

PERCEPCIONES DE LOS ESTUDIANTES SOBRE LA COMPETENCIA DE LOS PROFESORES DE ENSEÑANZA DEL INGLÉS (ELT) EN LA PROMOCIÓN DE HABILIDADES 6C: UN ESTUDIO DE DESARROLLO Y VALIDACIÓN DE ESCALA

PERCEPÇÕES DOS ESTUDANTES SOBRE A COMPETÊNCIA DE PROFESSORES DE ENSINO DE INGLÊS (ELT) NA PROMOÇÃO DE HABILIDADES 6C: UM ESTUDO DE DESENVOLVIMENTO E VALIDAÇÃO DE ESCALA

STUDENTS' PERCEPTIONS OF ELT TEACHERS' COMPETENCE IN PROMOTING 6C SKILLS: A SCALE DEVELOPMENT AND VALIDATION STUDY



Birsen TÜTÜNIŞ¹
e-mail: tutunisster@gmail.com



Kerim ÜNAL²
e-mail: kerimunal@mersin.edu.tr



Taner HOSSEINI³
e-mail: t.hosseini@iku.edu.tr



Özge KÖKSAL⁴
e-mail: o.koksal@iku.edu.tr



Ata ALKAYALAR⁵
e-mail: a.alkayalar@iku.edu.tr

Cómo citar este artículo:

Tütüniş, B., Ünal, K., Hosseini, T., Köksal, Ö., & Alkayalar, A. (2026). Percepciones de los estudiantes sobre la competencia de los profesores de enseñanza del inglés (ELT) en la promoción de habilidades 6C: un estudio de desarrollo y validación de escala. *Revista on line de Política e Gestão Educacional*, 30(esp1), e026016. <https://doi.org/10.22633/rpge.v30iesp1.20904>



| Enviado el: 10/01/2026
| Revisiones requeridas el: 16/01/2026
| Aprobado el: 05/02/2026
| Publicado el: 30/03/2026

Editor: Prof. Dr. Sebastião de Souza Lemes
Editor ejecutivo adjunto: Prof. Dr. José Anderson Santos Cruz

¹Universidad Kültür de Estambul, Campus Ataköy, Estambul y 34158, Türkiye.

²Universidad de Mersin, campus de Çiftlikköy, Mersin y 33110, Türkiye.

³Universidad Kültür de Estambul, Campus Ataköy, Estambul y 34158, Türkiye.

⁴Universidad Kültür de Estambul, Campus Ataköy, Estambul y 34158, Türkiye.

⁵Universidad Kültür de Estambul, Campus Ataköy, Estambul y 34158, Türkiye.

RESUMEN: Este estudio desarrolló y validó una escala para medir las percepciones de los estudiantes sobre la competencia de sus profesores en la implementación de las habilidades 6C: Pensamiento Crítico, Comunicación, Colaboración, Creatividad, Ciudadanía y Carácter. Dos muestras independientes de estudiantes participaron en el proceso de validación: 302 estudiantes para el análisis factorial exploratorio (AFE) y 319 para el análisis factorial confirmatorio (AFC). Los estudiantes abarcaron desde 5.º hasta 11.º grado en escuelas secundarias de Estambul. Los análisis estadísticos redujeron la escala a 37 ítems organizados en seis factores correspondientes al modelo 6C. Los resultados mostraron sólidas propiedades psicométricas. La fiabilidad global fue muy alta ($\alpha = 0,962$), y las fiabilidades de los factores individuales oscilaron entre 0,866 y 0,923. El análisis factorial confirmatorio demostró un excelente ajuste del modelo (RMSEA = 0,040; CFI = 0,959; TLI = 0,955). Un modelo de segundo orden confirmó que las seis habilidades pueden combinarse en un único constructo general de competencias del siglo XXI.

PALABRAS CLAVE: Habilidades 6C. Competencias del siglo XXI. Desarrollo de escala. Evaluación de los estudiantes. Eficacia docente.

RESUMO: Este estudo desenvolveu e validou uma escala para medir as percepções dos estudantes sobre a competência de seus professores na implementação das habilidades 6C: Pensamento Crítico, Comunicação, Colaboração, Criatividade, Cidadania e Caráter. Duas amostras distintas de estudantes participaram do processo de validação: 302 estudantes para a análise fatorial exploratória (AFE) e 319 para a análise fatorial confirmatória (AFC). Os estudantes variaram do 5º ao 11º ano em escolas secundárias de Istambul. As análises estatísticas reduziram a escala a 37 itens organizados em seis fatores correspondentes ao modelo 6C. Os resultados evidenciaram fortes propriedades psicométricas. A confiabilidade geral foi muito alta ($\alpha = 0,962$), e as confiabilidades dos fatores individuais variaram de 0,866 a 0,923. A análise fatorial confirmatória demonstrou excelente ajuste do modelo (RMSEA = 0,040; CFI = 0,959; TLI = 0,955). Um modelo de segunda ordem confirmou que as seis habilidades podem ser combinadas em um único construto geral de competências do século XXI.

PALAVRAS-CHAVE: Habilidades 6C. Competências do século XXI. Desenvolvimento de escala. Avaliação dos estudantes. Eficácia docente.

ABSTRACT: This study developed and validated a scale to measure students' perceptions of their teachers' competence in implementing 6C skills: Critical Thinking, Communication, Collaboration, Creativity, Citizenship, and Character. Two separate student samples participated in the validation process: 302 students for an exploratory factor analysis (EFA) and 319 for a confirmatory factor analysis (CFA). The students ranged in grade from 5th to 11th in Istanbul secondary schools. Statistical analyses reduced the scale to 37 items organized into six factors corresponding to the 6C framework. Results showed strong psychometric properties. The overall reliability was very high ($\alpha = 0.962$), and individual factor reliabilities ranged from 0.866 to 0.923. Confirmatory factor analysis demonstrated excellent model fit (RMSEA = 0.040, CFI = 0.959, TLI = 0.955). A second-order model confirmed that the six skills could be combined into one overall 21st-century skills construct.

KEYWORDS: 6C skills. 21st-century competencies. Scale development. Student evaluation. Teacher effectiveness.

INTRODUCCIÓN

Los rápidos cambios en la educación y las exigencias del aprendizaje del siglo XXI han incrementado la necesidad de docentes capaces de integrar eficazmente las seis competencias clave (pensamiento crítico, comunicación, colaboración, creatividad, ciudadanía y carácter) en la práctica docente. Si bien numerosos estudios analizan estas competencias, pocos han investigado la percepción del alumnado sobre la eficacia con la que los docentes las implementan. Esta brecha pone de manifiesto la necesidad de una escala fiable y válida que mida la competencia docente desde la perspectiva del alumnado. Por consiguiente, este estudio tiene como objetivo desarrollar una escala que capture la percepción del alumnado sobre la eficacia de los docentes en la promoción y aplicación de las seis competencias clave en el aula.

El siglo XXI ha traído consigo cambios tecnológicos y sociales que desafían fundamentalmente los paradigmas educativos tradicionales. A diferencia de las reformas educativas anteriores, centradas principalmente en el contenido curricular o los métodos de enseñanza, la transformación educativa contemporánea exige una reconceptualización completa de los objetivos de aprendizaje y los enfoques pedagógicos (Saavedra & Opfer, 2012). El cambio de la educación de la era industrial —caracterizada por la estandarización, la compartimentación y la transmisión del conocimiento— al aprendizaje de la era digital requiere el desarrollo de estudiantes capaces de desenvolverse en la complejidad, colaborar en diversos contextos, comunicarse eficazmente a través de múltiples modalidades, pensar críticamente sobre la abundancia de información, crear soluciones innovadoras y demostrar integridad en la toma de decisiones éticas.

Esta transformación refleja un reconocimiento social más amplio de que el rendimiento académico por sí solo es insuficiente para el éxito en entornos globales cada vez más complejos. Los desafíos modernos —desde el cambio climático y la desigualdad social hasta el rápido avance tecnológico y la proliferación de información— requieren ciudadanos capaces de resolver problemas de forma colaborativa, razonar éticamente y desarrollar un pensamiento adaptativo (Harari, 2018). En consecuencia, las instituciones educativas de todo el mundo están replanteando su propósito fundamental, pasando de preparar a los estudiantes para trayectorias profesionales predeterminadas a formar individuos adaptables, reflexivos e innovadores, capaces de generar valor en futuros impredecibles.

Entre los diversos marcos que abordan las necesidades educativas del siglo XXI, el modelo 6C de Michael Fullan representa un enfoque integral y con fundamentos pedagógicos

para la transformación educativa. A diferencia de los marcos que compartimentan las habilidades en categorías discretas, el modelo 6C —que abarca Carácter, Ciudadanía, Colaboración, Comunicación, Creatividad y Pensamiento Crítico— conceptualiza estas competencias como dimensiones interconectadas de un aprendizaje profundo que integra el desarrollo cognitivo, interpersonal e intrapersonal (Fullan & Scott, 2014a).

El marco teórico se fundamenta en la extensa investigación de Fullan sobre el cambio educativo y la mejora escolar, sintetizando décadas de ciencia de la implementación con una comprensión contemporánea de los procesos de aprendizaje. Esta base empírica distingue al modelo 6C de otros marcos más teóricos, ya que surgió de la observación directa de iniciativas exitosas de transformación educativa, en lugar de ejercicios abstractos de mapeo de competencias (Fullan et al., 2022). El énfasis del modelo en el “aprendizaje profundo” —un aprendizaje que involucra a los estudiantes intelectual, emocional y socialmente, desarrollando tanto conocimientos académicos como habilidades para la vida— refleja una comprensión sofisticada de cómo se produce un aprendizaje significativo en entornos complejos y tecnológicamente avanzados.

Además, la adopción global del marco 6C en diversos contextos educativos sugiere su validez transcultural y aplicabilidad práctica. Sistemas educativos en países que van desde Canadá y Australia hasta Finlandia y Singapur han implementado con éxito iniciativas basadas en el marco 6C, demostrando la adaptabilidad del marco a diferentes contextos culturales e institucionales, manteniendo al mismo tiempo sus principios pedagógicos fundamentales.

A pesar del creciente reconocimiento de la importancia del marco de las 6C y su cada vez mayor adopción en la práctica educativa, existen importantes deficiencias en nuestra capacidad para evaluar la competencia docente en la implementación efectiva de estas competencias. Los enfoques de evaluación actuales se basan principalmente en la autoevaluación del profesorado o en la observación externa, lo que omite la perspectiva crítica de los estudiantes que experimentan directamente estos enfoques pedagógicos en contextos de aula auténticos. Esta limitación resulta particularmente problemática, dado que las habilidades del siglo XXI son inherentemente experienciales y relacionales; su efectividad no depende de las intenciones del profesorado ni de las percepciones externas, sino de si los estudiantes desarrollan realmente estas competencias a través de las interacciones en el aula (Fullan & Scott, 2014b).

La ausencia de instrumentos validados para captar las percepciones de los estudiantes sobre la implementación de las 6C por parte de los docentes genera múltiples problemas para

la práctica y la investigación educativa. A nivel profesional, los docentes carecen de retroalimentación sistemática sobre su eficacia en el fomento de las habilidades del siglo XXI, lo que limita su capacidad de mejora específica. A nivel institucional, las escuelas no pueden evaluar de forma fiable la calidad de la implementación de las 6C en las aulas ni el impacto de las iniciativas de desarrollo profesional. Empíricamente, los investigadores no pueden estudiar las relaciones entre las prácticas docentes y el desarrollo de las competencias del siglo XXI en los estudiantes sin herramientas de medición válidas.

Este estudio aborda estas deficiencias cruciales mediante el desarrollo y la validación de una escala integral para medir la percepción que tienen los estudiantes de secundaria sobre la competencia de sus docentes en la implementación y promoción de las habilidades 6C. Al proporcionar el primer instrumento psicométricamente sólido diseñado específicamente para este fin, la investigación aporta tanto conocimiento teórico como herramientas prácticas para mejorar la práctica educativa del siglo XXI.

Estudios previos han explorado diversos marcos para las habilidades del siglo XXI, como el Marco P21, ATC21S y el modelo de Aprendizaje Profundo de Fullan. Sin embargo, la mayoría de los instrumentos existentes se centran en las competencias de los estudiantes o en la autoevaluación del profesorado. La investigación sobre las evaluaciones que hacen los estudiantes del desempeño de los docentes en las 6C es limitada. Además, las escalas anteriores varían considerablemente en su estructura, lo que dificulta su uso en diferentes contextos culturales. Este estudio aborda estas deficiencias mediante el diseño de una escala basada en el estudiante que se alinea con los modelos internacionales de las 6C y, al mismo tiempo, refleja la realidad del aula.

Partiendo de los fundamentos conceptuales y la brecha empírica descrita anteriormente, el presente estudio parte de la premisa de que las experiencias directas de los estudiantes en el aula constituyen una perspectiva esencial para evaluar de manera significativa la implementación de las competencias del siglo XXI por parte del profesorado. En consonancia con esta visión, el objetivo central de la investigación es construir y validar un instrumento de medición que capture las percepciones de los estudiantes de secundaria sobre la eficacia de sus docentes en la aplicación de los seis dominios del marco de las 6C. Se espera que el desarrollo de dicha herramienta ofrezca tanto fundamentos teóricos —al clarificar cómo se experimentan

las 6C en la práctica— como valor práctico para la retroalimentación en el aula, la mejora de los programas y la investigación futura.

Preguntas de investigación

En consecuencia, el estudio aborda las siguientes preguntas de investigación:

- RQ1: ¿Hasta qué punto se puede desarrollar un instrumento basado en el estudiante que mida de manera confiable la implementación de las seis competencias 6C por parte de los docentes desde el punto de vista de los estudiantes?
- RQ2: ¿El análisis factorial exploratorio respalda una estructura de seis factores coherente con las dimensiones teóricas del marco 6C?
- RQ3: ¿En qué medida los factores resultantes demuestran una consistencia interna y una validez convergente aceptables, de acuerdo con los criterios psicométricos establecidos?
- RQ4: ¿El análisis factorial confirmatorio valida aún más el modelo de seis factores y proporciona evidencia de un constructo de segundo orden que representa la competencia docente general del siglo XXI?

Estas cuestiones se examinan mediante un proceso de desarrollo de escalas en varias etapas, que comprende la generación de ítems a partir de datos de grupos focales, la revisión por expertos para la validación del contenido, el análisis factorial exploratorio para identificar la estructura latente y el análisis factorial confirmatorio para establecer la coherencia estructural. Este diseño secuencial permite contrastar los componentes teóricos del modelo 6C con indicadores empíricos derivados de las evaluaciones de los estudiantes, lo que garantiza na

correspondencia directa entre el marco conceptual, el enfoque metodológico y la contribución prevista del estudio.

MÉTODO

El desarrollo de la Escala de Evaluación Estudiantil de las Habilidades Docentes del Siglo XXI (6C) siguió un proceso sistemático de cinco fases diseñado para garantizar la validez del contenido, la alineación teórica y la aplicabilidad práctica en entornos educativos.

Proceso de desarrollo de escalas (diseño de investigación)

Fase 1: Revisión de la literatura y fundamentos teóricos

Se realizó una revisión exhaustiva de la investigación existente sobre los marcos de habilidades del siglo XXI, con especial énfasis en el modelo 6C, que abarca el pensamiento crítico, la creatividad, la colaboración, la comunicación, el carácter y la ciudadanía. Esta revisión estableció fundamentos teóricos e identificó indicadores conductuales clave para cada dominio de habilidades que pueden observarse en el aula. El análisis de la literatura reveló deficiencias en los instrumentos de evaluación desde la perspectiva del estudiante, lo que respalda la necesidad de una escala que capture las prácticas de implementación de los docentes desde el punto de vista del alumno.

Fase 2: Recopilación de datos cualitativos mediante grupos focales

Se realizó un estudio de grupo focal con 15 docentes seleccionados aleatoriamente de instituciones de educación secundaria. Los participantes completaron un protocolo de entrevista estructurada que constaba de 30 preguntas abiertas, cinco de las cuales estaban diseñadas específicamente para obtener descripciones detalladas de cómo se manifiesta cada una de las seis habilidades clave en la práctica docente. Las preguntas de la entrevista abarcaron tanto

comportamientos de enseñanza explícitos como estrategias de instrucción implícitas que promueven el desarrollo de las habilidades del siglo XXI.

Fase 3: Generación y categorización inicial de los elementos.

El análisis temático de las respuestas del grupo focal arrojó 160 ítems preliminares de la escala, categorizados sistemáticamente en los seis dominios de habilidades. La distribución inicial reflejó la complejidad y el alcance de cada área de competencia. Comunicación y Pensamiento Crítico generaron 30 ítems cada uno, representando las categorías más extensas. Creatividad y Carácter aportaron 29 ítems cada uno, mientras que Colaboración generó 24. Ciudadanía, si bien es importante, generó la menor cantidad de ítems, con 18, lo que refleja su alcance conceptual más específico dentro del contexto del aula.

El proceso de generación de ítems priorizó la especificidad conductual y la observabilidad en el aula. Cada ítem se diseñó para describir acciones concretas del profesor que los alumnos pudieran observar y evaluar fácilmente, evitando conceptos abstractos o estados mentales internos que serían difíciles de evaluar con precisión.

Fase 4: Análisis de contenido y perfeccionamiento de los elementos

Una fase rigurosa de análisis de contenido implicó una revisión sistemática por parte del equipo de investigación para eliminar elementos redundantes, ambiguos o teóricamente desalineados. Este proceso aplicó varios criterios de exclusión: (a) superposición semántica con elementos existentes, (b) alineación insuficiente con el dominio de la habilidad objetivo, (c) potencial de carga cruzada en múltiples dominios y (d) complejidad inapropiada para el público estudiantil al que va dirigido.

El proceso de refinamiento redujo el conjunto de ítems de 160 a 95, con tasas de eliminación que variaron según el dominio. La comunicación experimentó la reducción más significativa, con 15 ítems eliminados y 15 restantes, lo que representa la tasa de reducción más alta debido a la redundancia de ítems que abordan comportamientos comunicativos similares.

Pensamiento Crítico tuvo 14 ítems eliminados, conservando 16, mientras que Creatividad vio 12 ítems eliminados, quedando 17. Carácter tuvo 13 ítems eliminados y 16 conservados. Colaboración experimentó una reducción moderada con 8 ítems eliminados y 16 restantes. Ciudadanía mostró la tasa de eliminación más baja con solo 3 ítems eliminados y 15 restantes, lo que sugiere que los ítems iniciales para este dominio estaban bien diferenciados y

eran teóricamente distintos. En total, se eliminaron 65 ítems del conjunto original, resultando en 95 ítems para su posterior desarrollo.

Fase 5: Validación por expertos y selección final de los elementos.

La escala refinada de 95 ítems fue sometida a una validación exhaustiva por parte de seis especialistas: dos expertos en medición y evaluación, dos expertos en currículo e instrucción y tres expertos en enseñanza del idioma inglés. Este enfoque multidisciplinario garantizó la evaluación desde perspectivas tanto psicométricas como de contenido.

Los expertos evaluaron cada ítem mediante una escala Likert de 10 puntos, que iba de 1 (totalmente inapropiado) a 10 (totalmente apropiado). Los criterios de evaluación incluyeron la relevancia del contenido, la adecuación a la edad de los estudiantes de secundaria, la claridad del lenguaje y la alineación con los marcos teóricos. Los ítems con una puntuación media inferior a 7,0 se eliminaron sistemáticamente, garantizando así que solo permanecieran aquellos con un fuerte consenso entre los expertos.

Participantes

El estudio incluyó dos muestras independientes para el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC). La muestra para el AFE constó de 302 participantes, mientras que la muestra para el AFC incluyó 319 participantes. Las características demográficas de ambas muestras se presentan en la Tabla 1.

RESULTADO

Análisis preliminares

Antes de realizar los análisis factoriales, se llevó a cabo una serie de comprobaciones preliminares para evaluar si el conjunto de datos cumplía los supuestos necesarios para el análisis multivariante. Las estadísticas descriptivas y las pruebas de diagnóstico no mostraron evidencia de violaciones relacionadas con la normalidad, la multicolinealidad o la adecuación del muestreo. Además, el índice de Kaiser-Meyer-Olkin y la prueba de esfericidad de Bartlett indicaron que los datos eran adecuados para los procedimientos de análisis factorial. Los resultados de estos análisis preliminares se presentan en las Tablas 1 a 4. Las estadísticas descriptivas detalladas y las comprobaciones de supuestos, incluidos los índices de normalidad

y las intercorrelaciones, se presentan en las Tablas 2 a 4. Las características de los participantes para las muestras de AFE y AFC se presentan en la Tabla 1.

Tabla 1.

Características demográficas de los participantes en el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC)

Variable	Categoría	AFE (n)	AFE (%)	AFC (n)	AFC (%)
Género	Femenino	173	57.3	179	56.1
	Masculino	129	42.7	140	43.9
Edad	12	15	5.0	24	7.5
	13	29	9.6	38	11.9
	14	63	20.9	43	13.5
	15	76	25.2	78	24.5
	16	41	13.6	71	22.3
	17	36	11.9	29	9.1
	18	42	13.9	18	5.6

Nota. Resultados del autor.

Ambas muestras fueron similares en cuanto a la distribución por género y el rango de edad (12-18 años), con la mayoría de los participantes entre 14 y 16 años. Los niveles escolares abarcaron desde el quinto hasta el undécimo grado, con una distribución bastante uniforme entre los niveles de secundaria y bachillerato.

Procedimientos de recopilación de datos

Este estudio se realizó como parte de un Proyecto de Investigación Científica (PIC) financiado por la Universidad Kültür de Estambul. Se obtuvo la aprobación ética del Comité de Ética de la Universidad Kültür de Estambul y, antes de la recopilación de datos, la Dirección Provincial de Educación Nacional de Estambul autorizó la realización de la encuesta en las escuelas secundarias.

Los datos se recopilaron durante el semestre de primavera del curso 2024-2025 en escuelas secundarias públicas de Estambul. La escala de 60 ítems fue administrada en formato impreso por los profesores durante las clases habituales, y cada administración duró aproximadamente 25 minutos. Para garantizar la estandarización de los procedimientos de administración, un asistente de investigación impartió formación detallada a los profesores, explicándoles cómo administrar la escala y qué instrucciones dar a los alumnos. Los profesores

contaron con guiones estandarizados para mantener la coherencia en todas las aulas participantes.

Antes de la recopilación de datos, se informó a los estudiantes que su participación era voluntaria. Solo aquellos que aceptaron participar voluntariamente completaron los formularios de la escala. Para garantizar la confidencialidad, no se les solicitó información personal que los identificara, y la recopilación de datos demográficos se limitó a edad, género y nivel de grado. Se les aseguró a los estudiantes que sus respuestas serían confidenciales y que su participación o no participación no afectaría su rendimiento académico. Los formularios completados se recogieron inmediatamente después de su administración para preservar la integridad de los datos y el anonimato de los participantes.

PROCEDIMIENTOS DE ANÁLISIS DE DATOS

Prueba de supuestos

Antes de realizar el análisis factorial exploratorio, un análisis multivariado de los datos obtenidos, se probaron los supuestos básicos de este análisis. Inicialmente, se examinó la adecuación del tamaño de la muestra y la presencia de datos faltantes. Respecto al tamaño de la muestra, Kline enfatizó que podría ser 10 veces el número de ítems, pero debería ser al menos el doble (1994). Algunas fuentes (Büyüköztürk, 2006; Tabachnick & Fidell, 2001) afirman que el tamaño de la muestra debería ser al menos cinco veces el número total de ítems. Con base en esto, nuestra escala desarrollada contiene un total de 60 ítems, y cinco veces ese número, 300 observaciones, serían suficientes. En este estudio, se obtuvieron 312 observaciones; 10 fueron excluidas del análisis, quedando 302 para el análisis. Esto puede ser evidencia de que los datos son suficientes para el análisis factorial. No hay valores faltantes en los datos obtenidos en el estudio.

En segundo lugar, se examinó la normalidad de la distribución en el conjunto de datos obtenido. Se verificaron los valores de curtosis y asimetría de todos los datos. Se observó que los valores de asimetría ($-0,092 / 0,138 = -0,667$) y curtosis ($-0,956 / 0,275 = -3,476$) cumplían con el supuesto de distribución normal. Al examinar el valor Z, este se encuentra entre -1,913 y 1,781, y dado que el 95% de las observaciones se encuentran dentro del rango de -1,96 a +1,96, se puede afirmar que la distribución es normal (Alpar, 2012:140). Al evaluar los resultados de la prueba de Kolmogorov-Smirnov ($KS = 0,0453$; $p = 0,545 > 0,05$) y de Anderson-

Darling ($AD=0,943$, $p=0,017$), junto con la prueba de Shapiro-Wilk ($0,990$, $p=0,025$), se observa que la distribución generalmente cumple con el supuesto normal.

En tercer lugar, se examinó la multicolinealidad y, para ello, se verificaron los valores de VIF y Tolerancia (Tabla 2). Se espera que los valores de VIF sean <5 o 10 (Montgomery, 2001; Field, 2009; Alpar, 2012; Büyüköztürk, 2006) y que los valores de Tolerancia (1-SMC) sean $>0,20$ y cercanos a 1 (Büyüköztürk, 2006). Tras el análisis, se observó que los valores de VIF oscilaron entre $2,091$ y $4,682$ y todos fueron menores que 5 , mientras que los valores de Tolerancia (1-SMC) oscilaron entre $0,214$ y $0,478$ y todos fueron mayores que $0,20$. Estos resultados indican que no existe un problema de multicolinealidad entre las variables.

Tabla 2.

Valores VIF y de tolerancia

ítem	Tolerancia	VIF
C1	0,432	2,316
C2	0,332	3,016
C3	0,434	2,302
C4	0,432	2,312
C5	0,364	2,749
C6	0,336	2,976
C7	0,298	3,354
C8	0,359	2,783
C9	0,350	2,858
C10	0,433	2,311
CT1	0,353	2,832
CT2	0,321	3,114
CT3	0,320	3,126
CT4	0,315	3,176
CT5	0,349	2,865
CT6	0,369	2,710
CT7	0,289	3,459
CT8	0,316	3,168
CT9	0,353	2,830
CT10	0,400	2,499

ítem	Tolerancia	VIF
CR1	0,358	2,791
CR2	0,327	3,057
CR3	0,277	3,607
CR4	0,413	2,420
CR5	0,478	2,091
CR6	0,381	2,622
CR7	0,349	2,863
CR8	0,257	3,893
CR9	0,274	3,643
CR10	0,447	2,235
CH1	0,298	3,355
CH2	0,343	2,912
CH3	0,419	2,389
CH4	0,299	3,346
CH5	0,263	3,804
CH6	0,283	3,529
CH7	0,301	3,319
CH8	0,323	3,096
CH9	0,375	2,668
CH10	0,373	2,681
COM1	0,337	2,969
COM2	0,333	2,999
COM3	0,276	3,624
COM4	0,315	3,177
COM5	0,307	3,258
COM6	0,291	3,440
COM7	0,247	4,044
COM8	0,291	3,441
COM9	0,285	3,503
COM10	0,320	3,125
COL1	0,319	3,133

ítem	Tolerancia	VIF
COL2	0,229	4,366
COL3	0,226	4,424
COL4	0,264	3,784
COL5	0,214	4,682
COL6	0,232	4,306
COL7	0,256	3,900
COL8	0,287	3,486
COL9	0,290	3,447
COL10	0,392	2,548

Nota: C (Ciudadanía), CT (Pensamiento Crítico), CR (Creatividad), CH (Carácter), COM (Comunicación), COL (Colaboración).

En cuarto lugar, para detectar valores atípicos multivariados, se compararon los valores de Z y de distancia de Mahalanobis (Tabla 3) con el valor $X^2(60) = 99,607$ al nivel de significancia de 0,001. Los valores de la distancia de Mahalanobis oscilaron entre 0,880 y 106,625, y se excluyeron del análisis 10 observaciones que superaban el valor crítico.

Tabla 3.

Distancias de Mahalanobis

Observación	Distancia de Mahalanobis
O1	106.625
O2	106.054
O3	104.262
O4	103.326
O5	102.654
O6	102.069
O7	101.378
O8	100.406
O9	100.387
O10	99.871
O11	99.589
O12	99.466
O13	99.293

Observación	Distancia de Mahalanobis
O14	99.085
...	...
O318	1.336
O319	1.213
O320	1.213
O321	1.213
O322	0,880

Nota. Resultados del autor.

En quinto lugar, se examinaron los valores de Durbin-Watson (Tabla 4) para determinar si los datos eran independientes del error. Un valor de Durbin entre 0 y 4, cercano a 2, indica que los datos son independientes del error. En este análisis, el valor de Durbin-Watson de los datos fue de 1,89. Como resultado de la prueba de Durbin-Watson, el valor de autocorrelación calculado fue de 0,050 y el valor p de 0,33. Por lo tanto, no existe una autocorrelación significativa en el modelo de regresión para $n=302$ al nivel de significancia de 0,01 ($p>0,01$), lo que significa que el error puede considerarse independiente.

Tabla 4.

Valores de Durbin-Watson

Modelo	R	R ²	Durbin-Watson	Autocorrelación	p
1	0,525	0,275	1.89	0,050	0,33

Nota. Resultados del autor.

Finalmente, tras excluir del análisis las observaciones con valores extremos y aquellas que generaban problemas de multicolinealidad, se evaluó la factorabilidad de los datos, que comprendían las 302 observaciones restantes. Se examinaron el valor KMO (Medida de Adecuación del Muestreo de Kaiser-Meyer-Olkin) y el valor de la Prueba de Esfericidad de Bartlett (Tabla 5) para comprobar la significancia del valor KMO. El valor KMO fue de 0,965, y el resultado de la Prueba de Esfericidad de Bartlett mostró que el valor KMO era significativo ($\chi^2=17506$, $gl=1770$, $p<0,001$). Un valor KMO superior a 0,90 indica que el tamaño de la muestra era excelente (Alpar, 2012).

Tabla 5.

Valores de la prueba KMO y de Bartlett

Medida	Valor
Medida de adecuación del muestreo de Kaiser-Meyer-Olkin	0,965
Prueba de esfericidad de Bartlett – Aprox. χ^2	17.506
Prueba de esfericidad de Bartlett – gl	1.770
Prueba de esfericidad de Bartlett – Sig.	< 0,001

Nota. Resultados del autor.

El valor KMO superior a 0,95 indica que el conjunto de datos es extremadamente adecuado para el análisis factorial.

Análisis factorial exploratorio

Se realizó un análisis factorial exploratorio para examinar la estructura factorial subyacente de la escala preliminar. El análisis se llevó a cabo mediante extracción de máxima verosimilitud con rotación oblimin, de acuerdo con el supuesto teórico de factores correlacionados. Los resultados del AFE se presentan a continuación.

Tabla 6.

Cargas factoriales

Ítem	Factor 1 (COL)	Factor 2 (COM)	Factor 3 (CR)	Factor 4 (CH)	Factor 5 (CT)	Factor 6 (C)	Unicidad
COL7	0,852						0,250
COL8	0,837						0,244
COL6	0,776						0,250
COL5	0,774						0,220
COL1	0,696						0,377
0							
COL3	0,690						0,261
COL4	0,680						0,249
COL9	0,667						0,275

COL2	0,631						0,249
COL1	0,546						0,324
COM7		0,829					0,186
COM6		0,815					0,201
COM5		0,710					0,372
COM8		0,685					0,310
COM4		0,604					0,360
COM1 0		0,519					0,329
COM9		0,490					0,333
COM3		0,476					0,351
COM2		0,393					0,433
CR3			0,711				0,283
CR4			0,654				0,435
CR2			0,625				0,340
CR9			0,598				0,346
CR8			0,577				0,319
CR5			0,577				0,537
CR10			0,554				0,501
CR7			0,439				0,418
CR6			0,399				0,447
CR1			0,369				0,474
CH10			0,352				0,479
CH1			0,326				0,425
CH6				0,804			0,212
CH5				0,774			0,226
CH7				0,709			0,278
CH8				0,651			0,313
CH9				0,611			0,367
CH2				0,523			0,354
CH4				0,506			0,326
CH3				0,448			0,492

COM1							0,453
CT4					0,668		0,264
CT7					0,639		0,283
CT10					0,604		0,418
CT5					0,593		0,379
CT8					0,584		0,321
CT6					0,572		0,355
CT1					0,549		0,374
CT9					0,544		0,384
CT2					0,472		0,340
CT3					0,440		0,367
C1					0,435		0,485
C2					0,337	0,301	0,459
C6						0,704	0,354
C7						0,701	0,283
C5						0,683	0,386
C8						0,656	0,333
C4						0,584	0,500
C3						0,511	0,506
C10						0,382	0,509
C9				0,366	0,378		0,397

Nota. Se utilizó el método de extracción de máxima verosimilitud con rotación oblimin.

Como resultado del análisis inicial, se obtuvo una estructura de 6 factores, pero algunos ítems mostraron cargas factoriales bajas o cargas cruzadas en múltiples factores. Los índices de ajuste del modelo (Tabla 7) se calcularon como RMSEA = 0,0659 (0,0632-0,0691), TLI = 0,850.

Tabla 7.

Medidas de ajuste del modelo

Índice de ajuste	Valor	Umbral aceptable
RMSEA	0,0659	< 0,08
TLI	0,850	> 0,90

Nota. Resultados del autor.

Proceso de retirada de ítems

Durante el proceso de análisis factorial exploratorio, se excluyeron sistemáticamente un total de 23 ítems. En la primera fase de limpieza, se eliminaron tres ítems críticos. El ítem CH10 se excluyó del análisis debido a su baja carga factorial de 0,317 y a que también presentaba problemas de cargas cruzadas. El ítem COM1 se excluyó del análisis porque no presentaba cargas significativas en ningún factor y no contribuía a la estructura factorial, y el ítem C9 se eliminó debido a que presentaba problemas de cargas simultáneas en múltiples factores.

En la segunda fase de limpieza, se excluyeron del análisis tres ítems más con cargas factoriales bajas. El ítem C2 tenía una carga factorial de 0,371, el ítem COM2 de 0,378 y el ítem C10 de 0,369; estos valores se mantuvieron significativamente por debajo del umbral aceptable de 0,50. Estos ítems se excluyeron del análisis porque reducían la calidad general de la estructura factorial y no podían proporcionar una medición fiable.

En la tercera fase de limpieza, se priorizó la selección de ítems con bajos valores de comunalidad. El ítem CR10 fue eliminado debido a que los factores no lo explicaban adecuadamente, con un valor de comunalidad de 0,505. El ítem CR5 también fue eliminado debido a un valor de comunalidad de 0,535.

Elementos eliminados: COL1, COL2, COL3, COM1, COM2, COM3, CT1, CT2, CT3, CH1, CH2, CH3, CH4, CH10, C1, C2, C9, C10, CR1, CR5, CR6, CR7, CR10

Como resultado de este proceso de limpieza gradual, la escala inicial, compuesta por 60 ítems, se redujo a 37 ítems, y los índices de ajuste del modelo mejoraron significativamente.

Segundo análisis factorial (37 ítems)

Como resultado del análisis factorial realizado tras el proceso de eliminación de ítems, se observó que el valor KMO (Tabla 8) fue de 0,960 y el índice de compatibilidad entre ítems también fue significativo ($\chi^2=10171$, gl =666, $p<0,001$). El análisis factorial realizado mediante el método de extracción de residuos mínimos y la técnica de rotación oblimin revela las propiedades psicométricas de la estructura de seis dimensiones que representa las habilidades del siglo XXI.

Tabla 8.
KMO y Bartlett

Medida de adecuación del muestreo de Kaiser-Meyer-Olkin		0,960
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	10171
	Gl	666
	Sig.	< 0,001

Nota. Resultados del autor.

Las cargas factoriales presentadas en la Tabla 9 demuestran una clara estructura de seis factores que se alinea con el marco teórico de las habilidades 6C. Cada factor muestra cargas fuertes para los ítems dentro de sus dominios respectivos, con los ítems de Colaboración cargando en el Factor 1 (que oscila entre 0,616 y 0,872), los ítems de Comunicación en el Factor 2 (0,562 a 0,898), los ítems de Pensamiento Crítico en el Factor 3 (0,595 a 0,708), los ítems de Carácter en el Factor 4 (0,694 a 0,876), los ítems de Ciudadanía en el Factor 5 (0,550 a 0,790) y los ítems de Creatividad en el Factor 6 (0,532 a 0,752). El método de extracción de residuos mínimos combinado con la rotación oblimin reveló con éxito esta estructura factorial interpretable, con la mayoría de los ítems demostrando cargas sustanciales superiores a 0,60 en sus factores primarios y cargas cruzadas mínimas en otros factores.

Tabla 9.
Cargas factoriales

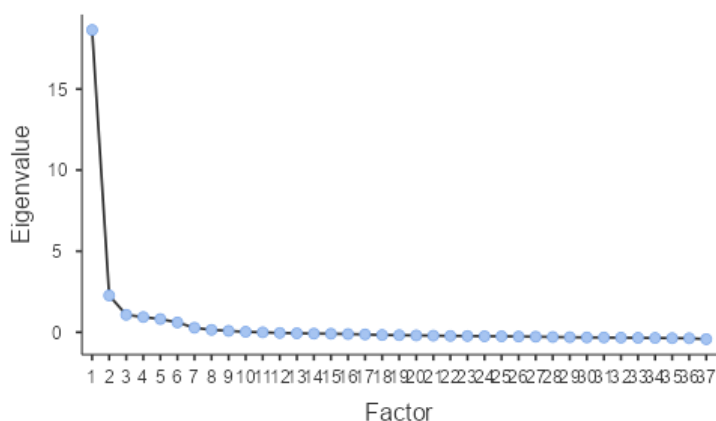
	Factor						Unicidad
	1	2	3	4	5	6	
COL8	0,872						0,212
COL7	0,833						0,252
COL6	0,756						0,242
COL9	0,725						0,238
COL5	0,716						0,233
COL10	0,708						0,354
COL4	0,616						0,279
COM7		0,898					0,175
COM6		0,839					0,216

COM8		0,725					0,305
COM5		0,708					0,381
COM4		0,644					0,371
COM10		0,585					0,317
COM9		0,562					0,320
CT7			0,708				0,258
CT10			0,661				0,388
CT4			0,651				0,296
CT8			0,649				0,287
CT9			0,625				0,358
CT6			0,621				0,341
CT5			0,595				0,385
CH6				0,876			0,174
CH7				0,756			0,259
CH8				0,732			0,287
CH5				0,701			0,283
CH9				0,694			0,321
C5					0,790		0,298
C6					0,657		0,363
C4					0,651		0,448
C7					0,630		0,304
C8					0,595		0,359
C3					0,550		0,491
CR3						0,752	0,244
CR4						0,670	0,421
CR2						0,621	0,345
CR9						0,579	0,348
CR8						0,532	0,344
<i>Nota: Se utilizó el método de extracción de “residuo mínimo” en combinación con una rotación de “oblimin”.</i>							

Nota. Resultados del autor.

El gráfico de sedimentación (Figura 1) proporciona una confirmación visual de la solución de seis factores, mostrando un claro punto de inflexión después del sexto valor propio que respalda la retención de seis factores para la rotación. Este patrón de valores propios refuerza la expectativa teórica de seis dominios de habilidades distintos pero relacionados en el modelo de medición.

Figura 1.
Cargas factoriales



Nota. Resultados del autor.

Varianza explicada

Los resultados de la explicación de la varianza (Tabla 10) obtenidos del análisis factorial revelan que la estructura de seis factores explica una porción significativa de la varianza total. El primer factor se posiciona como el más fuerte, explicando el 14,27% de la varianza total con una carga factorial de 5,28. El segundo factor tiene una tasa de explicación de la varianza del 13,68% con un valor de carga de 5,06, y los dos primeros factores juntos explican el 27,9% de la varianza total. El tercer y cuarto factor muestran un desempeño casi igual, con tasas de explicación de la varianza del 11,26% y 11,24% con cargas factoriales de 4,16, respectivamente. Estos cuatro factores juntos explican el 50,4% de la varianza total, formando la estructura principal de la escala. El quinto factor contribuye con el 9,66% de la varianza con un valor de carga de 3,58, mientras que el sexto factor muestra la menor contribución al explicar el 8,81% de la varianza con un valor de carga de 3,26. La tasa de explicación de la varianza total de los seis factores se calculó en un 68,9%. Esta tasa es bastante satisfactoria para los estudios de análisis factorial en el campo de las ciencias sociales y demuestra la sólida validez estructural de la escala que representa las habilidades del siglo XXI. La distribución

relativamente equilibrada de la varianza entre los factores indica que ninguno predomina excesivamente y que cada dimensión ocupa un lugar significativo en la escala.

Tabla 10.

Cargas de varianza

Factor	Cargas SS	% de variación	Porcentaje acumulado
1	5,28	14,27	14,3
2	5,06	13,68	27,9
3	4,16	11,26	39,2
4	4,16	11,24	50,4
5	3,58	9,66	60,1
6	3,26	8,81	68,9

Nota. Resultados del autor.

Distribución y valores de carga de la estructura factorial

Factor de Colaboración (COL): Este factor contiene el conjunto de ítems más completo y está representado por siete ítems. En esta dimensión, liderada por el ítem COL8 con una carga factorial de 0,872, las cargas factoriales oscilan entre 0,616 y 0,872. Si bien la carga factorial relativamente baja del ítem COL4 (0,616) llama la atención, se encuentra en un nivel aceptable. Los valores de unicidad que permanecen en el rango de 0,212 a 0,354 indican que el factor común explica en gran medida los ítems.

Factor de Comunicación (COM): En este factor operacionalizado con seis ítems, el ítem COM7 presenta el valor más alto de toda la escala, con una carga factorial de 0,898. La distribución de las cargas factoriales entre 0,562 y 0,898 refleja la consistencia de la medición de las habilidades comunicativas. El ítem COM9, con la carga más baja (0,562), indica que su representación factorial es limitada.

Factor de Pensamiento Crítico (PC): En este factor, representado por seis ítems, el ítem PC7 destaca con una carga factorial de 0,708. La distribución de las cargas factoriales en el rango de 0,595 a 0,708 indica una medición homogénea de las habilidades de pensamiento crítico. El ítem PC5, con la carga factorial más baja (0,595), demuestra ser el representante más débil en esta dimensión.

Factor de Carácter (CC): En este factor de cinco ítems, destaca la posición dominante de CC6 con una carga factorial de 0,876. La variación de las cargas factoriales entre 0,694 y

0,876 demuestra que este factor es una de las estructuras más consistentes. Todos los ítems que presentan cargas superiores a 0,694 revelan las sólidas características de medición de la dimensión de carácter.

Factor de Ciudadanía (C): En este factor operacionalizado con seis ítems, destaca la carga factorial de C5 de 0,790. Considerando la distribución de las cargas factoriales en el rango de 0,550 a 0,790, se observa que el ítem C3 se encuentra en un valor límite de 0,550, pero se mantiene en un nivel aceptable.

Factor de Creatividad (CR): En este factor, representado por cinco ítems, se observa que CR3 presenta una fuerte representación con una carga factorial de 0,752. El ítem CR8, al tener el valor más bajo en toda la escala con una carga factorial de 0,532, indica que el poder explicativo factorial de este ítem es limitado.

Indicadores de fiabilidad psicométrica

El uso de la rotación oblimin permite correlaciones entre factores, lo que refleja la naturaleza relacional de las habilidades del siglo XXI en la vida real. En general, los valores de unicidad que se mantienen mayoritariamente por debajo de 0,50 demuestran que el poder explicativo factorial de los ítems es adecuado.

Estos hallazgos revelan que la estructura teórica de seis factores de la escala de habilidades del siglo XXI ha sido validada empíricamente y puede utilizarse como una herramienta de medición válida en las evaluaciones educativas.

Tabla 11.

Medidas de ajuste del modelo

RMSEA	Intervalo de confianza del 90% para RMSEA		TLI	BIC	Prueba del modelo		
	Más bajo	Superior			χ^2	gl	p
0,0645	0,0594	0,0699	0,911	-1585	1037	459	<.001

Nota. Resultados del autor.

Los índices de ajuste del modelo (Tabla 11) generalmente proporcionan resultados aceptables. El valor RMSEA es 0,0645, manteniéndose por debajo del umbral de 0,08 e

indicando un nivel de ajuste aceptable. El valor TLI calculado fue 0,911 y se mantuvo por encima del umbral de 0,90, lo que revela que el ajuste del modelo es adecuado. La prueba de chi-cuadrado fue estadísticamente significativa, pero esto es previsible en muestras de gran tamaño, y otros índices de ajuste proporcionan información más fiable.

Tabla 12.

Correlaciones entre factores

	COL	COM	CT	CH	C	CR
COL	—	0,694	0,387	0,551	0,457	0,543
COM		—	0,473	0,614	0,527	0,615
CT			—	0,615	0,499	0,351
CH				—	0,516	0,417
C					—	0,413
CR						—

Nota. Resultados del autor. COL=Colaboración, COM=Comunicación, CT=Pensamiento crítico, CH=Carácter, C=Ciudadanía, CR=Creatividad.

Las correlaciones entre factores (Tabla 12) oscilan entre 0,351 y 0,694. La correlación más alta se observa entre los factores Colaboración y Comunicación ($r=0,694$), mientras que la más baja se registra entre los factores Pensamiento Crítico y Creatividad ($r=0,351$). Estas correlaciones demuestran que existen relaciones significativas entre los factores, pero se mantiene la validez discriminante.

Tabla 13.

Valores KMO generales e individuales

Factor	Ítems	KMO (Mín-Máx)
KMO general	Todos los ítems	0,960
COL (Colaboración)	Ítems COL	0,951 - 0,973
COM (Comunicación)	ítems COM	0,959 - 0,976
TC (Pensamiento Crítico)	Ítems de CT	0,955 - 0,970
CH (Carácter)	Ítems CH	0,947 - 0,971
C (Ciudadanía)	Ítems C	0,939 - 0,971
CR (Creatividad)	Ítems CR	0,939 - 0,962

Nota. Resultados del autor.

Las pruebas de supuestos respaldan la idoneidad del análisis factorial. El valor KMO (Tabla 13) se calculó en 0,960, lo que indica que el tamaño de la muestra es excelente. Los valores MSA individuales oscilan entre 0,934 y 0,977, lo que revela que se alcanza la adecuación de la muestra para todos los ítems. La prueba de esfericidad de Bartlett fue estadísticamente significativa y confirmó que existe una estructura de correlación adecuada para el análisis factorial entre las variables.

Análisis factorial confirmatorio

Para el análisis factorial confirmatorio, se utilizaron datos de 319 participantes. El tamaño de la muestra (319) es adecuado para este análisis. Se empleó el método de estimación de máxima verosimilitud y se calcularon errores estándar robustos. El modelo convergió en 116 iteraciones, lo que indica que el proceso de estimación fue exitoso.

Tabla 14.

Índices de ajuste del modelo

Índice de ajuste	Valor	Criterios de aceptación	Resultado
χ^2 (escalado)	926.00	-	-
gl	611	-	-
valor p	< 0,001	> 0,05	Significativo
RMSEA (escalado)	0,040	≤ 0.08 aceptable, ≤ 0.06 bueno	Buen ajuste

Límite inferior de RMSEA	0,036	-	-
Límite superior de RMSEA	0,044	-	-
Valor p de RMSEA	1.000	> 0,05	Buen ajuste
CFI (escalado)	0,959	≥ 0.95 excelente, ≥ 0.90 aceptable	Ajuste excelente
TLI (a escala)	0,955	≥ 0.95 excelente, ≥ 0.90 aceptable	Ajuste excelente
SRMR	0,050	≤ 0.08 buen ajuste	Buen ajuste
GFI	0,886	≥ 0.90 aceptable	Límite

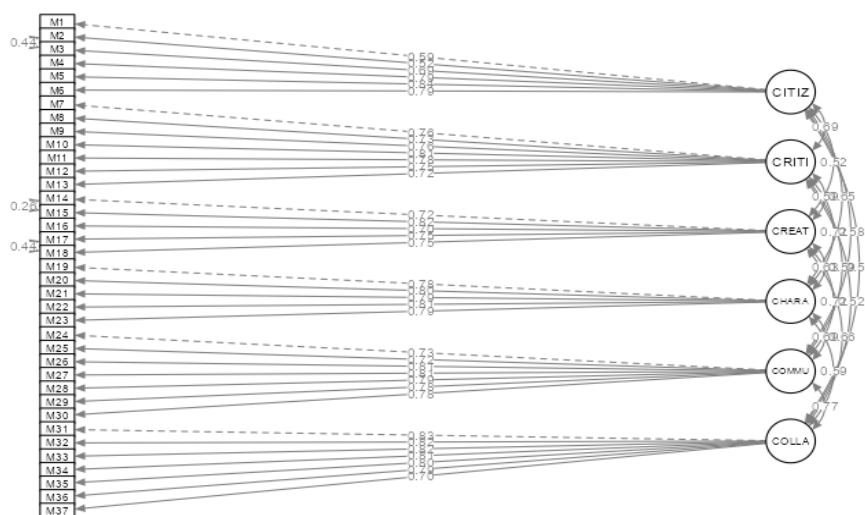
Nota. Resultados del autor.

Los índices de ajuste del modelo (Tabla 14) evalúan qué tan bien se ajusta el modelo propuesto a los datos observados. Los valores de RMSEA (0,040), CFI (0,959), TLI (0,955) y SRMR (0,050) revelan que el modelo presenta un ajuste excelente. El resultado significativo de la prueba de chi-cuadrado ($p < 0,001$) es una situación normal en muestras de gran tamaño y no requiere, por sí solo, el rechazo del modelo.

Diagrama de trayectorias

Figura 2.

Diagrama de trayectorias del análisis factorial confirmatorio de la escala de habilidades del siglo XXI



Nota. Resultados del autor.

El diagrama de trayectorias (Figura 2) presenta visualmente la estructura de 6 factores de la escala de habilidades del siglo XXI de 37 ítems. El diagrama muestra los factores CIUDADANÍA (IT1-IT6), PENSAMIENTO CRÍTICO (IT7-IT13), CREATIVIDAD (IT14-IT18), CARÁCTER (IT19-IT23), COMUNICACIÓN (IT24-IT30) y COLABORACIÓN (IT31-IT37), así como las cargas factoriales de los ítems pertenecientes a estos factores. Las cargas factoriales estandarizadas oscilan entre 0,525 y 0,844, con 35 de los 37 ítems presentando cargas factoriales elevadas superiores a 0,70. Las correlaciones interfactoriales oscilan entre 0,521 y 0,769, observándose la correlación más alta entre los factores COMUNICACIÓN y COLABORACIÓN ($r=0,769$), y la correlación más baja entre los factores PENSAMIENTO CRÍTICO y COLABORACIÓN ($r=0,521$).

Estos resultados revelan que la estructura teórica del modelo de seis factores está fuertemente respaldada por los datos empíricos. La mayoría de las cargas factoriales superiores a 0,70 demuestran que cada factor explica adecuadamente sus propios ítems. Las correlaciones de nivel moderado-alto entre los factores (inferiores a 0,85) confirman que las habilidades del

siglo XXI son estructuras relacionadas pero distinguibles. Tres modificaciones de covarianza de error realizadas en el modelo (IT2↔IT3, IT14↔IT15, IT17↔IT18) mejoraron el ajuste del modelo, y se obtuvieron excelentes índices de ajuste (RMSEA=0,040, CFI=0,959, TLI=0,955). Estas modificaciones se justificaron teóricamente, ya que implicaban ítems consecutivos con características de contenido similares. Estos hallazgos demuestran que la escala puede medir las habilidades del siglo XXI de manera válida y fiable. Todas las cargas factoriales son estadísticamente significativas al nivel $p < 0,001$. Se muestran los coeficientes estandarizados.

Cargas factoriales y niveles de significancia

La Tabla 15 muestra las cargas factoriales y los parámetros estadísticos asociados para todos los ítems dentro de sus respectivos factores. Las cargas factoriales estandarizadas (β) representan la fuerza de asociación entre cada ítem y su constructo subyacente, junto con los errores estándar, los estadísticos z y los niveles de significancia que confirman la validez de la estructura factorial.

CIUDADANÍA

Tabla 15.

Cargas factoriales

Nombre del factor	Ítem	Carga factorial (β)	Error estándar	valor z	valor p
CIUDADANÍA	IT1	0,593	-	-	-
	IT2	0,525	0,097	9.92	< 0,001
	IT3	0,695	0,107	11.43	< 0,001
	IT4	0,789	0,120	11.04	< 0,001
	IT5	0,844	0,129	10,78	< 0,001
	IT6	0,789	0,129	10.59	< 0,001
PENSAMIENTO CRÍTICO	IT7	0,757	-	-	-
	IT8	0,729	0,058	15.01	< 0,001
	IT9	0,762	0,062	15.44	< 0,001
	IT10	0,814	0,061	17.07	< 0,001
	IT11	0,779	0,066	15.30	< 0,001
	IT12	0,745	0,062	15,64	< 0,001
	IT13	0,724	0,061	14.70	< 0,001
CREATIVIDAD	IT14	0,721	-	-	-
	IT15	0,822	0,067	17,75	< 0,001
	IT16	0,700	0,082	12.34	< 0,001

	IT17	0,748	0,075	14.25	< 0,001
	IT18	0,751	0,077	14.20	< 0,001
CARÁCTER	IT19	0,781	-	-	-
	IT20	0,803	0,058	17.01	< 0,001
	IT21	0,789	0,060	15.80	< 0,001
	IT22	0,806	0,060	16.19	< 0,001
	IT23	0,790	0,057	17.38	< 0,001
COMUNICACIÓN	IT24	0,735	-	-	-
	IT25	0,723	0,073	14.33	< 0,001
	IT26	0,810	0,074	15.20	< 0,001
	IT27	0,811	0,071	15.51	< 0,001
	IT28	0,787	0,074	14.77	< 0,001
	IT29	0,779	0,077	14.25	< 0,001
COLABORACIÓN	IT30	0,776	0,072	15.15	< 0,001
	IT31	0,826	-	-	-
	IT32	0,823	0,052	19.72	< 0,001
	IT33	0,818	0,047	21.57	< 0,001
	IT34	0,813	0,049	20,75	< 0,001
	IT35	0,801	0,052	18,94	< 0,001
	IT36	0,787	0,053	18.51	< 0,001
	IT37	0,700	0,050	17.55	< 0,001

Nota. Resultados del autor.

Todos los ítems del factor CIUDADANÍA son estadísticamente significativos ($p < 0,001$). Las cargas factoriales oscilan entre 0,525 y 0,844. El IT1 se fija en 1,00, ya que es el ítem de referencia. El IT5 presenta la carga factorial más alta (0,844), mientras que el IT2 presenta la más baja (0,525). Todas las cargas factoriales son superiores a 0,50, lo que indica que son aceptables.

El factor PENSAMIENTO CRÍTICO presenta cargas factoriales elevadas (entre 0,724 y 0,814). Todos los ítems son altamente significativos ($z > 14$). El ítem IT10 tiene la carga factorial más alta (0,814), mientras que el IT13 tiene la más baja (0,724). Este factor muestra una estructura homogénea.

Las cargas factoriales del factor CREATIVIDAD oscilan entre 0,700 y 0,822. IT15 presenta la carga factorial más alta (0,822) y el valor z más alto (17,75). Se realizó una modificación entre IT17 e IT18 (covarianza = 1,327). Todas las cargas factoriales superan el nivel aceptable.

El factor CARÁCTER presenta cargas factoriales muy consistentes (entre 0,781 y 0,806). Las cargas factoriales son muy similares entre sí, lo que indica una estructura factorial homogénea. IT22 presenta la carga factorial más alta (0,806). Todos los valores *z* son superiores a 15.

COMUNICACIÓN es el factor más grande, con 7 ítems. Las cargas factoriales oscilan entre 0,723 y 0,811. Los ítems IT26 e IT27 presentan las cargas más altas (0,810 y 0,811, respectivamente). Este factor muestra una fuerte consistencia interna. Todos los ítems son altamente significativos.

El factor COLABORACIÓN presenta las cargas factoriales más sustanciales (entre 0,700 y 0,826). IT31, IT32 y IT33 tienen cargas factoriales particularmente altas (>0,80). Este factor presenta los valores *z* más elevados (entre 17,55 y 21,57), lo que refleja una estructura factorial muy sólida. Si bien M37 tiene la carga factorial más baja, aún se encuentra en un nivel aceptable de 0,70.

Análisis de confiabilidad

Tabla 16.

Análisis de confiabilidad

Factor	alfa de Cronbach	Fiabilidad compuesta (ω)	AVE
CIUDADANÍA	0,866	0,833	0,503
PENSAMIENTO CRÍTICO	0,905	0,905	0,577
CREATIVIDAD	0,876	0,835	0,563
CARÁCTER	0,895	0,895	0,630
COMUNICACIÓN	0,913	0,913	0,600
COLABORACIÓN	0,923	0,923	0,633

Nota. Resultados del autor.

Los resultados del análisis de fiabilidad (Tabla 16) muestran que todos los factores presentan una alta consistencia interna. Los valores del coeficiente alfa de Cronbach oscilan entre 0,866 y 0,923, todos por encima del criterio de 0,70. Los valores de fiabilidad compuesta (ω) son igualmente altos. Los valores de AVE (Varianza Media Extraída) superan el criterio de 0,50, lo que respalda la validez convergente. COLABORACIÓN presenta la mayor fiabilidad, mientras que CIUDADANÍA tiene un valor de AVE en el umbral (0,503).

Correlaciones entre factores

Tabla 17.
Correlaciones entre factores

	Ciudadanía	Pensamiento crítico	Creatividad	Carácter	Comunicación	Colaboración
Ciudadanía	1.000	0,692**	0,522**	0,649**	0,577**	0,540**
Pensamiento crítico	0,692**	1.000	0,590**	0,718**	0,590**	0,521**
Creatividad	0,522**	0,590**	1.000	0,630**	0,716**	0,656**
Carácter	0,649**	0,718**	0,630**	1.000	0,694**	0,593**
Comunicación	0,577**	0,590**	0,716**	0,694**	1.000	0,769**
Colaboración	0,540**	0,521**	0,656**	0,593**	0,769**	1.000

Nota. Resultados del autor. ** $p < 0,001$

La matriz de correlación interfactorial (Tabla 17) muestra las relaciones entre los seis factores. Las correlaciones oscilan entre 0,521 y 0,769. La correlación más alta se da entre COMUNICACIÓN y COLABORACIÓN ($r = 0,769$). La correlación más baja se da entre PENSAMIENTO CRÍTICO y COLABORACIÓN ($r = 0,521$). Todas las correlaciones son estadísticamente significativas ($p < 0,001$). Las correlaciones inferiores a 0,85 indican que los factores son distinguibles.

HTMT (heterotrait-monotrait)

Tabla 18.
HTMT (heterotrait-monotrait)

	Ciudadanía	Pensamiento crítico	Creatividad	Carácter	Comunicación	Colaboración
Ciudadanía	-	0,680	0,558	0,651	0,607	0,580
Pensamiento crítico	0,680	-	0,575	0,720	0,582	0,522
Creatividad	0,558	0,575	-	0,615	0,705	0,639
Carácter	0,651	0,720	0,615	-	0,687	0,589

Comunicación	0,607	0,582	0,705	0,687	-	0,773
Colaboración	0,580	0,522	0,639	0,589	0,773	-

Nota. Los valores de HTMT deben ser inferiores a 0,90. Todos los valores cumplen este criterio.

La matriz de ratios HTMT (Heterotrait-Monotrait) (Tabla 18) se utiliza para evaluar la validez discriminante de los factores. Todos los valores HTMT están por debajo del criterio de 0,90, lo que indica que se alcanza la validez discriminante entre los factores. El valor HTMT más alto se encuentra entre COMUNICACIÓN y COLABORACIÓN (0,773). El valor HTMT más bajo se encuentra entre PENSAMIENTO CRÍTICO y COLABORACIÓN (0,522). Estos resultados respaldan que los factores son constructos distinguibles. Estos resultados indican que el modelo de 6 factores es una estructura estadísticamente aceptable y confiable. Este modelo puede utilizarse como un instrumento válido y confiable para medir las habilidades del siglo XXI.

Resultados del análisis factorial confirmatorio de segundo orden (AFC de segundo orden)

El análisis probó si los seis factores de primer orden (F1-F6) se pueden agregar bajo un único factor de orden superior (SIXCS).

Índices de ajuste del modelo

El modelo AFC de segundo orden demuestra un ajuste excelente (Tabla 19). Los valores de RMSEA (0,042), CFI (0,954), TLI (0,951) y SRMR (0,059) revelan que el modelo se ajusta muy bien a los datos. Estos resultados respaldan la afirmación de que los primeros seis factores de orden pueden agruparse de manera significativa en un único factor de orden superior. Los índices de ajuste son muy similares a los del AFC de primer orden, lo que indica que la estructura jerárquica es apropiada.

Tabla 19.

Índices de ajuste del modelo

Índice de ajuste	Valor	Criterios de aceptación	Resultado
------------------	-------	-------------------------	-----------

χ^2 (escalado)	969.00	-	-
gl	620	-	-
valor p	< 0,001	> 0,05	Significativo
RMSEA (escalado)	0,042	≤ 0.08 aceptable, ≤ 0.06 bueno	Buen ajuste
Límite inferior de RMSEA	0,038	-	-
Límite superior de RMSEA	0,046	-	-
Valor p de RMSEA	0,999	> 0,05	Buen ajuste
CFI (escalado)	0,954	≥ 0.95 excelente, ≥ 0.90 aceptable	Ajuste excelente
TLI (a escala)	0,951	≥ 0.95 excelente, ≥ 0.90 aceptable	Ajuste excelente
SRMR	0,059	≤ 0.08 buen ajuste	Buen ajuste
GFI	0,880	≥ 0.90 aceptable	En el umbral

Nota. Resultados del autor.

*Cargas factoriales de segundo orden***Tabla 20.***Cargas de segundo orden (SIXCS → Factores de primer orden)*

Factor	Carga factorial (β)	Error estándar	Límite inferior	Límite superior	valor z	valor p
SIXCS → F1 (CIUDADANÍA)	0,734	-	0,661	0,807	-	-
SIXCS → F2 (PENSAMIENTO CRÍTICO)	0,765	0,152	0,698	0,832	8.79	< 0,001
SIXCS → F3 (CREATIVIDAD)	0,791	0,158	0,712	0,869	8.53	< 0,001
SIXCS → F4 (CARÁCTER)	0,828	0,186	0,770	0,885	8.44	< 0,001
SIXCS → F5 (COMUNICACIÓN)	0,861	0,175	0,794	0,928	8.42	< 0,001
SIXCS → F6 (COLABORACIÓN)	0,785	0,180	0,712	0,859	8.57	< 0,001

Nota. Resultados del autor.

La Tabla 20 muestra las cargas factoriales del factor SIXCS de orden superior sobre los seis factores de primer orden. Todas las cargas factoriales son superiores a 0,70 y estadísticamente muy significativas ($p < 0,001$). COMUNICACIÓN tiene la carga factorial más alta (0,861), mientras que CIUDADANÍA tiene la más baja (0,734). Estas altas cargas factoriales respaldan firmemente la idea de que los seis dominios de habilidades son, de hecho, componentes de un único constructo de orden superior denominado «habilidades del siglo XXI». F1 sirve como factor de referencia; por lo tanto, no se dispone de un valor z.

Análisis de confiabilidad

Tabla 21.
Análisis de confiabilidad

Factor	alfa de Cronbach	Fiabilidad compuesta (ω)	AVE	Resultado
F1 (CIUDADANÍA)	0,866	0,836	0,506	Buena fiabilidad, AVE en el umbral
F2 (PENSAMIENTO CRÍTICO)	0,905	0,905	0,578	Muy buena fiabilidad
F3 (CREATIVIDAD)	0,876	0,835	0,563	Muy buena fiabilidad
F4 (CARÁCTER)	0,895	0,895	0,629	Muy buena fiabilidad
F5 (COMUNICACIÓN)	0,913	0,913	0,599	Excelente fiabilidad
F6 (COLABORACIÓN)	0,923	0,923	0,633	Excelente fiabilidad

Nota. Resultados del autor.

Los resultados del análisis de fiabilidad (Tabla 21) son idénticos a los del análisis factorial confirmatorio de primer orden y confirman que todos los factores presentan una alta consistencia interna. Todos los valores de alfa de Cronbach son superiores a 0,86, lo que indica excelentes niveles de fiabilidad. Los valores de AVE superan el criterio de 0,50, lo que respalda la validez convergente.

Comparación de modelos: de primer orden frente a de segundo orden

Tabla 22.
De primer orden frente a de segundo orden

Índice de ajuste	AFC de primer orden	AFC de segundo orden	Diferencia	Evaluación
χ^2 (escalado)	926	969	+43	Pequeño incremento
gl	611	620	+9	Más restricciones

RMSEA (escalado)	0,040	0,042	+0,002	Casi idéntico
CFI (escalado)	0,959	0,954	-0,005	Disminución mínima
TLI (a escala)	0,955	0,951	-0,004	Disminución mínima
SRMR	0,050	0,059	+0,009	Pequeño incremento

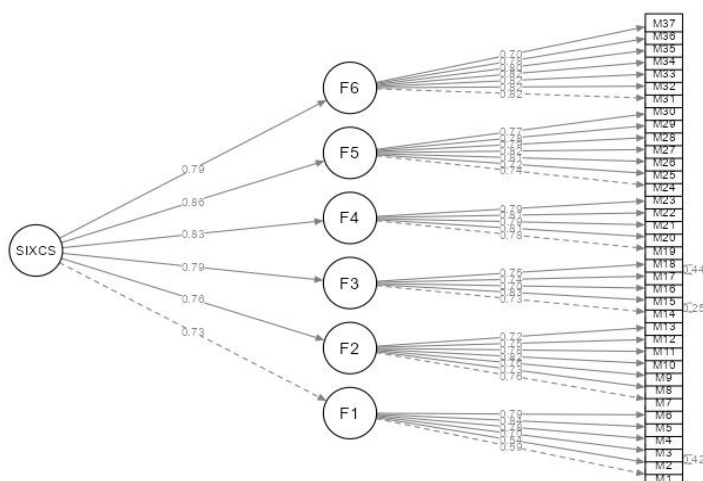
Nota. Resultados del autor.

La comparación de modelos revela que el AFC de segundo orden presenta un ajuste muy similar al del AFC de primer orden (Tabla 22). Las mínimas diferencias en los índices de ajuste ($\Delta\text{RMSEA} = 0,002$, $\Delta\text{CFI} = 0,005$) respaldan la idoneidad de la estructura jerárquica para los datos. Estos resultados indican que, según el principio de parsimonia, se puede preferir el modelo de segundo orden, más sencillo. Se obtuvo un modelo más simple y teóricamente más significativo con una mínima pérdida de ajuste.

Diagrama de trayectorias

Figura 3.

Diagrama de trayectorias AFC de segundo orden



Nota. Resultados del autor.

El diagrama de trayectorias del AFC de segundo orden (Figura 3) visualiza claramente la estructura jerárquica de los factores. El factor de orden superior SIXCS se ubica en el lado izquierdo del diagrama, mientras que los seis primeros factores de orden (F1-F6) se sitúan en el lado derecho. Las flechas que parten del factor SIXCS hacia los factores F1-F6 contienen las cargas factoriales (que oscilan entre 0,734 y 0,861), las cuales indican el grado en que el factor de orden superior explica los subfactores. Las flechas que se extienden desde cada factor de primer orden hacia sus respectivos ítems muestran las relaciones ítem-factor.

Esta estructura jerárquica respalda la idea de que las habilidades del siglo XXI son subdimensiones que se unen bajo un único constructo general, manteniendo sus características únicas. Los factores COMUNICACIÓN (0,861) y CARÁCTER (0,828) representan la estructura general con mayor fuerza, mientras que el factor CIUDADANÍA (0,734) aporta la contribución más baja, pero aún significativa. El modelo presenta una estructura teóricamente más clara, libre de correlaciones directas entre factores de primer orden, y demuestra que una única estructura de orden superior puede explicar las habilidades del siglo XXI. Todas las

cargas factoriales son estadísticamente significativas al nivel $p < 0,001$. Índices de ajuste del modelo: $\chi^2=969$, $gl=620$, $RMSEA=0,042$, $CFI=0,954$, $TLI=0,951$, $SRMR=0,059$.

Método de puntuación

El modelo AFC de segundo orden ofrece ventajas significativas sobre el modelo de primer orden en cuanto a la puntuación de la escala de habilidades del siglo XXI. Esta estructura jerárquica permite la puntuación independiente de las seis subdimensiones (CIUDADANÍA, PENSAMIENTO CRÍTICO, CREATIVIDAD, CARÁCTER, COMUNICACIÓN, COLABORACIÓN) y el cálculo de una puntuación SIXCS general que abarca todos los ítems. En las evaluaciones individuales, las puntuaciones de las subdimensiones (calculadas promediando los ítems de cada factor) se utilizan para un análisis detallado del perfil, mientras que la puntuación SIXCS general (promedio de los 37 ítems) resume el nivel general de habilidades del siglo XXI de cada individuo en un solo valor. Este enfoque de puntuación de doble nivel proporciona flexibilidad tanto para identificar áreas de desarrollo como para realizar clasificaciones intragrupalas, mientras que la puntuación ponderada basada en las cargas factoriales (utilizando cargas factoriales que oscilan entre 0,734 y 0,861) ofrece una evaluación general más precisa. Los índices de ajuste del modelo, al ser casi idénticos a los del AFC de primer orden ($\Delta CFI = 0,005$), respaldan la validez psicométrica de este método de puntuación y avalan la preferencia por el modelo de segundo orden en aplicaciones prácticas.

Escala de puntuación

La puntuación de la escala de habilidades del siglo XXI (Tabla 23) puede implementarse mediante dos enfoques diferentes basados en los resultados del análisis factorial confirmatorio de segundo orden. En el enfoque de puntuación basado en factores, cada una de las seis subdimensiones (ciudadanía, pensamiento crítico, creatividad, carácter, comunicación y colaboración) se calcula por separado tomando la media aritmética de sus respectivos ítems. Este enfoque proporciona un análisis de perfil detallado para determinar en qué áreas de habilidades los individuos son fuertes o débiles, y proporciona un rango de puntuación de 1,00 a 9,00 para cada factor. El número de ítems varía según los factores: el factor de ciudadanía contiene seis ítems, los factores de pensamiento crítico y comunicación contienen siete ítems

cada uno, los factores de creatividad y carácter contienen cinco ítems cada uno, y el factor de colaboración contiene siete ítems.

Tabla 23.

Rangos de puntuación de la Escala de Habilidades del Siglo XXI

Factor	Número de ítems	Números de los ítems	Rango de puntuación total	Rango de puntuación promedio	Carga factorial
Ciudadanía (F1)	6	IT1-IT6	6-54	1.00-9.00	0,734
Pensamiento crítico (F2)	7	IT7-IT13	7-63	1.00-9.00	0,765
Creatividad (F3)	5	IT14-IT18	5-45	1.00-9.00	0,791
Carácter (F4)	5	IT19-IT23	5-45	1.00-9.00	0,828
Comunicación (F5)	7	IT24-IT30	7-63	1.00-9.00	0,861
Colaboración (F6)	7	IT31-IT37	7-63	1.00-9.00	0,785
general de SIXCS	37	IT1-IT37	37-333	1.00-9.00	-

Nota. Resultados del autor.

El método de puntuación general de segundo orden consiste en calcular la puntuación del factor SIXCS como la media de las puntuaciones de los 37 ítems, resumiendo así el nivel general de habilidades del siglo XXI del individuo en un único valor. Esta puntuación general puede calcularse mediante la media aritmética simple (suma de todos los ítems dividida entre 37) o mediante una media ponderada basada en las cargas factoriales ($F1 \times 0,734 + F2 \times 0,765 + F3 \times 0,791 + F4 \times 0,828 + F5 \times 0,861 + F6 \times 0,785$). El método de puntuación ponderada proporciona resultados psicométricamente más sensibles, ya que considera las contribuciones de los factores a la estructura de orden superior y aumenta la validez teórica de la puntuación general. Ambos sistemas de puntuación pueden interpretarse dentro de los rangos de 1,00-3,00 (bajo), 3,01-5,00 (moderado), 5,01-7,00 (bueno) y 7,01-9,00 (alto) (Tabla 24). La escala final se presenta en el Apéndice.

Tabla 24.*Criterios de interpretación de la puntuación*

Rango de puntuación	Nivel	Descripción
1.00-3.00	Bajo	Alta necesidad de desarrollo en el área de habilidades
3.01-5.00	Moderado	Nivel de competencia moderado en el área de habilidades.
5.01-7.00	Bien	Buen nivel de competencia en el área de habilidades.
7.01-9.00	Alto	Alto nivel de competencia en el área de especialización.

Nota. Resultados del autor.***Análisis de confiabilidad***

Los resultados del análisis de fiabilidad (Tabla 25) indican que la escala de habilidades del siglo XXI demuestra una alta consistencia interna tanto a nivel general como factorial. El coeficiente alfa de Cronbach de la escala global fue de 0,962, lo que indica una fiabilidad excelente. El valor ω de McDonald también fue de 0,962, lo que confirma el coeficiente alfa. Al examinar los datos a nivel factorial, los coeficientes de fiabilidad de las seis subdimensiones oscilan entre 0,866 y 0,923, observándose la menor fiabilidad en el factor de Ciudadanía ($\alpha=0,866$) y la mayor en el factor de Colaboración ($\alpha=0,923$). Si bien los factores Pensamiento Crítico ($\alpha=0,905$), Comunicación ($\alpha=0,913$) y Colaboración ($\alpha=0,923$) tienen una fiabilidad excelente, los factores Ciudadanía ($\alpha=0,866$), Creatividad ($\alpha=0,876$) y Carácter ($\alpha=0,895$) tienen una buena fiabilidad.

Los resultados del análisis a nivel de ítems muestran que todos los ítems contribuyen a la escala y no es necesario eliminar ninguno. Los valores de correlación ítem-resto oscilan entre 0,501 y 0,780, con todos los ítems por encima del umbral crítico de 0,30. Aunque el ítem IT2 (0,501) tiene la correlación ítem-resto más baja en la escala general (Tabla 25), este valor aún se encuentra en un nivel aceptable. Los ítems con las correlaciones ítem-resto más altas fueron identificados como IT29 (0,706), IT27 (0,691) e IT31 (0,696). El cambio en el alfa de Cronbach al eliminar ítems sigue siendo mínimo (oscilando entre 0,960 y 0,962), y no se recomienda eliminar ningún ítem. Estos resultados respaldan que todos los ítems de la escala funcionan

adecuadamente según la estructura teórica y que la escala es un instrumento fiable en términos de propiedades psicométricas.

Tabla 25.

Resultados del análisis de fiabilidad de la Escala de Habilidades del Siglo XXI

Factor	Número de ítems	Mediana	Desviación estándar	alfa de Cronbach	McDonald's ω	Rango de correlación ítem-resto	
						Más bajo	Máximo
Escala general (SIXCS)	37	5.90	1,66	0,962	0,962	0,501	0,706
Ciudadanía (F1)	6	5.64	1,97	0,866	0,868	0,587 (IT2)	0,730 (IT3)
Pensamiento crítico (F2)	7	6.33	1.91	0,905	0,905	0,685 (IT13)	0,769 (IT10)
Creatividad (F3)	5	5.63	2.15	0,876	0,877	0,646 (IT16)	0,767 (IT15)
Carácter (F4)	5	6.20	2.12	0,895	0,895	0,722 (IT19)	0,752 (IT20)
Comunicación (F5)	7	5.87	2.08	0,913	0,913	0,693 (IT25)	0,769 (IT27)
Colaboración (F6)	7	5.71	2.19	0,923	0,924	0,674 (IT37)	0,780 (IT31)

Nota. Resultados del autor. Se aceptó un umbral de correlación ítem-resto de 0,30. Todos los ítems superan este umbral. Los criterios de evaluación de la fiabilidad se interpretan de la siguiente manera: los valores alfa de 0,90 o superiores indican una fiabilidad excelente; los valores entre 0,80 y 0,89 indican una buena fiabilidad; los valores entre 0,70 y 0,79 indican una fiabilidad aceptable; y los valores inferiores a 0,70 indican una baja fiabilidad.

Evaluación de los docentes en la implementación de las habilidades del siglo XXI según el género del estudiante

Los análisis realizados por género del estudiante (Tabla 26) revelan que los estudiantes de ambos sexos evalúan las competencias de sus docentes en la implementación de habilidades del siglo XXI en niveles similares. Los resultados de la prueba t de muestras independientes muestran que no existen diferencias estadísticamente significativas entre los grupos en cinco de las seis dimensiones de habilidades y en la puntuación total general ($p > .05$). El único hallazgo notable es una diferencia marginalmente significativa observada en la dimensión de pensamiento crítico ($p = .064$, d de Cohen = 0.210); en esta dimensión, las estudiantes evalúan a sus docentes más alto que los estudiantes (45.54 vs 42.74, diferencia media = 2.80). La mayor diferencia en otras dimensiones se observa en el área de creatividad, donde los estudiantes evalúan a sus docentes ligeramente más alto que las estudiantes (28.50 vs 27.92). Sin embargo, esta diferencia no es estadísticamente significativa ($p = .631$).

Tabla 26.

Evaluaciones de la competencia de los docentes en la implementación de habilidades del siglo XXI según el género del estudiante

Área de habilidades	Género del estudiante	n	Media	DE (Desviación Estándar)	gl	valor t	valor p	d de Cohen	Tamaño del efecto	Resultado
6C HABILIDADES (TODAS)	Femenino	179	220.22	64,24	317	0,611	0,541	0,069	Muy pequeño	No significativo
	Masculino	140	216.00	57,59						
Ciudadanía	Femenino	179	33,93	12,20		0,205	0,837	0,023		

	Masculino	140	33,73	11.35	317				Muy pequeño	No significativo
Pensamiento crítico	Femenino	179	45.54	13.60	317	1.861	0,064	0,210	Pequeño	Marginalmente significativo
	Masculino	140	42,74	13.04	317					
Creatividad	Femenino	179	27,92	11.30	317	-0,481	0,631	-0,054	Muy pequeño	No significativo
	Masculino	140	28.50	9,93	317					
Carácter	Femenino	179	31.43	11.00	317	0,743	0,458	0,084	Muy pequeño	No significativo
	Masculino	140	30.53	10.00	317					
Comunicación	Femenino	179	41,54	15.20	317	0,570	0,569	0,064	Muy pequeño	No significativo
	Masculino	140	40,61	13,76	317					
Colaboración	Femenino	179	39,93	16.20	317	-0,048	0,961	-0,005	Muy pequeño	No significativo
	Masculino	140	40.01	14.24	317					

Nota. Resultados del autor. Criterios de interpretación del tamaño del efecto (d de Cohen): $d < 0,20$ = Efecto muy pequeño; $0,20 \leq d < 0,50$ = Efecto pequeño; $0,50 \leq d < 0,80$ = Efecto medio; $d \geq 0,80$ = Efecto grande.

Estos hallazgos ofrecen importantes implicaciones educativas respecto a la interacción docente-alumno. El hecho de que los estudiantes, tanto hombres como mujeres, perciban de manera similar las competencias pedagógicas de sus docentes en general indica que estos exhiben enfoques neutrales en cuanto al género en la gestión del aula y las prácticas de enseñanza. La diferencia marginalmente significativa en la dimensión del pensamiento crítico podría indicar que las estudiantes tienden a evaluar las prácticas de sus docentes de manera más positiva en esta área de habilidades. Sin embargo, los tamaños del efecto muy pequeños en

todas las dimensiones (d de Cohen $< 0,20$) muestran que estas diferencias son prácticamente insignificantes. El desequilibrio en los tamaños de muestra ($N_1=179$, $N_2=140$) y la violación del supuesto de homogeneidad de varianza en algunas dimensiones son factores que deben considerarse al interpretar los resultados. En la evaluación general, estos resultados respaldan que las competencias de los docentes en la implementación de las habilidades 6C se perciben independientemente del género del estudiante y proporcionan un indicador positivo en términos de igualdad de género en los entornos educativos.

Evaluación de los docentes en la implementación de habilidades del siglo XXI por nivel de grado

Los resultados del análisis ANOVA unidireccional (Tabla 27) por nivel de grado muestran que no existen diferencias estadísticamente significativas en las evaluaciones que los estudiantes de 5.º a 11.º grado hacen de las competencias de implementación de habilidades del siglo XXI de sus docentes. Según los resultados de la prueba ANOVA de Welch, los valores p superaron 0,05 en las seis dimensiones de habilidades (ciudadanía, pensamiento crítico, creatividad, carácter, comunicación, colaboración) y en la puntuación total general (rango $p = 0,254 - 0,717$). Estos hallazgos revelan que los estudiantes de diferentes niveles de grado perciben las competencias pedagógicas de sus docentes en niveles similares. Las pruebas post hoc de Tukey tampoco detectaron diferencias significativas en ninguna de las comparaciones por pares, lo que respalda que las diferencias entre niveles de grado son aleatorias.

Tabla 27.

Evaluaciones de competencias de los docentes en la implementación de habilidades del siglo XXI por nivel de grado

Área de habilidades	<i>F</i>	<i>gl</i>	<i>p</i>
Total general (6 habilidades C)	1,020	6,109	0,416
Ciudadanía	1,320	6,111	0,254
Pensamiento crítico	1,172	6,109	0,326
Creatividad	1,131	6,109	0,349
Carácter	1,044	6,110	0,401
Comunicación	0,616	6,109	0,717
Colaboración	0,973	6,109	0,447

Nota. Resultados del autor. Se utilizó el análisis de varianza de Welch debido a la desigualdad en el tamaño de los grupos. Todos los valores $p > 0,05$, lo que indica que no existen diferencias significativas entre los niveles de grado.

Al examinar las estadísticas descriptivas (Tabla 28), se observa que los estudiantes de sexto grado generalmente evaluaron a sus maestros con las puntuaciones más altas (total general: 234.3, pensamiento crítico: 47.6, carácter: 33.7, comunicación: 44.5), mientras que los estudiantes de undécimo grado realizaron las evaluaciones más bajas en la mayoría de las dimensiones (total general: 205.9, pensamiento crítico: 41.1, carácter: 29.1, colaboración: 36.1). Esta tendencia puede interpretarse como un aumento en las expectativas de los maestros a medida que aumenta la edad de los estudiantes y los criterios de evaluación se vuelven más estrictos. Cabe destacar que los estudiantes de quinto grado dieron las puntuaciones más altas en las dimensiones de creatividad y ciudadanía. Desde una perspectiva educativa, estos resultados muestran que los maestros exhiben un desempeño pedagógico consistente en diferentes niveles de grado e implementan prácticas efectivas en la enseñanza de habilidades del siglo XXI, independientemente del grupo de edad.

Tabla 28.

Estadísticas descriptivas de las habilidades del siglo XXI por nivel de grado

Nivel de grado	n	Total general	Ciudadanía	Pensamiento crítico	Creatividad	Carácter	Comunicación	Colaboración
5°	18	226,7 (60,70)	36,6 (9,67)	43,7 (13,31)	31,3 (9,64)	31,9 (9,55)	39,8 (13,85)	43,5 (14,05)
6°	57	234,3 (70,29)	36,0 (12,56)	47,6 (12,76)	29,7 (12,41)	33,7 (11,05)	44,5 (15,26)	42,8 (16,35)
7°	72	215,9 (52,42)	34,8 (11,47)	44,0 (12,71)	26,8 (9,65)	29,9 (10,07)	40,8 (12,62)	39,6 (13,29)
8°	38	223,3 (56,28)	34,2 (9,17)	45,4 (12,67)	30,0 (10,54)	32,2 (10,32)	39,8 (14,99)	41,7 (15,39)
9°	54	211,1 (60,02)	31,2 (12,08)	43,4 (13,68)	26,8 (9,78)	30,0 (11,18)	40,3 (15,27)	39,4 (13,99)

10°	3 8	214,2 (67,00)	33,6 (13,82)	43,9 (16,27)	26,7 (10,68)	31,1 (9,43)	41,1 (14,98)	37,9 (15,81)
11°	4 2	205,9 (62,34)	31,3 (11,69)	41,1 (12,66)	28,4 (11,77)	29,1 (11,52)	39,9 (15,60)	36,1 (18,42)

Nota. Resultados del autor. Los valores representan M (DE). N total = 319. No se encontraron diferencias significativas entre los niveles de grado en ninguna medida utilizando comparaciones post hoc HSD de Tukey.

Los hallazgos respaldan investigaciones previas que sugieren que las habilidades 6C requieren la participación activa del docente y un diseño instruccional intencional. Los estudiantes reportaron mayor competencia en áreas donde los docentes utilizaron tareas colaborativas y centradas en la comunicación, en consonancia con el modelo de aprendizaje profundo de Fullan. La escala validada proporciona nuevas evidencias de que los estudiantes pueden ofrecer juicios significativos sobre la implementación de las habilidades 6C por parte de los docentes, complementando los autoinformes docentes encontrados en estudios anteriores.

DISCUSIÓN

El objetivo principal de este estudio fue desarrollar y validar una escala integral para evaluar la competencia docente en la implementación y promoción de las habilidades 6C desde la perspectiva del estudiante. El proceso sistemático de desarrollo en cinco fases dio como resultado un instrumento psicométricamente sólido de 37 ítems que demuestra excelentes características de fiabilidad y validez, lo que permite subsanar una importante deficiencia en la evaluación educativa.

Propiedades psicométricas y validación teórica

El proceso de desarrollo de la escala aportó pruebas contundentes tanto de la coherencia teórica como de la validez empírica del marco 6C. La exitosa identificación de seis factores distintos, correspondientes a las dimensiones teóricas de Ciudadanía, Pensamiento Crítico, Creatividad, Carácter, Comunicación y Colaboración, proporciona un sólido respaldo empírico para conceptualizarlas como competencias separadas pero relacionadas. Este hallazgo es

particularmente significativo, ya que demuestra que estos constructos teóricos pueden observarse y distinguirse en la práctica real del aula.

El excelente coeficiente de fiabilidad global ($\alpha = 0,962$) y la robusta fiabilidad a nivel factorial (que oscila entre 0,866 y 0,923) superan con creces los estándares convencionales para instrumentos de medición educativa. Estos resultados indican que la escala proporciona una medición consistente y fiable en los seis dominios de competencia, con el factor Colaboración demostrando la mayor consistencia interna ($\alpha = 0,923$) y el factor Ciudadanía mostrando la menor fiabilidad, aunque aún aceptable ($\alpha = 0,866$).

Los resultados del análisis factorial confirmatorio proporcionan evidencia contundente de la validez estructural del instrumento. Los excelentes índices de ajuste (RMSEA = 0,040, CFI = 0,959, TLI = 0,955, SRMR = 0,050) demuestran que el modelo de seis factores representa adecuadamente la estructura subyacente de las competencias docentes en la implementación de las habilidades 6C. Además, la validación exitosa del modelo de segundo orden con un deterioro mínimo en los índices de ajuste (Δ RMSEA = 0,002, Δ CFI = 0,005) respalda la proposición de que las seis áreas de competencia pueden conceptualizarse como manifestaciones de un constructo de orden superior que representa la competencia pedagógica general del siglo XXI.

Estructura factorial e interrelaciones

Los patrones de carga factorial revelan información importante sobre cómo los estudiantes perciben los diferentes aspectos de las competencias docentes del siglo XXI. El factor Comunicación mostró la relación más fuerte con el factor de orden superior (carga = 0,861), lo que sugiere que los estudiantes consideran las habilidades comunicativas fundamentales para una pedagogía eficaz en el siglo XXI. Le sigue de cerca el factor Carácter (carga = 0,828), lo que indica que los estudiantes valoran mucho a los docentes que ejemplifican comportamientos éticos y morales.

Las correlaciones interfactoriales, que oscilan entre 0,351 y 0,694, revelan patrones interesantes de relaciones entre las seis habilidades C. La correlación particularmente fuerte entre Colaboración y Comunicación ($r = 0,694$) coincide con las expectativas, ya que estas habilidades están intrínsecamente interconectadas en la práctica docente. Por el contrario, la correlación relativamente más débil entre Pensamiento Crítico y Creatividad ($r = 0,351$) sugiere que representan enfoques pedagógicos más distintos, lo que respalda su conceptualización por separado en el marco teórico.

Percepciones de los estudiantes e invarianza demográfica

La ausencia de diferencias significativas entre géneros y niveles educativos ofrece información valiosa sobre la universalidad de las percepciones estudiantiles respecto a la pedagogía eficaz del siglo XXI. Estos hallazgos sugieren que las prácticas docentes eficaces en la implementación de las habilidades 6C trascienden las barreras demográficas y se reconocen de manera consistente en poblaciones estudiantiles diversas. Esta invariabilidad entre los distintos grupos demográficos refuerza la utilidad del instrumento en diversos contextos educativos y con distintos grupos de estudiantes.

La significación marginal observada en la dimensión de Pensamiento Crítico entre estudiantes mujeres y hombres ($p = 0,064$, d de Cohen = 0,210) exige una interpretación cuidadosa. Si bien este hallazgo se acerca a la significación estadística, el pequeño tamaño del efecto sugiere una importancia práctica limitada. Este patrón podría reflejar diferencias sutiles en cómo los estudiantes de distinto género perciben o valoran la enseñanza del pensamiento crítico, aunque tales interpretaciones deben hacerse con cautela dado el nivel de significación límite.

La estabilidad de las percepciones en todos los niveles educativos (de 5.º a 11.º grado) es particularmente destacable, ya que sugiere que las competencias 6C son valoradas de manera consistente por los estudiantes, independientemente de su etapa de desarrollo o nivel académico. Este hallazgo respalda la aplicabilidad universal del marco en todos los niveles de la educación secundaria.

Rendimiento de los elementos y refinamiento de la escala

El riguroso proceso de eliminación de ítems, que redujo los 160 ítems iniciales a los 37 finales, demuestra la importancia del refinamiento sistemático de la escala. La eliminación de 23 ítems durante la fase de análisis factorial exploratorio se basó en criterios psicométricos claros, como bajas cargas factoriales, problemas de cargas cruzadas y valores de comunalidad insuficientes. Este proceso dio como resultado un instrumento más parsimonioso y psicométricamente sólido.

La distribución final de los ítems entre los factores (Ciudadanía: 6 ítems; Pensamiento crítico: 7 ítems; Creatividad: 5 ítems; Carácter: 5 ítems; Comunicación: 7 ítems; Colaboración: 7 ítems) refleja tanto consideraciones teóricas como el desempeño empírico. La distribución

relativamente equitativa sugiere que cada dominio de competencia está adecuadamente representado, manteniendo al mismo tiempo la eficiencia de la medición.

Implicaciones prácticas para la evaluación educativa

El desarrollo de esta escala ofrece importantes beneficios prácticos para la práctica educativa y el desarrollo profesional docente. El instrumento proporciona un marco estructurado para recopilar retroalimentación sistemática sobre la implementación de las competencias del siglo XXI por parte del profesorado, lo que puede contribuir tanto al crecimiento profesional individual como a las iniciativas de mejora institucional. La perspectiva del alumnado aporta un valor único, ya que los estudiantes son los destinatarios finales de estos enfoques pedagógicos.

El sistema de puntuación de doble nivel, posible gracias al modelo de segundo orden, resulta especialmente útil para los profesionales de la educación. La disponibilidad de puntuaciones específicas para cada factor y una puntuación global de competencias 6C permite una evaluación diagnóstica más precisa, a la vez que proporciona una métrica de evaluación integral. Esta flexibilidad favorece enfoques diferenciados de desarrollo profesional, facilitando intervenciones específicas en áreas de competencia concretas sin perder de vista la eficacia pedagógica general.

Fortalezas e innovaciones metodológicas

Este estudio se distingue de otros intentos previos de desarrollar escalas por diversas razones metodológicas. El uso de muestras independientes para los análisis factoriales exploratorios y confirmatorios proporciona evidencia sólida de la estabilidad y generalización de la estructura factorial. El proceso sistemático de desarrollo en cinco fases, que comienza con una extensa revisión bibliográfica y la recopilación de datos cualitativos, garantiza una sólida validez de contenido y una fundamentación teórica firme.

El enfoque en las percepciones de los estudiantes como evaluadores de la competencia docente representa una estrategia innovadora en el campo. En lugar de basarse únicamente en las autoevaluaciones del profesorado o en observaciones externas, este estudio capta la experiencia auténtica en el aula desde la perspectiva del alumno, ofreciendo una visión de cómo los estudiantes experimentan realmente los enfoques pedagógicos.

Limitaciones y consideraciones metodológicas

Al interpretar estos hallazgos, conviene tener en cuenta varias limitaciones. La muestra se obtuvo exclusivamente de escuelas secundarias turcas en una región geográfica específica, lo que podría limitar la generalización de los resultados a otros contextos culturales y educativos. La validez del instrumento en diferentes sistemas educativos y marcos culturales requiere una investigación más profunda.

La dependencia de las percepciones de los estudiantes como única fuente de datos representa tanto una fortaleza como una limitación. Si bien las perspectivas estudiantiles brindan información valiosa sobre las prácticas en el aula, es posible que no abarquen todos los aspectos de la competencia docente en la implementación de las habilidades del siglo XXI. Los estudiantes no siempre reconocen ni evalúan con precisión ciertas estrategias pedagógicas, en particular aquellas cuyos efectos son más sutiles o a largo plazo.

Este estudio desarrolló y validó una escala confiable para medir la percepción de los estudiantes sobre la competencia docente en la implementación de las habilidades 6C. La escala contribuye a la literatura al ofrecer una herramienta de evaluación centrada en el estudiante que se alinea con los principales marcos del siglo XXI. Sus sólidas propiedades psicométricas sugieren que puede respaldar la mejora educativa, el desarrollo docente y futuras investigaciones sobre la instrucción basada en competencias.

El diseño transversal del estudio limita las conclusiones sobre la estabilidad de la estructura factorial a lo largo del tiempo y la sensibilidad del instrumento a los cambios en la práctica docente. Además, el estudio no establece relaciones entre las competencias docentes medidas con esta escala y los resultados reales del aprendizaje de los estudiantes en habilidades del siglo XXI.

CONSIDERACIONES FINALES

Este estudio desarrolló y validó con éxito un instrumento integral para evaluar la competencia docente en la implementación y promoción de las habilidades 6C desde la perspectiva del estudiante. La escala resultante, compuesta por 37 ítems, demuestra excelentes propiedades psicométricas y proporciona un sólido respaldo empírico al marco teórico de las competencias de aprendizaje del siglo XXI. La robusta estructura factorial del instrumento, su

alta fiabilidad y su invarianza demográfica lo convierten en una valiosa herramienta tanto para la investigación como para la práctica en la educación contemporánea.

Contribuciones teóricas

Los hallazgos del estudio contribuyen significativamente a la comprensión teórica de las habilidades del siglo XXI al proporcionar la primera validación empírica del marco de las 6C mediante observaciones de la práctica docente realizadas por estudiantes. La exitosa extracción de seis factores distintos que corresponden a las dimensiones teóricas, junto con la validación de un modelo jerárquico de segundo orden, ofrece evidencia convincente tanto de la singularidad como de la unidad de estas competencias en la práctica educativa.

La estructura jerárquica revelada por el análisis factorial confirmatorio de segundo orden sugiere que, si bien las seis competencias representan dominios pedagógicos distintos, también reflejan un enfoque unificado de la educación del siglo XXI. Este hallazgo tiene importantes implicaciones para comprender cómo funcionan estas habilidades en la práctica y cómo deben desarrollarse y evaluarse.

Aplicaciones prácticas e impacto educativo

Desde un punto de vista práctico, la escala ofrece múltiples aplicaciones para la mejora educativa. Puede servir como herramienta de diagnóstico para identificar necesidades de desarrollo profesional, como instrumento de evaluación formativa para monitorear el crecimiento docente y como métrica de evaluación de la eficacia institucional en la implementación de pedagogías del siglo XXI. Sus excelentes características de confiabilidad garantizan una medición consistente y fidedigna en diversos contextos.

El diseño de este instrumento como herramienta de evaluación estudiantil es particularmente valioso, ya que captura la experiencia auténtica en el aula desde la perspectiva del alumno. Este enfoque se alinea con las filosofías educativas contemporáneas que enfatizan la participación y la autonomía del estudiante en la evaluación educativa, a la vez que ofrece perspectivas prácticas sobre la eficacia de los enfoques pedagógicos.

El sistema de puntuación ofrece flexibilidad para diferentes propósitos de evaluación. Los líderes educativos pueden usar las puntuaciones específicas de cada factor para identificar áreas que requieren desarrollo profesional específico, mientras que la puntuación global de las 6C proporciona una medida integral de la eficacia pedagógica. Este enfoque de doble nivel respalda tanto la evaluación diagnóstica como la sumativa.

Implicaciones para el desarrollo docente

La escala proporciona un marco concreto para conceptualizar y evaluar las competencias docentes del siglo XXI. Al hacer observables y medibles estos conceptos abstractos a través de las percepciones de los estudiantes, el instrumento puede orientar la reflexión docente y el desarrollo profesional. Los indicadores de comportamiento específicos integrados en los ítems de la escala ofrecen objetivos claros para la mejora pedagógica.

El hecho de que las percepciones de los estudiantes se mantengan consistentes en todos los grupos demográficos sugiere que la implementación efectiva de las habilidades 6C es universalmente reconocible, independientemente de las características de los estudiantes. Esto respalda el desarrollo de programas estandarizados de desarrollo profesional, al tiempo que mantiene la confianza en que dichos programas serán relevantes para poblaciones estudiantiles diversas.

Direcciones futuras de investigación

Este estudio abre diversas vías para futuras investigaciones. Los estudios de validación transcultural son esenciales para establecer la aplicabilidad del instrumento en diferentes sistemas educativos y contextos culturales. No se puede dar por sentada la aplicabilidad universal del marco sin una verificación empírica en diversos entornos.

Las investigaciones longitudinales que examinen la estabilidad temporal de las competencias docentes y la sensibilidad del instrumento a las intervenciones de desarrollo profesional mejorarían su utilidad para la evaluación de programas. Dichos estudios podrían determinar si la escala puede medir eficazmente el crecimiento a lo largo del tiempo y responder a las iniciativas de mejora específicas.

Investigar la relación entre las evaluaciones estudiantiles realizadas con esta escala y las medidas objetivas de los resultados del aprendizaje estudiantil proporcionaría evidencia valiosa sobre la validez predictiva del instrumento. Establecer vínculos empíricos entre la implementación de las habilidades 6C por parte del profesorado y el rendimiento estudiantil en los dominios de competencias tradicionales y del siglo XXI fortalecería la base teórica del marco.

REFERENCIAS

- Fullan, M., & Scott, G. (2014a). *Change leader: Learning to do what matters most*. Jossey-Bass.
- Fullan, M., & Scott, G. (2014b). *New pedagogies for deep learning (NPDL) white paper*. Collaborative Impact SPC.
- Fullan, M., Quinn, J., & McEachen, J. (2018). *Deep learning: Engage the world, change the world*. Corwin.
- Fullan, M., Quinn, J., & McEachen, J. (2022). *Education reimaged: The future of learning*. Springer.
- Gelen, İ. H. (2017). Investigating the effects of 21st-century skill training on student success. *Journal of Educational Sciences*, 5(2), 98–111.
- Göksun, T., & Kurt, A. (2017). Türkiye’de 21. yüzyıl becerileri ve eğitimde uygulamaları. *Uluslararası Eğitim Bilimleri Dergisi*, 5(3), 45–58.
- Harari, Y. N. (2018). *21 lessons for the 21st century*. Spiegel & Grau.
- Saavedra, A. R., & Opfer, V. D. (2012). Learning 21st-century skills requires 21st-century teaching. *Phi Delta Kappan*, 94(2), 8–13.

CRediT Author Statement

- ☐ **Agradecimientos:** Este estudio se realizó como parte de un Proyecto de Investigación Científica (PIC).
 - ☐ **Financiación:** Este estudio fue realizado por la Universidad Kültür de Estambul.
 - ☐ **Conflictos de intereses:** Los autores declaran que no existen conflictos de intereses.
 - ☐ **Aprobación ética:** Se obtuvo la aprobación ética del Comité de Revisión Ética de la Universidad Kültür de Estambul.
 - ☐ **Disponibilidad de datos y materiales:** Los datos y materiales utilizados en este estudio están disponibles previa solicitud al autor correspondiente.
 - ☐ **Contribuciones de los autores:** Todos los autores contribuyeron sustancialmente a este trabajo. Participaron en la conceptualización, el diseño de la investigación, la recopilación de datos, el análisis y la interpretación. Todos los autores participaron en la redacción y revisión del manuscrito y aprobaron la versión final para su publicación.
-

Procesamiento y edición: Editora Ibero-Americana de Educação
Corrección de pruebas, formato, estandarización y traducción

