. doi:10.1111/j.1745-3984.2001.tb01131.x
- Escobar, J. & Cuervo, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*, 6, 27–36.

- Etkina, E., Van Heuvelen, A., White-Brahmia, S., Brookes, D. T., Gentile, M., Murthy, S., & Warren, A. (2006). Scientific abilities and their assessment. *Physical Review Special Topics - Physics Education Research*, 2(2). doi:10.1103/physrevstper.2.020103.
- Feist, G. J. (2006a). How Development and Personality Influence Scientific Thought, Interest, and Achievement. *Review of General Psychology*, 10(2), 163–182. doi:10.1037/1089-2680.10.2.163
- Feist, G. J. (2006b). *The Psychology of Science and the Origins of the Scientific Mind*. Yale University Press.
- Feist, G. J. (2014). Psychometric studies of scientific talent and eminence. In D. K. Simonton (Ed.), *The Wiley handbook of genius* (p. 62–86). Wiley Blackwell. <https://doi.org/10.1002/9781118367377.ch4>
- Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific Culture and Educational Research. *Educational Researcher*, 31(8), 4–14. doi:10.3102/0013189x031008004
- Fontaines Ruiz, T. & Urdaneta, G. (2009). Culturas de formación y formación de investigadores educativos. *Ra Ximhai*, 5 (3), 357 -371.
- Fraser, B. J. (1982). *Test of science related attitudes*. Australian Council for Educational Research.
- Fugelsang, J. A., Stein, C. B., Green, A. E., & Dunbar, K. N. (2004). Theory and data interactions of the scientific mind: Evidence from the molecular and the cognitive laboratory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 58(2), 86–95.
- García Sedeño, M., & García-Tejera, M. C. (2014). Estimate of the Content Validity on a Scale to Assess Gender Violence Rating Supported in Adolescents. *Acción Psicológica*, 10(2), 41–58. <https://doi.org/10.5944/ap.10.2.11823>
- Garza-Almanza, V. (2016). Periodismo científico en México. Necesidades y Propuestas. *Cultura Científica Y Tecnológica*, (58). Recuperado a partir de <http://erevistas.uacj.mx/ojs/index.php/culcyt/article/view/1082>
- Germann, P. J. (1994). Testing a model of science process skills acquisition: An interaction with parents' education, preferred language, gender, science attitude, cognitive development, academic ability, and biology knowledge. *Journal of Research in Science Teaching*, 31(7), 749–783. doi:10.1002/tea.3660310707
- Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2007). Nonscience majors learning science: A theoretical model of motivation. *Journal of Research in Science Teaching*, 44(8), 1088–1107. doi:10.1002/tea.20181

- Glynn, S. M., Brickman, P., Armstrong, N. & Taasoobshiraz, G. (2011). Science Motivation Questionnaire II: Validation with Science Majors and Nonscience Majors. *Journal of Research in Science Teaching*, 48 (10), 1159–1176.
- Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2012). From Log Files to Assessment Metrics: Measuring Students' Science Inquiry Skills Using Educational Data Mining. *Journal of the Learning Sciences*, 22(4), 521–563. doi:10.1080/10508406.2013.837391.
- Godin, B. & Gingras, Y. (2000). What is scientific and technological culture and how is it measured? A multidimensional model. *Public Understanding of Science*, 9 (1).
- Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, 102(4), 700–717. <https://doi.org/10.1037/a0026659>.
- Haladyna, T. M. & Rodríguez, M.C. (2013). *Developing and Validating Test Items*. doi:10.4324/9780203850381.
- Hartman, R. O., Dieckmann, N. F., Sprenger, A. M., Stastny, B. J., & DeMarree, K. G. (2017). Modeling Attitudes Toward Science: Development and Validation of the Credibility of Science Scale. *Basic and Applied Social Psychology*, 39(6), 358–371. doi:10.1080/01973533.2017.1372284.
- Hemmings, B. & Kay, R. (2010). Research self-efficacy, publication output, and early career development. *International Journal of Educational Management*, 24 (7), 562-574. <https://doi.org/10.1108/09513541011079978>.
- Hernández-Sampieri, R., Fernández-Collado, C. & Baptista-Lucio, P. (2014). Definición conceptual o constitutiva. En *Metodología de la Investigación* (6ª ed., pp. 119-125). México: McGraw-Hill.
- Hillman, S. J., Zeeman, S. I., Tilburg, C. E., & List, H. E. (2016). My Attitudes Toward Science (MATS): the development of a multidimensional instrument measuring students' science attitudes. *Learning Environments Research*, 19(2), 203–219. doi:10.1007/s10984-016-9205-x.
- Hodson, D. (2014). Learning Science, Learning about Science, Doing Science: Different goals demand different learning methods. *International Journal of Science Education*, 36(15), 2534–2553. doi:10.1080/09500693.2014.899722
- Holland, J. L., Fritzsche, B. A., & Powell, A. B. (2005). *Búsqueda autodirigida SDS: Manual técnico*. México : El Manual Moderno.
- Instituto Nacional de Estadística y Geografía. (noviembre de 2021). Encuesta sobre la Percepción Pública de la Ciencia y la Tecnología (ENPECYT) 2017. <https://www.inegi.org.mx/programas/enpecyt/2017/>.

- Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, *41*, 13–24. doi:10.1016/j.cedpsych.2014.11.0.
- Jegede, O. J. (1997). School science and the development of scientific culture: a review of contemporary science education in Africa. *International Journal of Science Education*, *19*(1), 1–20. doi:10.1080/0950069970190101.
- Kawamoto, S., Nakayama, M., & Saijo, M. (2013). A survey of scientific literacy to provide a foundation for designing science communication in Japan. *Public Understanding of Science*, *22*(6), 674–690. <https://doi.org/10.1177/0963662511418893>.
- Keller, M. M., Neumann, K., & Fischer, H. E. (2016). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*, *54*(5), 586–614. doi:10.1002/tea.21378
- Klopfer, L. E. (1973). A Structure for the Affective Domain in Relation to Science Education. Disponible en: <https://files.eric.ed.gov/fulltext/ED079071.pdf>.
- Koballa, T. R. & Glynn, S. M. (2009). Attitudinal and motivational constructs in Science Learning en S. K. Abell y N.G. Lederman (Ed.), *Handbook of Research on Science Education* (2nd ed., pp. 75–102). Routledge.
- Knaggs, C. M., & Sondergeld, T. A. (2015). Science as a Learner and as a Teacher: Measuring Science Self-Efficacy of Elementary Preservice Teachers. *School Science and Mathematics*, *115*(3), 117–128. doi:10.1111/ssm.12110
- Kruit, P. M., Oostdam, R. J., van den Berg, E., & Schuitema, J. A. (2018). Assessing students' ability in performing scientific inquiry: instruments for measuring science skills in primary education. *Research in Science & Technological Education*, 1–27. doi:10.1080/02635143.2017.1421530.
- Kuhn, D. (2012). *Enseñar a pensar*. Amorrortu Editores.
- Lawshe, C. H. (1975). "A quantitative approach to content validity", *Personnel Psychology*, (28) 4, 563–575.
- Lee, S., & Kim, S.H. (2018). Scientific Knowledge and Attitudes Toward Science in South Korea: Does Knowledge Lead to Favorable Attitudes? *Science Communication*, *40*(2), 147–172. doi:10.1177/1075547017753189.
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*, 79–122.

- Lent, R. W., Brown, S. D., Sheu, H.-B., Schmidt, J., Brenner, B. R., Gloster, C. S., Wilkins, G., Schmidt, L. C., Lyons, H., & Treistman, D. (2005). Social Cognitive Predictors of Academic Interests and Goals in Engineering: Utility for Women and Students at Historically Black Universities. *Journal of Counseling Psychology, 52*(1), 84–92. <https://doi.org/10.1037/0022-0167.52.1.84>
- Lent, R. W., Sheu, H.-B., Miller, M. J., Cusick, M. E., Penn, L. T., & Truong, N. N. (2018). Predictors of science, technology, engineering, and mathematics choice options: A meta-analytic path analysis of the social–cognitive choice model by gender and race/ethnicity. *Journal of Counseling Psychology, 65*(1), 17–35. <https://doi.org/10.1037/cou0000243>
- Lippa, R. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people–things dimension. *Journal of Personality and Social Psychology, 74*(4), 996–1009. <https://doi.org/10.1037/0022-3514.74.4.996>.
- Livinți, R., Gunnesch-Luca, G., & Iliescu, D. (2021). Research self-efficacy: A meta-analysis. *Educational Psychologist, 56*(3), 215–242. doi:10.1080/00461520.2021.188610.
- Macedoi, B. (2016). *Educación científica*. Foro Abierto de Ciencias de América Latina y el Caribe.
- Maier, M. F., Greenfield, D. B., & Bulotsky-Shearer, R. J. (2013). Development and validation of a preschool teachers' attitudes and beliefs toward science teaching questionnaire. *Early Childhood Research Quarterly, 28*(2), 366–378. doi:10.1016/j.ecresq.2012.09.003.
- Mallinckrodt, B. & Gelso CH. J. (2002). Impact of Research Training Environment and Holland Personality Type: A 15-Year Follow-Up of Research Productivity. *Journal of Counseling Psychology, 49* (1), 60–70.
- Meliá Navarro, J. M. (1990). *La construcción de la psicometría como ciencia teórica y aplicada*. Cristóbal Serrano.
- Messick, S. (1989). *Validity*. In R. L. Linn (Ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement* (p. 13–103). Macmillan Publishing Co, Inc; American Council on Education.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Midtgarden T. (2020). Peirce's Classification of the Sciences. *Knowledge Organization, 47* (3), 267-278, doi: 10.5771/0943-7444-2020-3-267

- Molina, M., Carriazo, J. & Casas, J. (2013). Estudio transversal de las actitudes hacia la ciencia en estudiantes de grados quinto a undécimo. Adaptación y aplicación de un instrumento para valorar actitudes. *Tecné, Episteme y Didaxis: TED*, (33). <https://doi.org/10.17227/01213814.33ted103.122>
- Munby, H. (1997). Issues of validity in science attitude measurement. *Journal of Research in Science Teaching*, 34(4), 337–341. doi:10.1002/(sici)1098-2736(199704)34:4<337::aid-tea4>3.0.co;2-s
- Navarro, M., Forster, C., González, C., & González-Pose, P. (2016). Attitudes toward science: measurement and psychometric properties of the Test of Science-Related Attitudes for its use in Spanish-speaking classrooms. *International Journal of Science Education*, 38(9), 1459–1482. doi:10.1080/09500693.2016.1195521.
- Nitko, A.J. (1994). A Model for Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations for Certification and Selection of Students. International Conference on Educational Evaluation and Assessment (2nd, Pretoria, South Africa, July 1994).
- Organización para la Cooperación y el Desarrollo Económicos (2022). El programa PISA de la OCDE. Qué es y para qué sirve. Organización para la Cooperación y el Desarrollo Económicos, París.
- Osborne, J., Simon, S. & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25 (9), 1049-1079.
- Ossa-Cornejo, C. J., Palma-Luengo, M. R., Lagos-San Martín, N. G., Quintana-Abello, I. M., & Díaz-Larenas, C. H. (2017). Análisis de instrumentos de medición del pensamiento crítico, *Ciencias Psicológicas*, 11(1), 19-28. <https://doi.org/10.22235/cp.v11i2.1343>
- Padilla González, J., Patiño Barba, L. & Herrera, S. (2020). *¿Qué ciencia necesita el ciudadano?. SOMEDICYT y FIBONACCI.*
- Pajares, F. (1996). Self-Efficacy Beliefs in Academic Settings. *Review of Educational Research*, 66(4), 543–578. doi:10.3102/00346543066004543
- Pedrosa, I., Suarez, J., & García, E. (2014). Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación. *Acción psicológica*, 10 (2), 3-18.
- Perales, R. G. (2018). Diseño y construcción de un instrumento de evaluación de la competencia matemática: aplicabilidad práctica de un juicio de expertos. *Ensaio: Avaliação e Políticas Públicas Em Educação*, 26(99), 347–372. doi:10.1590/s0104-40362018002601263
- Perera, L. D. H. (2014). Parents' Attitudes Towards Science and their Children's Science Achievement. *International Journal of Science Education*, 36(18), 3021–3041. doi:10.1080/09500693.2014.949900

- Pérez Morán, J. C. (2010). *Evaluación criterial del área metodológica de la carrera de Psicología de la UABC* [Tesis de Maestría, Universidad Autónoma de Baja California]. Repositorio del Instituto de Investigación y Desarrollo Educativo.
- Pérez Morán, J. C. (2014). *Análisis del aspecto sustantivo de la validez de constructo de una prueba de habilidades cuantitativas* [Tesis de Doctorado, Universidad Autónoma de Baja California]. Repositorio del Instituto de Investigación y Desarrollo Educativo.
- Pérez Tamayo, R. (2009). Ciencia, conocimiento e identidad nacional. *REencuentro. Análisis de Problemas Universitarios*, (56), 12-16. Disponible en: <https://www.redalyc.org/articulo.oa?id=34011860003>
- Phillips, J. C., & Russell, R. K. (1994). Research Self-Efficacy, the Research Training Environment, and Research Productivity among Graduate Students in Counseling Psychology. *The Counseling Psychologist*, 22(4), 628–641. doi:10.1177/0011000094224008
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. doi:10.1002/nur.20147
- Potvin, P., & Hasni, A. (2014). Interest, motivation and attitude towards science and technology at K-12 levels: a systematic review of 12 years of educational research. *Studies in Science Education*, 50(1), 85–129. doi:10.1080/03057267.2014.881626
- Ramírez, C., Reyna, M., García, A. Ortiz, X. & Valdez, P. (2011). Formación científica de los egresados de tres programas de maestría en ciencias: seguimiento a 10 años (1999-2009). *Revista de la educación superior*, 2 (158), 91-103.
- Reyes Cruz, M. & Gutiérrez Arceo, J. M. (2015). Sentido de autoeficacia en investigación de estudiantes de posgrado. *Sinéctica*, (45), 1-15. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-109X2015000200011&lng=es&tlng=.
- Rizeq, J., Flora, D. B., & Toplak, M. E. (2020). An examination of the underlying dimensional structure of three domains of contaminated mindware: paranormal beliefs, conspiracy beliefs, and anti-science attitudes. *Thinking & Reasoning*, 27(2), 187–211. doi:10.1080/13546783.2020.1759688
- Rodríguez, W., Jiménez, R, & Caicedo-Maya, C. (2007). Protocolo de actitudes relacionadas con la ciencia: adaptación para Colombia. *Psychologia. Avances de la disciplina*, 1(2), 85-100.
- Rodríguez de la Vega-Elu, M. (2021). *Desarrollo y validación de una prueba para el diagnóstico cognitivo de los aprendizajes del inglés en educación secundaria* [Tesis de Doctorado, Universidad Autónoma de Baja California]. Repositorio del Instituto de Investigación y Desarrollo Educativo.

- Rodríguez Méndez, M.C., Peña Calvo, J.V. & Inda Cxaro, M. (2012). Creencias de autoeficacia y elección femenina de estudios científico-tecnológicos: una revisión teórica de su relación. *Teoría de la Educación. Revista Interuniversitaria*, 24 (1), 81-104. <https://doi.org/10.14201/10333>.
- Romine, W. L., Sadler, T. D., & Kinslow, A. T. (2016). Assessment of scientific literacy: Development and validation of the Quantitative Assessment of Socio-Scientific Reasoning (QuASSR). *Journal of Research in Science Teaching*, 54(2), 274–295. doi:10.1002/tea.21368
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1), 54–67. doi:10.1006/ceps.1999.1020
- Schunk, D. H. (1991). Self-Efficacy and Academic Motivation. *Educational Psychologist*, 26(3-4), 207–231. doi:10.1080/00461520.1991.9653133
- Siegel, M. A. & Ranney, M. A. (2003). Developing the Changes in Attitude about the Relevance of Science (CARS) Questionnaire and Assessing Two High School Science Classes. *Journal of research in Science Teaching*, 40 (8), 757–775.
- Sireci, S. G. (1998). "The Construct of Content Validity", *Social Indicators Research*, (45), 83-117.
- Stekolschik, G., Gallardo, S. y Draghi, C. (2007). La comunicación pública de la ciencia y su rol en el estímulo de la vocación científica. *Redes*, 12 (25), 165-180.
- Stets, J. E., Brenner, P. S., Burke, P J.& Serpe, R. T. (2016). The science identity and entering a science occupation. *Social Science Research*. doi: 10.1016/j.ssresearch.2016.10.016
- Stone, E. M. (2014). Guiding Students to Develop an Understanding of Scientific Inquiry: A Science Skills Approach to Instruction and Assessment. *CBE—Life Sciences Education*, 13(1), 90–101. doi:10.1187/cbe-12-11-0198
- Summers, R. & Abd-El-Khalick, F. (2018). Development and Validation of an Instrument to Assess Student Attitudes Toward Science Across Grades 5 Through 1. *Journal of research in Science Teaching*, 55 (2), 172–205. doi 10.1002/tea.21416
- Schwarzer, R., & Hallum, S. (2008). Perceived Teacher Self-Efficacy as a Predictor of Job Stress and Burnout: Mediation Analyses. *Applied Psychology*, 57(s1), 152–171. doi:10.1111/j.1464-0597.2008.00359.x
- Swenson-Britt, E., & Berndt, A. (2013). Development and Psychometric Testing of the Nursing Research Self-Efficacy Scale (NURSES). *Journal of Nursing Measurement*, 21(1), 4–22. doi:10.1891/1061-3749.21.1.4

- Thompson, S., Johnstone, C. & Thurlow, M. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). University of Minnesota, National Center on Educational Outcomes.
- Thyer, B. A. (2008). The quest for evidence-based practice?: We are all positivists! *Research on Social Work Practice, 18*(4), 339–345. <https://doi.org/10.1177/1049731507313998>
- Toma, R. B., & Lederman, N. G. (2020). A Comprehensive Review of Instruments Measuring Attitudes Toward Science. *Research in Science Education*. doi:10.1007/s11165-020-09967-1.
- Tristán-López, A. (2008). “Modificación al modelo de Lawshe para el dictamen de validez de contenido de un instrumento objetivo”, *Avances en Medición, 6* (1), 37-48.
- Trujillo, G., & Tanner, K. D. (2014). Considering the Role of Affect in Learning: Monitoring Students’ Self-Efficacy, Sense of Belonging, and Science Identity. *CBE—Life Sciences Education, 13*(1), 6–15. doi:10.1187/cbe.13-12-0241
- Tuan *, H., Chin, C. y Shieh, S. (2005). The development of a questionnaire to measure students’ motivation towards science learning. *International Journal of Science Education, 27*(6), 639–654. doi:10.1080/0950069042000323737
- Tytler, R., & Osborne, J. (2011). Student Attitudes and Aspirations Towards Science. *Second International Handbook of Science Education, 597–625*. doi:10.1007/978-1-4020-9041-7_41
- Udo, M.K., Ramsey, G.P. & Mallow, J.V. (2004). Science Anxiety and Gender in Students Taking General Education Science Courses. *J Sci Educ Technol, 13*, 435–446. <https://doi.org/10.1007/s10956-004-1465-z>
- UNESCO (2020). Investigadores dedicados a investigación y desarrollo (por cada millón de personas). Disponible en: <https://datos.bancomundial.org/indicador/SP.POP.SCIE.RD.P6?view=chart>.
- UNESCO (2021). La ciencia al servicio de un futuro sostenible. Disponible en: <https://es.unesco.org/themes/ciencia-al-servicio-futuro-sostenible>
- UNESCO (2021). Educación en ciencias: cómo despertar vocaciones. Disponible en: <https://es.unesco.org/news/educacion-ciencias-como-despertar-vocaciones>
- Unfried, A., Faber, M., Stanhope, D. S., & Wiebe, E. (2015). The Development and Validation of a Measure of Student Attitudes Toward Science, Technology, Engineering, and Math (S-STEM). *Journal of Psychoeducational Assessment, 33*(7), 622–639. doi:10.1177/0734282915571160.
- Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic Versus Extrinsic Goal Contents in Self-Determination Theory: Another Look at the Quality of Academic Motivation. *Educational Psychologist, 41*(1), 19–31. doi:10.1207/s15326985ep4101_4

- Van Aalderen-Smeets, S., & Walma van der Molen, J. (2013). Measuring Primary Teachers' Attitudes Toward Teaching Science: Development of the Dimensions of Attitude Toward Science (DAS) Instrument. *International Journal of Science Education*, 35(4), 577–600. doi:10.1080/09500693.2012.755576.
- Vázquez Alonso, A. & Manassero Mas, M. A. (1997). Una evaluación de las actitudes relacionadas con la ciencia. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas*, 15 (2), 199-213.
- Vázquez Alonso, A. & Manassero Mas, M. A. (2009a). La relevancia de la educación científica: actitudes y valores de los estudiantes relacionados con la ciencia y la tecnología. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas*, 27 (1) 33-48.
- Vázquez Alonso, A. & Manassero Mas, M. A. (2009b). Patrones actitudinales de la vocación científica y tecnológica en chicas y chicos de secundaria. *Revista iberoamericana de educación*, 50 (4) 2-15.
- Vázquez Alonso, A. & Manassero Mas, M. A. (2009c). Expectativas sobre un trabajo futuro y vocaciones científicas en estudiantes de educación secundaria. *Revista electrónica de investigación educativa*, 11 (1) 1-20.
- Vázquez-Alonso, A. & Manassero-Mas, M.A. (2015). La elección de estudios superiores científico-técnicos: análisis de algunos factores determinantes en seis países. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 12(2), 264-277. DOI: 10498/17251
- Vázquez-Alonso, A. & Manassero-Mas, M.A. (2018). Más allá de la comprensión científica: educación científica para desarrollar el pensamiento. *Revista Electrónica de Enseñanza de las Ciencias*, 17 (2), 309-336.
- Velayutham, S., Aldridge, J., & Fraser, B. (2011). Development and Validation of an Instrument to Measure Students' Motivation and Self-Regulation in Science Learning. *International Journal of Science Education*, 33(15), 2159–2179. doi:10.1080/09500693.2010.541529
- Wang, T., & Berlin, D. (2010). Construction and Validation of an Instrument to Measure Taiwanese Elementary Students' Attitudes toward Their Science Class. *International Journal of Science Education*, 32(18), 2413–2428. doi:10.1080/09500690903431561
- Wareing, C. (1982). Developing the wasp: Wareing attitudes toward science protocol. *Journal of Research in Science Teaching*, 19(8), 639–645. doi:10.1002/tea.3660190803
- Willson, V. L. (1983). A meta-analysis of the relationship between science achievement and science attitude: Kindergarten through college. *Journal of Research in Science Teaching*, 20(9), 839–850. doi:10.1002/tea.3660200906

Yuan, S., Ma, W., & Besley, J. C. (2019). Should Scientists Talk About GMOs Nicely? Exploring the Effects of Communication Styles, Source Expertise, and Preexisting Attitude. *Science Communication*, 41(3), 267–290. <https://doi.org/10.1177/1075547019837623>

APÉNDICES

Apéndice I.

Ficha Técnica de la Encuesta de Motivación, Actitudes y Autoeficacia en el Aprendizaje de las Ciencias en Estudiantes de Educación Superior (EMAAC-ESu)

Tabla de contenidos

I. Nombre de la encuesta	3
II. Instrumentos y formato que integran la encuesta	3
III. Propósitos y usos de la EMAAAC-ESu	3
IV. Antecedentes del estudio y medición de la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias	4
V. Referentes teóricos y metodológicos de la encuesta	6
VI. Descripción de los referentes de la Escala de Motivación en el Aprendizaje de las Ciencias (EMAC)	6
VII. Descripción de los referentes de la Escala de Actitudes hacia la Ciencia (EAC) ...7	
VIII. Descripción de los referentes de la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC)	9
IX. Enfoque de medición	10
Apéndice 1. Tabla de contenido de los ítems de la Escala de Motivación en el aprendizaje de las ciencias (EMAC)	11
Apéndice 2. Tabla de contenido de los ítems de la Escala de Actitudes hacia la Ciencia (EAC)	13
Apéndice 3. Tabla de contenido de los ítems de la Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC)	16
Referencias	18

- **Nombre de la encuesta:** Encuesta de Motivación, Actitudes y Autoeficacia en el Aprendizaje de las Ciencias en estudiantes de Educación Superior (EMAAAC-ESu).
- **Instrumentos y formato que integran la encuesta:** La EMAAAC-ESu está conformada por cuatro instrumentos. (1) Sección de datos de Identificación, (2) Escala de Motivación en el Aprendizaje de las Ciencias (EMAC), (3) Escala de Actitudes hacia la Ciencia (EAC), (3) Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC). Esta es la primera versión de la EMAAAC-ESu, está adaptada en formato físico (papel y lápiz) y posteriormente se adaptará en formato digital (google forms) para su fácil acceso y aplicación.
- **Propósitos y usos de la EMAAAC-ESu:** La EMAAAC-ESu tiene el propósito de explorar la motivación en el aprendizaje de las ciencias, las actitudes hacia la ciencia y la autoeficacia en el aprendizaje del método científico de estudiantes de Educación Superior (ES) en el nivel licenciatura. En especial con dicha escala, se desea ofrecer información a estudiantes que favorezca su toma de decisiones respecto al desarrollo y formación de su vocación científica. Asimismo, con la EMAAAC-ESu se puede aportar información a docentes, orientadores vocacionales y directivos para el diseño de estrategias pedagógicas, planes curriculares e intervenciones educativas que atiendan el desarrollo y la formación científica de sus estudiantes. El uso de la escala puede auxiliar también a los comités de posgrado, con la finalidad de tener más información del perfil de los profesionistas que buscan formarse como investigadores. Esta herramienta ayudará a los investigadores, tutores o asesores a conocer el estado del dominio afectivo y de autopercepción de los estudiantes en el aprendizaje de las ciencias, necesario para la formación de profesionales y nuevos científicos. Por su parte, a nivel institucional y a nivel del sistema educativo estatal y nacional, con la aplicación a gran escala de la EMAAAC-ESu, se puede ofrecer información para el diseño e implementación de políticas educativas en CyT encaminadas al fomento de vocaciones científicas en educación superior.
- **Antecedentes del estudio y medición de la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias:** Entre los antecedentes que se consideran relevantes en el estudio y medición de la motivación, actitudes y autoeficacia en el aprendizaje de las ciencias, se mencionan en especial las aportaciones de Fraser (1982) con el Test Of Science Related Attitudes (TOSRA), el Protocolo de Actitudes hacia la Ciencia (PAC) de Vázquez y

Manassero (1997), el Inventario de Búsqueda Autodirigida SDS de Holland, Fritzsche y Powell (2005), el capítulo sobre actitudes y motivación en el aprendizaje de las ciencias de Koballa y Glynn (2007) dentro del Handbook of Research on Science Education, la prueba Science Motivation Learning Questionnaire II (Glynn et al., 2011), el New students' Adaptive Learning Engagement in Science questionnaire (SALES) de Velayutham, Aldrige y Fraser (2011) y la prueba PISA (OCDE, 2013). Además, la revisión teórica de Osborne et al. (2003) y las revisiones de instrumentos de medición de actitud hacia la ciencia de Blalock et al. (2008) y Toma y Lederman (2020). Un antecedente clásico de la medición de las actitudes hacia la ciencia es el Test Of Science Related Attitudes (TOSRA) de Fraser (1982). Esta prueba está diseñada para estudiantes de 7 a 10 años y para su aplicación grupal. Su uso se extiende principalmente para docentes, evaluadores de currículo y para investigadores que quieran evaluar los cambios en la actitud hacia la ciencia de los estudiantes antes y después de una intervención o periodo escolar. Por otro lado, también es relevante mencionar el Protocolo de Actitudes relacionadas con la Ciencia (PAC) traducido y adaptado por Vázquez y Manassero (1997). Esta prueba consta de 50 ítems con respuesta tipo Likert organizado en cuatro subescalas: imagen global, aspectos sociales, aspectos escolares y características científicas. Otro gran antecedente, es la prueba de preferencias vocacionales Búsqueda Autodirigida SDS de Holland, Fritzsche y Powell (2005). Este instrumento es auxiliar en la toma de decisiones de los estudiantes en su elección profesional. El SDS mide intereses, actitudes y aptitudes, a la vez que las organiza en tipos de personalidades, una de estas personalidades es la I de Investigador que se asocia a intereses científicos, actitudes y aptitudes para el desenvolvimiento en ambientes de ciencia. Aunque el SDS es una prueba de preferencias vocacionales vigente, algunos reactivos dentro de la personalidad investigadora hacen referencia directa a las ciencias naturales y exactas como la física y la química. Otro antecedente relevante que se consultó para la realización de este instrumento de medición, es el capítulo relacionado a las actitudes y motivación de los estudiantes en el aprendizaje de las ciencias desarrollado por Koballa y Glynn (2007), dentro del Handbook of Research on Science Education. Asimismo, en 2011 Glynn y colaboradores desarrollaron el Science Motivation Learning Questionnaire II (SMLQ-II), mismo que Ardura y Pérez Bitrián (2018) tradujeron y adaptaron al español para la medición de la motivación en el aprendizaje de la química en estudiantes de

secundaria. También se puede mencionar el New students' Adaptive Learning Engagement in Science questionnaire (SALES) de Velayutham, Aldrige y Fraser (2011), para la medición de la autoregulación de los estudiantes en el aprendizaje de las ciencias. Además, en la investigación del desarrollo de las vocaciones científicas, se puede mencionar el trabajo de Lent et al. (2018) por sus aportaciones en el estudio de las variables asociadas a la elección de una carrera científica. Otro antecedente de medición del aprendizaje de las ciencias es el Programa Internacional de Evaluación de Alumnos (PISA, por sus siglas en inglés) una evaluación realizada por la Organización para la Cooperación y Desarrollo Económicos (OCDE). Esta prueba es aplicada a estudiantes de 15 años y mide tres componentes principales: competencia matemática, competencia lectora y competencia científica. A su vez, la competencia científica mide tanto el componente afectivo (intereses y actitudes) como el componente cognitivo de la competencia científica de los estudiantes. Este último en tres principales dominios: explicar un fenómeno científicamente, diseñar y evaluar investigación científica y analizar e interpretar datos científicamente (OCDE, 2013). Cabe agregar, que los reactivos de la competencia científica de la prueba PISA se han utilizado anteriormente para evaluar a estudiantes de educación superior. Por ejemplo, en el trabajo de Falicoff y Domínguez (2014) se realizó un estudio con el objetivo de medir el componente cognitivo de la prueba PISA en estudiantes de licenciatura inscritos en bioquímica y biotecnología. En esta evaluación, se retomaron los reactivos de PISA sin ninguna modificación y se agregaron otros reactivos diseñados para el estudio. Los investigadores precisaron que el objetivo de la medición era el componente cognitivo y no el afectivo de la competencia científica.

- **Referentes teóricos y metodológicos de la encuesta:** La EMAAAC-ESu tiene cinco grandes referentes principales: el Test Of Science Related Attitudes (TOSRA, Fraser, 1982) basado en la definición de Klopfer (1971), el constructo Contaminated Mindware (Rizeq et al., 2020), la definición de Motivación en el Aprendizaje de las Ciencias y su sistematización en el instrumento de medición Science Motivation Questionnaire II (SMQ-II, Glynn et al. 2011), el componente de Competencia Científica de la prueba PISA (OCDE, 2013) y la definición de la Autoeficacia Percibida de Bandura (1997) y Schunk y Pajares (2002).

- **Descripción de los referentes de la EMAC:** La Escala de Motivación hacia el Aprendizaje de las Ciencias (EMAC) tiene 29 ítems ($k=29$) organizados en cuatro subescalas: motivación intrínseca ($k=7$), motivación extrínseca ($k=5$), autodeterminación ($k=3$) y autoeficacia ($k=14$). La EMAC está diseñada en un formato de escala ordinal con cuatro categorías de respuesta: Totalmente en desacuerdo = 1, En desacuerdo = 2, De acuerdo = 3 y Totalmente de acuerdo = 4. En la Tabla 2 se presentan las subescalas, indicadores y posición en la encuesta de los ítems de la escala. La EMAC se desarrolló con base en el Science Motivation Questionnaire II desarrollado por Glynn y colaboradores (2011). Se define la motivación hacia el aprendizaje de las ciencias como un estado interno que despierta, dirige y sostiene la conducta del aprendizaje de las ciencias en los estudiantes (Glynn et al., 2011). En cuanto a las subescalas, se entiende la motivación intrínseca como el gusto o la satisfacción inherente por el aprendizaje de las ciencias (Simpkins et al., 2006; Glynn, et al. 2011) o bien, una tendencia natural de perseguir intereses (Deci, 1996; Reeve, 1996; Ryan y Deci, 2000 como se citó en Glynn y Koballa, 2007). Asimismo, los estudiantes que tienen una elevada motivación intrínseca no requieren de un premio o recompensa tangible, sino que el mismo proceso de aprendizaje es motivador (Glynn y Koballa, 2007). La subescala de motivación extrínseca se refiere a que el estudiante percibe el aprendizaje de las ciencias como un medio para lograr una meta u objetivo tangible, como podría ser el obtener un premio, una buena calificación, el finalizar una carrera, obtener un empleo o un posgrado (Mazlo et al., 2002, Glynn y Koballa, 2007, Glynn et al., 2011). La subescala de autodeterminación se refiere a la capacidad de tener opciones y cierto grado de control sobre las acciones que se realizan y la manera en la que se realizan (Deci et al., 2011). En este tema en particular, se refiere al control que perciben los estudiantes sobre su propio aprendizaje en contenidos científicos (Black & Deci, 2000; Glynn et al., 2011). La última subescala de autoeficacia se refiere a la confianza de los estudiantes en su propia capacidad para realizar con éxito determinadas actividades, o bien, actividades relacionadas con el aprendizaje de las ciencias (Lawson et al. 2007; Glynn et al. 2011). Para el desarrollo de la EMAC se tomaron en cuenta algunos aspectos como el contexto universitario de los estudiantes, la prudencia de las actividades académicas mencionadas en los ítems y su aplicabilidad en carreras dentro de las ciencias naturales, exactas, sociales, educativas y del comportamiento. Se procuró que las actividades de investigación no fueran atribuidas a

ningún sexo o género en específico. Se tomó en cuenta la mención y consideración de materias dentro del eje de metodología de la investigación comunes dentro de la Educación Superior (ES) en el SEM y se realizó un reajuste a la subescala de motivación extrínseca incluyendo como fines tangibles además de la obtención de una licenciatura, se consideraron el perseguir y obtener un posgrado, así como buenas oportunidades de empleo al egresar de la carrera.

- **Descripción de los referentes de la EAC:** La Escala de Actitudes hacia la Ciencia (EAC) se compone de 46 ítems (k=46) organizados en 10 subescalas: implicaciones sociales de la ciencia (k=4), normalidad de los científicos (k=5), actitud hacia la investigación (k=5), adopción de actitudes científicas (k=4), el disfrute de las experiencias de aprendizaje de las ciencias (k=5), el interés en el aprendizaje de las ciencias en tiempo de ocio (k=5), el interés por una especialización científica, o por un trabajo relacionado con la ciencia (k=6), actitudes anticientíficas (k=4), creencias conspiranoicas (k=3) y creencias paranormales (k=5). La EAC está diseñada en un formato de escala ordinal con cuatro categorías de respuesta: Totalmente en desacuerdo = 1, En desacuerdo = 2, De acuerdo = 3 y Totalmente de acuerdo = 4. En la Tabla 2 se presentan las subescalas, indicadores y posición en la encuesta de los ítems de la escala. La EAC se inspira en dos referentes principales: el instrumento de medición TOSRA (Fraser, 1982), basado en la definición de actitudes hacia la ciencia de Klopfer (1971) y el constructo de Contaminated Mindware (MC) sistematizado en el trabajo de investigación de Rizeq, Flora y Toplak (2020). Para el diseño y desarrollo de la EAC se retomaron las subescalas del TOSRA de Fraser (1982), basadas en Klopfer (1971). Dichas subescalas son siete y se definen de la siguiente manera: la subescala de implicaciones sociales de la ciencia, que mide la actitud hacia los beneficios y los problemas sociales que acompañan al progreso científico; la subescala de normalidad de los científicos, que mide las actitudes favorables y desfavorables hacia los científicos y su estilo de vida. La tercera subescala de actitud hacia la investigación, que mide el agrado hacia la experimentación científica y hacia la investigación como formas de obtener información sobre el mundo. La subescala de adopción de actitudes científicas que, de acuerdo con Fraser (1982) mide lo mismo que la tercera subescala, con excepción de que mide actitudes científicas específicas. Vázquez y Manassero Mas (1995) definen las actitudes científicas como “un conjunto de rasgos emanados de las características que el método científico

impone a las actividades de la investigación científica realizadas por los científicos” (p. 341). En la EAC se tomaron en cuenta la objetividad, el pensamiento crítico y el escepticismo, de acuerdo con la propuesta de Vázquez y Manassero Mas (1995). La subescala del disfrute de las experiencias de aprendizaje de las ciencias, que mide el gusto de los estudiantes por el aprendizaje de las ciencias en un entorno más académico; la subescala del interés en el aprendizaje de las ciencias en el tiempo libre, que mide el interés de los estudiantes por aprender ciencias en el tiempo libre, de ocio o como entretenimiento y el interés en una especialización científica, que mide el interés de los estudiantes por una especialización en investigación científica y en la investigación como opción laboral. Las tres subescalas subsecuentes, se basan en el trabajo de Rizeq, Flora y Toplak (2020) sobre el Contaminated Mindware (CM) de Stanovich (2009). Para el diseño y desarrollo de la EAC se tomaron en cuenta tres subescalas que son: las actitudes anticientíficas, las creencias conspiranoicas y las creencias paranormales. Se definen las actitudes anticientíficas como una postura de oposición y rechazo ante los métodos de la ciencia (Holton, 1993) así como la percepción de que los métodos científicos son poco confiables (Hartman, 2017). La subescala de creencias conspiranoicas se refiere a suposiciones o conjeturas innecesarias de conspiración aun cuando otras causas son más probables (Aaronovitch, 2009 como se citó en Rizeq et al., 2020). La última subescala de creencias paranormales se entiende como la creencia en fenómenos que violan la comprensión científica actual (Tobacyk y Milford, 1983) como las creencias en psicoquinesis, brujería, supersticiones, espiritualismo (por ejemplo, la comunicación con personas fallecidas) y la precognición como la creencia en la astrología (Rizeq et al., 2020). Estas últimas subescalas se desarrollaron con la finalidad de obtener evidencias de validez discriminantes. Para el diseño y desarrollo de los ítems de la EAC se tomaron en cuenta algunos aspectos relevantes como el contexto universitario de los estudiantes, la prudencia de las actividades académicas mencionadas en los ítems y su aplicabilidad en carreras dentro de las ciencias naturales, exactas, sociales, educativas y del comportamiento. Se procuró que las actividades de investigación no fueran atribuidas a ningún sexo o género en específico. Se tomó en cuenta la mención y consideración de materias dentro del eje de metodología de la investigación comunes dentro de la Educación Superior (ES) en el SEM y la inclusión de actividades y formatos comúnmente utilizados por los estudiantes como el uso de redes sociales y plataformas de entretenimiento.

- Descripción de los referentes de la EAPAMC:** La Escala de Autoeficacia Percibida en el Aprendizaje del Método Científico (EAPAMC) se compone de 26k organizados en tres subescalas: explicar un fenómeno científicamente ($k=7$), diseñar y evaluar investigación científica ($k=5$) y analizar e interpretar datos científicamente ($k=5$). La EAPAMC está diseñada en un formato de escala ordinal con cuatro categorías de respuesta: Nada en absoluto = 1, Poco = 2, Moderado = 3 y Fuerte = 4. En la Tabla 3 se presentan las subescalas, indicadores y posición en la encuesta de los ítems de la escala. La EAPAMC se basa en dos referentes teóricos principales: el componente de competencia o alfabetización científicas de la prueba PISA (2013) y la definición de Autoeficacia Percibida de Bandura (1997). Dicho autor, define la autoeficacia como la creencia en la propia capacidad de organizar y ejecutar las acciones necesarias para alcanzar los logros deseados. En el aprendizaje de las ciencias, de acuerdo con Glynn et al. (2007) se refiere a la evaluación que el estudiante realiza acerca de sus propias capacidades y competencias personales para desenvolverse en el campo científico. Cabe mencionar también a Hemmings y Kay (2010) que definen la autoeficacia en la investigación como la percepción o estimación personal de qué tan bien puede uno ejecutar un conjunto de tareas relacionadas con la investigación científica. En la prueba PISA (OCDE, 2013) se define la Science Literacy o Alfabetización Científica por su traducción al español como la capacidad para involucrarse en ideas, temas y debates de ciencia y tecnología de manera activa y reflexiva. Para ser partícipe de temas de ciencia y tecnología, la OCDE señala que se requieren tres competencias: explicar un fenómeno científicamente, diseñar y evaluar investigación científica y analizar e interpretar datos científicamente (OCDE, 2013). Cada una de estas competencias tiene indicadores específicos de lo que sería ideal que los estudiantes pudieran realizar a una edad específica. En el desarrollo de la EAPAMC se retomaron los indicadores de estas competencias de la prueba PISA con la finalidad de explorar el grado de dominio que los estudiantes perciben en su propia ejecución de ejercicios relacionados a la actividad científica. Cabe aclarar que se obviaron las dimensiones con respecto al conocimiento del contenido, procedimiento y epistemología, así como las propuestas de contexto de aplicación que propone la prueba PISA para evaluar estas competencias.
- Enfoque de medición:** El enfoque de la medición parte desde una postura realista-constructivista de la validez llamada validez unitaria o validez de constructo propuesta por

Messick (1995). Dicho autor define la validez unitaria o validez de constructo como el grado en que la evidencia empírica y los fundamentos teóricos apoyan la idoneidad y la pertinencia de las interpretaciones y acciones sobre la base de los puntajes de las pruebas u otros modos de evaluación (Messick, 1989b como se citó en Messick, 1995). Para el diseño de la encuesta se consideraron los Estándares para las pruebas educativas y psicológicas (AERA, APA, NCME, 2014). En este documento se señala que la validez de una prueba hace referencia “al grado en que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba para usos propuestos de las pruebas” (pp. 11). Por lo tanto, el proceso de validación consiste en la recopilación de evidencias con base científica que permita hacer interpretaciones sobre los puntajes de las pruebas, de manera que la evidencia respalda el uso que se le dará a la prueba (AERA, APA, NCME, 2014).

Apéndice II.

Manual de procedimientos para la validación del contenido y del diseño y equidad de la EMAAAC-ESu

Estimado miembro del Comité de Validación de la Encuesta (CVE):

Este es un breve instructivo sobre el procedimiento de recopilación de evidencias de validez de constructo en el aspecto del contenido y del diseño y equidad de las escalas de la **Encuesta de Motivación, Actitudes y Autoeficacia en el Aprendizaje de las Ciencias para estudiantes de Educación Superior (EMAAAC-ESu)**.

PROCEDIMIENTO

Paso 1. Primero, es importante que se familiarice con los cuatro criterios para la validación del contenido con los que se calcula el Índice de Validez de Contenido (IVC, Lawshe, 1975; Pedrosa et al., 2013) (ver tabla 1); y con los siete elementos de la Evaluación de Diseño Universal (EDU) propuestos por Thompson, Johnstone y Thurlow (pp. 5-20, 2002) y adaptados por Pérez-Morán (pp.122-123, apéndice 3, 2014) (ver tabla 2).

Tabla 1. Criterios para la validación del contenido para el cálculo del IVC

No.	Criterio	Descripción
1	Relevancia	<i>¿El ítem es importante para medir aspectos representativos y sustantivos de los dominios, constructos o temas del instrumento?</i>
2	Precisión	<i>¿El ítem es exacto para medir el dominio, constructo o tema de interés sin presentar algún tipo de sesgo o estar contaminado por otro rasgo diferente al constructo de interés?</i>
3	Claridad	<i>¿El ítem es claro, legible y comprensible en todos los elementos que lo componen?</i>
4	Simplicidad	<i>¿El ítem es sencillo y simple, conteniendo lo esencial para ser respondido?</i>

Tabla 2. Elementos de la Evaluación de Diseño Universal (EDU) propuestos por Thompson, Johnstone y Thurlow (2002) y adaptados por Pérez-Morán (2014)

No.	Elemento	Descripción
1	Definición precisa de constructo	<p><i>¿Se encuentra bien definido y delimitado el rasgo a medir del constructo en cada uno de los ítems y escalas en su conjunto?</i></p> <p>Este elemento o principio del EDU se refiere a la definición clara y precisa del constructo, mismo que señala que en el proceso de diseño y desarrollo de pruebas se debe definir de manera cuidadosa el constructo que se quiere medir. Asimismo, se trata de evitar todas aquellas barreras cognitivas, sensoriales, físicas y emocionales que no son relevantes para el constructo, es decir, aquellos aspectos que representan varianza irrelevante.</p>
2	Inclusión poblacional	<p><i>¿Los ítems y las escalas en su conjunto son entendibles para diversas poblaciones?</i></p> <p>Este elemento o principio del EDU se refiere a la inclusión poblacional, mismo que señala que todos los estudiantes evaluados deben tener la misma oportunidad de demostrar su posición con respecto al constructo que el instrumento quiere medir, sin importar aspectos propios del sustentante e irrelevantes para el constructo como las habilidades cognitivas, el idioma o los antecedentes culturales.</p>
3	Accesibilidad e imparcialidad	<p><i>¿Los ítems y escalas se pueden adaptar a otros formatos para aplicarse a otras poblaciones?</i></p> <p>Este elemento o principio del EDU se refiere a la imparcialidad de los ítems, mismo que señala que los ítems deben ser accesibles y libres de ambigüedades. Por ejemplo, <i>¿el ítem y las escalas dependen de imágenes que pudieran ser una complicación irrelevante para una persona con debilidad visual?</i></p>
4	Acomodación flexible de los contenidos	<p><i>¿Los ítems y escalas presentan contenidos adecuados de acuerdo con el idioma, edad y nivel educativo de los participantes?</i></p> <p>Este elemento o principio del EDU señala que las pruebas deben facilitar las adaptaciones posteriores. Es decir, no deben contener gráficos o imágenes irrelevantes, textos diagonales o verticales, ítems que dependan de imágenes para resolverse o imágenes o texto irrelevante para el constructo que se desea medir.</p>
5	Procedimientos e instrucciones simples, claras e intuitivas	<p><i>¿Las instrucciones de las escalas son sencillas, claras e intuitivas?</i></p> <p>Este elemento o principio del EDU indica que las instrucciones y procedimientos deben ser fáciles de entender para los participantes independientemente de su dominio del idioma, nivel de desarrollo o nivel de concentración en el que se encuentren.</p>
6	Comprensibilidad	<p><i>¿El contenido de la base de los ítems es comprensible para la diversidad de posibles sustentantes?</i></p>

		Este elemento o principio del EDU señala que en los ítems se deberá cuidar que las oraciones no sean muy largas o confusas y que las palabras utilizadas sean familiares para quien las lee. Es decir, la respuesta no deberá ser influenciada por la falta de comprensión de las instrucciones del instrumento.
7	Máxima legibilidad	<p><i>¿El texto de la base de los ítems y de las instrucciones de las escalas es legible para la diversidad de posibles sustentantes?</i></p> <p>Este elemento o principio del EDU señala la importancia de que los elementos de la prueba sean fácilmente descifrables (legibles) para los sustentantes. Este elemento es aplicable también para figuras y tablas y formatos en el que se insertan los ítems.</p>

Paso 2. Consecutivamente, revise los documentos en formato PDF: **2) Ficha técnica del EMAAAC-ESu** y **3) Formato de aplicación EMAAAC-ESu** que se encuentran en la carpeta de Google Drive de trabajo del Comité de Validación de la Encuesta (CVE) (<https://drive.google.com/drive/folders/1FauXsT-2To3RKwLO1zti2BUY2abOdKiX?usp=sharing>).

La **Ficha Técnica de la EMAAAC-ESu** contiene: (a) la descripción de los usos y propósitos de la encuesta, (b) los antecedentes del estudio y medición de los constructos de interés, (c) los fundamentos teóricos y metodológicos, (d) el modelo de medición utilizado, (e) y las especificaciones de la estructura y de los ítems de la **EMAAAC-ESu** con la descripción de las subescalas, los indicadores, el número de posición de los reactivos y los tipos y códigos de respuesta.

Paso 3. Inserte comentarios y sugerencias de mejora en el **3) Formato de aplicación EMAAAC-ESu**. En especial, considere los cuatro criterios para el cálculo del IVC y los siete elementos del EDU al momento de comentar los ítems que presentan errores o problemas en alguno de los criterios señalados. Cualquier otro comentario que considere relevante para mejorar la calidad técnica de las escalas y de los distintos elementos de la **EMAAAC-ESu** será bienvenido. Es importante señalar que en este paso otros miembros del panel de jueces tendrán acceso a ambos documentos y comentarán bajo los mismos criterios enriqueciendo la información para los propósitos del estudio.

Paso 4. Inicie la evaluación de los ítems de las escalas. Para ello, ingrese al *Protocolo de validación del contenido de las escalas de la EMAAAC-ESU* haciendo clic en el vínculo de **Google Forms**: <https://forms.gle/749XjuZextCQD8hN7>. En este protocolo de validación, usted deberá asignar una

puntuación del 1 a 4 a cada uno de los ítems de la **EMAAAC-ESu** según considere el nivel de calidad técnica con base en los cuatro criterios para el cálculo del IVC:

- Relevancia: 1=Irrelevante, 2=Poco relevante, 3= Relevante, y 4=Muy relevante
- Precisión: 1 =Impreciso, 2=Poco preciso, 3=Preciso, y 4=Muy preciso
- Claridad: 1=Confuso, 2=Poco claro, 3=Claro, y 4=Muy claro
- Simplicidad: 1=Complejo, 2=Poco simple, 3=Simple, 4=Muy simple

Por ejemplo, si se estuviera evaluando la relevancia de un ítem el 1 representaría que el ítem es irrelevante y requiere una revisión mayor o en su caso eliminarlo, el 2 que el ítem necesita una revisión a profundidad o requiere una revisión menor, el 3 que el ítem es relevante quizás con algunas leves mejoras, y el 4 que el ítem es muy relevante para medir el constructo en cuestión y no requiere de ninguna mejora.

Paso 5. Una vez finalizada la validación de todos los reactivos de las escalas de la **EMAAAC-ESu** notifique por correo electrónico (lucia.marquez@uabc.edu.mx) a la Lic. Marla Márquez que concluyó la emisión de sugerencias y el protocolo de validación de contenido de la encuesta

