

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

## 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

## 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



Birsen TÜTÜNİŞ<sup>1</sup>

e-mail: tutunisster@gmail.com



Kerim ÜNAL<sup>2</sup>

e-mail: kerimunal@mersin.edu.tr



Taner HOSSEINI<sup>3</sup>

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



Özge KÖKSAL<sup>4</sup>

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



Ata ALKAYALAR<sup>5</sup>

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

### How to reference this paper:

Tütüniş, B., Ünal, K., Hosseini, T., Köksal, Ö., & Alkayalar, A. (2026). Students' perceptions of ELT teachers' competence in promoting 6C skills: a scale development and validation study. *Revista on line de Política e Gestão Educacional*, 30(esp1), e026016. <https://doi.org/10.22633/rpge.v30iesp1.20904>



| Submitted: 10/01/2026

| Revisions required: 16/01/2026

| Approved: 05/02/2026

| Published: 30/03/2026

**Editor:** Prof. Dr. Sebastião de Souza Lemes

**Deputy Executive Editor:** Prof. Dr. José Anderson Santos Cruz

<sup>1</sup> Istanbul Kültür University, Ataköy Campus, Istanbul and 34158, Türkiye.

<sup>2</sup> Mersin University, Çiftlikköy Campus, Mersin and 33110, Türkiye.

<sup>3</sup> Istanbul Kültür University, Ataköy Campus, Istanbul and 34158, Türkiye.

<sup>4</sup> Istanbul Kültür University, Ataköy Campus, Istanbul and 34158, Türkiye.

<sup>5</sup> Istanbul Kültür University, Ataköy Campus, Istanbul and 34158, Türkiye.

**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.

**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.

**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.

## INTRODUCTION

The rapid changes in education and the demands of 21st-century learning have increased the need for teachers who can effectively integrate 6C skills—critical thinking, communication, collaboration, creativity, citizenship, and character—into classroom practices. Although many studies examine 21st-century skills, fewer have investigated students’ perceptions of how well teachers implement these competencies. This gap highlights the need for a reliable and valid scale that measures teachers’ competence from the students’ viewpoint. Therefore, this study aims to develop a scale that captures students’ perceptions of teachers’ effectiveness in promoting and applying 6C skills in classroom contexts.

The 21st-century has brought about technological and social changes that fundamentally challenge traditional educational paradigms. Unlike previous educational reforms focused primarily on curriculum content or instructional methods, contemporary educational transformation demands a complete reconceptualization of learning objectives and pedagogical approaches (Saavedra & Opfer, 2012). The shift from industrial-age education—characterized by standardization, compartmentalization, and knowledge transmission—to digital-age learning requires developing learners who can navigate complexity, collaborate across diverse contexts, communicate effectively through multiple modalities, think critically about information abundance, create innovative solutions, and demonstrate character in ethical decision-making.

This transformation reflects broader societal recognition that academic achievement alone is insufficient for success in increasingly complex global environments. Modern challenges—from climate change and social inequality to rapid technological advancement and information proliferation—require citizens capable of collaborative problem-solving, ethical reasoning, and adaptive thinking (Harari, 2018). Educational institutions worldwide are consequently reimagining their fundamental purpose, shifting from preparing students for predetermined career paths to developing adaptable, thoughtful, and innovative individuals capable of creating value in unpredictable futures.

Among various frameworks addressing 21st-century educational needs, Michael Fullan’s 6C model represents a distinctively comprehensive and pedagogically-grounded approach to educational transformation. Unlike frameworks that compartmentalize skills into discrete categories, the 6C model—encompassing Character, Citizenship, Collaboration, Communication, Creativity, and Critical Thinking—conceptualizes these competencies as

interconnected dimensions of deep learning that integrate cognitive, interpersonal, and intrapersonal development (Fullan & Scott, 2014a).

The framework's theoretical foundation rests on Fullan's extensive research into educational change and school improvement, synthesizing decades of implementation science with contemporary understanding of learning processes. This empirical grounding distinguishes the 6C model from more theoretical frameworks, as it emerged from direct observation of successful educational transformation initiatives rather than abstract competency mapping exercises (Fullan et al., 2022). The model's emphasis on "deep learning"—learning that engages students intellectually, emotionally, and socially while developing both academic knowledge and life skills—reflects a sophisticated understanding of how meaningful learning occurs in complex, technology-rich environments.

Furthermore, the 6C framework's global adoption across diverse educational contexts suggests its cross-cultural validity and practical applicability. Educational systems in countries ranging from Canada and Australia to Finland and Singapore have successfully implemented 6C-based initiatives, demonstrating the framework's adaptability to different cultural and institutional contexts while maintaining its core pedagogical principles

Despite growing recognition of the 6C framework's importance and increasing adoption in educational practice, significant gaps exist in our ability to assess teacher competency in implementing these competencies effectively. Current assessment approaches rely primarily on teacher self-evaluation or external observation, missing the critical perspective of students who directly experience these pedagogical approaches in authentic classroom contexts. This limitation is particularly problematic given that 21st-century skills are inherently experiential and relational—their effectiveness depends not on teacher intentions or external perceptions, but on whether students actually develop these competencies through classroom interactions (Fullan & Scott, 2014b).

The absence of validated instruments for capturing student perceptions of teacher 6C implementation creates multiple problems for educational practice and research. Professionally, teachers lack systematic feedback on their effectiveness in fostering 21st-century skills, limiting their capacity for targeted improvement. Institutionally, schools cannot reliably assess the quality of 6C implementation across classrooms or evaluate the impact of professional development initiatives. Empirically, researchers cannot investigate relationships between

teaching practices and student development of 21st-century competencies without valid measurement tools.

This study addresses these critical gaps by developing and validating a comprehensive scale for measuring secondary school students' perceptions of their teachers' competence in implementing and promoting 6C skills. By providing the first psychometrically robust instrument specifically designed for this purpose, the research contributes both theoretical understanding and practical tools for improving 21st-century educational practice.

Previous studies have explored various frameworks for 21st-century skills, including the P21 Framework, ATC21S, and Fullan's Deep Learning model. However, most existing instruments focus either on student competencies or teacher self-evaluations. There is limited research on students' evaluations of teachers' 6C performance. Earlier scales also vary widely in structure, making them difficult to use across different cultural contexts. This study addresses these gaps by designing a student-based scale that aligns with international 6C models while reflecting classroom realities.

Building on the conceptual grounding and the empirical gap outlined above, the present study proceeds from the assumption that students' direct classroom experiences provide an essential lens through which teacher implementation of 21st-century competencies can be meaningfully evaluated. In line with this view, the central purpose of the research is to construct and validate a measurement instrument that captures secondary students' perceptions of their teachers' effectiveness in enacting the six domains of the 6C framework. The development of such a tool is expected to offer both theoretical traction—by clarifying how the 6Cs are

experienced in practice—and practical value for classroom-level feedback, programme improvement, and future research.

### *Research Questions*

Accordingly, the study addresses the following research questions:

- RQ1: To what extent can a student-based instrument be developed that reliably measures teachers' implementation of the six 6C competencies from the learners' viewpoint?
- RQ2: Does exploratory factor analysis support a six-factor structure consistent with the theoretical dimensions of the 6C framework?
- RQ3: To what degree do the resulting factors demonstrate acceptable internal consistency and convergent validity in line with established psychometric criteria?
- RQ4: Does confirmatory factor analysis further validate the six-factor model and provide evidence of a second-order construct representing overall 21st-century teaching competence?

These questions are examined through a multi-stage scale-development process comprising item generation from focus group data, expert review for content validation, exploratory factor analysis to identify latent structure, and confirmatory factor analysis to establish structural coherence. This sequential design allows the theoretical components of the 6C model to be tested against empirical indicators derived from student evaluations, thereby

ensuring a direct correspondence between the conceptual framework, the methodological approach, and the intended contribution of the study.

## METHOD

The development of the Student Evaluation Scale for Teachers' 21st-century Skills (6C) followed a systematic five-phase process designed to ensure content validity, theoretical alignment, and practical applicability in educational settings.

### *Scale Development Process (Research Design)*

#### Phase 1: Literature Review and Theoretical Foundation

An extensive review of existing research on 21st-century skills frameworks was conducted, with particular focus on the 6C model encompassing Critical Thinking, Creativity, Collaboration, Communication, Character, and Citizenship. This review established theoretical foundations and identified key behavioral indicators for each skill domain that could be observed in classroom settings. The literature analysis revealed gaps in student-perspective evaluation instruments, supporting the need for a scale that captures teachers' implementation practices from the learner's viewpoint.

#### Phase 2: Qualitative Data Collection Through Focus Groups

A focus group study was conducted with 15 teachers randomly selected from secondary education institutions. Participants completed a structured interview protocol consisting of 30 open-ended questions, with five questions specifically designed to elicit detailed descriptions of how each 6C skill manifests in classroom practice. The interview questions captured both explicit teaching behaviors and implicit instructional strategies that promote 21st-century skills development.

#### Phase 3: Initial Item Generation and Categorization

Thematic analysis of the focus group responses yielded 160 preliminary scale items, systematically categorized into the six skill domains. The initial distribution reflected the complexity and scope of each competency area. Communication and Critical Thinking each

generated 30 items, representing the most extensive categories. Creativity and Character contributed 29 items each, while Collaboration yielded 24 items. Citizenship, though important, generated the fewest items at 18, reflecting its more focused conceptual scope within the classroom context.

The item generation process prioritized behavioral specificity and classroom observability. Each item was crafted to describe concrete teacher actions that students could readily observe and evaluate, avoiding abstract concepts or internal mental states that would be difficult for students to assess accurately.

#### Phase 4: Content Analysis and Item Refinement

A rigorous content analysis phase involved systematic review by the research team to eliminate redundant, ambiguous, or theoretically misaligned items. This process applied several exclusion criteria: (a) semantic overlap with existing items, (b) insufficient alignment with the target skill domain, (c) cross-loading potential across multiple domains, and (d) inappropriate complexity for the intended student audience.

The refinement process reduced the item pool from 160 to 95 items, with elimination rates varying across domains. Communication experienced the most substantial reduction, with 15 items removed and 15 remaining, representing the highest reduction rate due to redundancy in items addressing similar communication behaviors.

Critical Thinking had 14 items eliminated, retaining 16 items, while Creativity saw 12 items removed, leaving 17 remaining. Character had 13 items eliminated and 16 retained. Collaboration experienced a moderate reduction with 8 items removed and 16 remaining. Citizenship showed the lowest elimination rate with only 3 items removed and 15 remaining, suggesting that the initial items for this domain were well-differentiated and theoretically distinct. Overall, 65 items were eliminated from the original pool, resulting in 95 items for further development.

#### Phase 5: Expert Validation and Final Item Selection

The refined 95-item scale underwent comprehensive expert validation involving six specialists: two Measurement and Evaluation experts, two Curriculum and Instruction experts,

and three English Language Teaching experts. This multidisciplinary approach ensured evaluation from both psychometric and content perspectives.

Experts evaluated each item using a 10-point Likert scale ranging from 1 (completely inappropriate) to 10 (completely appropriate). The evaluation criteria included content relevance, age-appropriateness for secondary students, clarity of language, and alignment with theoretical frameworks. Items receiving mean ratings below 7.0 were systematically removed, ensuring that only items with strong expert consensus remained.

### *Participants*

The study involved two separate samples for exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The EFA sample consisted of 302 participants, while the CFA sample included 319 participants. The demographic characteristics of both samples are presented in Table 1.

## **RESULT**

### *Preliminary Analyses*

Before conducting the factor analyses, a series of preliminary checks was carried out to assess whether the dataset met the assumptions required for multivariate analysis. Descriptive statistics and diagnostic tests showed no evidence of violations related to normality, multicollinearity, or sampling adequacy. In addition, the Kaiser–Meyer–Olkin index and Bartlett’s test of sphericity indicated that the data were suitable for factor analytic procedures. The outcomes of these preliminary analyses are reported in Tables 1–4. Detailed descriptive statistics and assumption checks, including normality indices and intercorrelations, are reported in Tables 2–4. Participant characteristics for the EFA and CFA samples are presented in Table 1.

**Table 1.**  
*Demographic Characteristics of EFA and CFA Participants*

| Variable | Category | EFA (n) | EFA (%) | CFA (n) | CFA (%) |
|----------|----------|---------|---------|---------|---------|
| Gender   | Female   | 173     | 57.3    | 179     | 56.1    |
|          | Male     | 129     | 42.7    | 140     | 43.9    |
| Age      | 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  |

*Note.* Author's results.

Both samples were similar in gender distribution and age range (12–18 years), with most participants aged 14–16. Grade levels ranged from 5th to 11th grade, with fairly even distribution across middle and high school levels.

### *Data Collection Procedures*

This study was conducted as part of a BAP (Scientific Research Project) funded by İstanbul Kültür University. Ethical approval was obtained from the Ethical Review Board of İstanbul Kültür University, and prior to data collection, permission was granted by the Istanbul Provincial Directorate of National Education to conduct the survey in secondary schools.

Data were collected during the 2024-2025 Spring term from public secondary schools in İstanbul. The 60-item scale was administered in print format by class teachers during regular class periods, with each administration taking approximately 25 minutes to complete. To ensure standardized administration procedures, a research assistant provided detailed training to class teachers, explaining how to administer the scale and what instructions to give students. Teachers were equipped with standardized scripts to maintain consistency across all participating classrooms.

Prior to data collection, students were informed that their participation was voluntary. Only students who willingly agreed to participate completed the scale forms. To maintain confidentiality, students were not required to provide identifying personal information, and demographic data collection was limited to age, gender, and grade level only. Students were assured that their responses would remain confidential and that participation or non-

participation would not affect their academic standing. The completed forms were collected immediately after administration to maintain data integrity and participant anonymity.

## DATA ANALYSIS PROCEDURES

### *Testing Assumptions*

Before conducting exploratory factor analysis, a multivariate analysis of the obtained data, the basic assumptions of this analysis were first tested. Initially, the adequacy of the sample size and the presence of missing data were examined. Regarding sample size, Kline emphasized that it could be 10 times the number of items but should be at least twice the number of items (1994). Some sources (Büyüköztürk, 2006; Tabachnick & Fidell, 2001) state that the sample size should be at least five times the total number of items. Based on this, our developed scale contains a total of 60 items, and five times that number, 300 observations, would be sufficient. In this study, 312 observations were obtained; 10 were excluded from the analysis, leaving 302 for analysis. This can be evidence that the data is sufficient for factor analysis. There are no missing values in the data obtained in the study.

Second, the normality of the distribution in the obtained dataset was examined. Kurtosis and skewness values of all values were checked. It was observed that skewness ( $-0.092 / 0.138 = -0.667$ ) and kurtosis ( $-0.956 / 0.275 = -3.476$ ) values met the normal distribution assumption. When examining the Z value, the Z value ranges between -1.913 and 1.781, and since 95% of the observations are within the -1.96 to +1.96 point range, the distribution can be said to be normal (Alpar, 2012:140). When evaluating the Kolmogorov-Smirnov test ( $K-S = 0.0453$ ;  $p = 0.545 > 0.05$ ) and Anderson-Darling ( $AD = 0.943$ ,  $p = 0.017$ ) results, along with the Shapiro-Wilk test ( $0.990$ ,  $p = 0.025$ ), it is observed that the distribution generally meets the normal assumption.

Third, multicollinearity was examined, and, for this purpose, VIF and Tolerance values (Table 2) were checked. VIF values are expected to be  $< 5$  or  $10$  (Montgomery, 2001; Field, 2009; Alpar, 2012; Büyüköztürk, 2006) and Tolerance (1-SMC) values are expected to be  $> .20$  and close to 1 (Büyüköztürk, 2006). After the analysis, it was observed that VIF values ranged between 2.091 and 4.682 and all were less than 5, while Tolerance (1-SMC) values ranged

between 0.214 and 0.478 and all were greater than 0.20. These results indicate that there is no multicollinearity problem among the variables.

**Table 2.**  
*VIF and Tolerance Values*

| Item | Tolerance | 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 |
| 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 |

| Item  | Tolerance | VIF   |
|-------|-----------|-------|
| 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 |
| 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 |

| Item  | Tolerance | VIF   |
|-------|-----------|-------|
| COL9  | 0.290     | 3.447 |
| COL10 | 0.392     | 2.548 |

*Note.* C (Citizenship), CT (Critical Thinking), CR (Creativity), CH (Character), COM (Communication), COL (Collaboration).

Fourth, to detect multivariate outliers, Z and Mahalanobis distance values (Table 3) were compared with the  $X^2(60) = 99.607$  value at the 0.001 significance level. Mahalanobis distance values ranged between 0.880 and 106.625, and 10 observations that were greater than the critical value were excluded from the analysis.

**Table 3.**  
*Mahalanobis Distances*

| Observation | Mahalanobis Distance |
|-------------|----------------------|
| 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               |
| O14         | 99.085               |
| ...         | ...                  |
| O318        | 1.336                |
| O319        | 1.213                |
| O320        | 1.213                |
| O321        | 1.213                |
| O322        | 0.880                |

*Note.* Author's results.

Fifth, Durbin-Watson values (Table 4) were examined to see whether the data were error-independent. A Durbin value between 0 and 4, and close to 2, indicates that the data are error-independent. In this analysis, the Durbin-Watson value of the data was found to be 1.89. As a result of the Durbin-Watson test, the autocorrelation value was calculated as 0.050 and the p value as 0.33. Therefore, there is no significant autocorrelation in the regression model for  $n=302$  at the 0.01 level ( $p>0.01$ ), meaning the error can be said to be independent.

**Table 4.**  
*Durbin–Watson Values*

| Model | R     | R <sup>2</sup> | Durbin–Watson | Autocorrelation | P    |
|-------|-------|----------------|---------------|-----------------|------|
| 1     | 0.525 | 0.275          | 1.89          | 0.050           | 0.33 |

*Note.* Author's results.

Finally, after excluding observations with extreme values and observations creating multicollinearity problems from the analysis, the factorability of the data consisting of the remaining 302 observations was tested. The KMO (The Kaiser-Meyer-Olkin Measure of Sampling Adequacy) value and Bartlett's Test of Sphericity value (Table 5) were examined to test the significance of the KMO value. The KMO value was 0.965, and Bartlett's Test of Sphericity result showed that the KMO value was significant ( $\chi^2=17506$ ,  $df=1770$ ,  $p<0.001$ ). A KMO value above 0.90 indicated that the sample size was excellent (Alpar, 2012).

**Table 5.**  
*KMO and Bartlett's Test Values*

| Measure                                          | Value   |
|--------------------------------------------------|---------|
| Kaiser–Meyer–Olkin Measure of Sampling Adequacy  | 0.965   |
| Bartlett's Test of Sphericity – Approx. $\chi^2$ | 17.506  |
| Bartlett's Test of Sphericity – Df               | 1.770   |
| Bartlett's Test of Sphericity – Sig.             | < 0.001 |

*Note.* Author's results.

The KMO value being above 0.95 indicates that the dataset is extremely suitable for factor analysis.

### *Exploratory Factor Analysis*

Exploratory factor analysis was conducted to examine the underlying factor structure of the preliminary scale. The analysis was performed using maximum likelihood extraction with oblimin rotation, in line with the theoretical assumption of correlated factors. The results of the EFA are presented below.

**Table 6.**  
*Factor Loadings*

| <b>Item</b> | <b>Factor 1<br/>(COL)</b> | <b>Factor 2<br/>(COM)</b> | <b>Factor 3<br/>(CR)</b> | <b>Factor 4<br/>(CH)</b> | <b>Factor 5<br/>(CT)</b> | <b>Factor 6<br/>(C)</b> | <b>Uniquene<br/>ss</b> |
|-------------|---------------------------|---------------------------|--------------------------|--------------------------|--------------------------|-------------------------|------------------------|
| COL7        | 0.852                     |                           |                          |                          |                          |                         | 0.250                  |
| COL8        | 0.837                     |                           |                          |                          |                          |                         | 0.244                  |
| COL6        | 0.776                     |                           |                          |                          |                          |                         | 0.250                  |
| COL5        | 0.774                     |                           |                          |                          |                          |                         | 0.220                  |
| COL1<br>0   | 0.696                     |                           |                          |                          |                          |                         | 0.377                  |
| 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<br>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 |

*Note.* Maximum likelihood extraction method with oblimin rotation was used.

As a result of the initial analysis, a 6-factor structure was obtained, but some items showed low factor loadings or cross-loaded on multiple factors. Model fit indices (Table 7) were calculated as RMSEA=0.0659 (0.0632-0.0691), TLI=0.850.

**Table 7.**  
*Model Fit Measures*

| Fit Index | Value  | Acceptable Threshold |
|-----------|--------|----------------------|
| RMSEA     | 0.0659 | < .08                |
| TLI       | 0.850  | > .90                |

*Note.* Author's results.

### *Item Removal Process*

During the exploratory factor analysis process, a total of 23 items were systematically excluded from the analysis. In the first cleaning phase, three critical items were removed. Item CH10 was excluded from the analysis due to having a very low factor loading of 0.317 and also showing cross-loading problems. Item COM1 was excluded from the analysis because it could not load significantly on any factor and could not contribute to the factor structure, and item C9 was removed due to showing simultaneous loading problems on multiple factors.

In the second cleaning phase, three more items with weak factor loadings were excluded from the analysis. Item C2 had a factor loading of 0.371, item COM2 had 0.378, and item C10 had 0.369, and these values remained significantly below the acceptable threshold of 0.50.

These items were excluded from the analysis on the grounds that they reduced the overall quality of the factor structure and could not provide reliable measurement.

In the third cleaning phase, focus was placed on items showing low communality values. Item CR10 was removed because it showed that the factors could not adequately explain this item with a communality value of 0.505. Item CR5 was also removed due to a communality value of 0.535.

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

As a result of this stepwise cleaning process, the initial scale consisting of 60 items was reduced to 37 items, and the model fit indices were significantly improved.

### *Second Factor Analysis (37 Items)*

As a result of the factor analysis conducted after the item removal process, it was observed that the KMO value (Table 8) was 0.960 and the inter-item compatibility index was also significant ( $\chi^2=10171$ ,  $df=666$ ,  $p<0.001$ ). The factor analysis conducted using the minimum residual extraction method and oblimin rotation technique reveals the psychometric properties of the six-dimensional structure representing 21st-century skills.

**Table 8.**

*KMO and Bartlett's Test Results*

|                                                 |                    |         |
|-------------------------------------------------|--------------------|---------|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy |                    | 0,960   |
| Bartlett's Test of Sphericity                   | Approx. Chi-Square | 10171   |
|                                                 | Df                 | 666     |
|                                                 | Sig.               | < 0,001 |

*Note.* Author's results.

The factor loadings presented in Table 9 demonstrate a clear six-factor structure that aligns with the theoretical framework of the 6C skills. Each factor shows strong loadings for items within their respective domains, with Collaboration items loading on Factor 1 (ranging from 0.616 to 0.872), Communication items on Factor 2 (0.562 to 0.898), Critical Thinking items on Factor 3 (0.595 to 0.708), Character items on Factor 4 (0.694 to 0.876), Citizenship items on Factor 5 (0.550 to 0.790), and Creativity items on Factor 6 (0.532 to 0.752). The minimum residual extraction method combined with oblimin rotation successfully revealed this

interpretable factor structure, with most items demonstrating substantial loadings above 0.60 on their primary factors and minimal cross-loadings on other factors.

**Table 9.**  
*Factor Loadings*

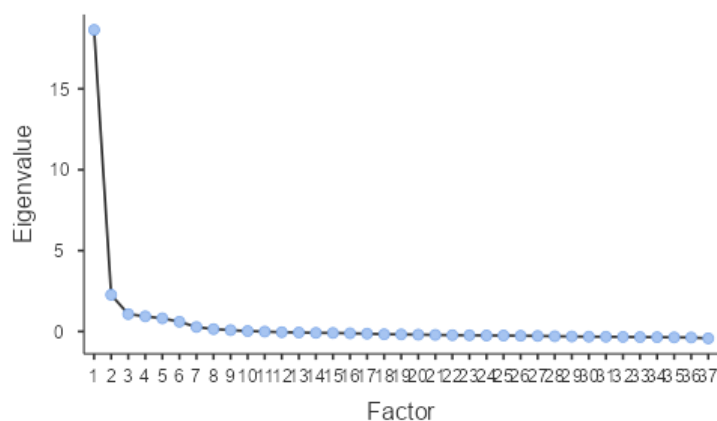
|       | Factor |       |       |       |   |   | Uniqueness |
|-------|--------|-------|-------|-------|---|---|------------|
|       | 1      | 2     | 3     | 4     | 5 | 6 |            |
| COL8  | 0.872  |       |       |       |   |   | 0.212      |
| COL7  | 0.833  |       |       |       |   |   | 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>Note. 'Minimum residual' extraction method was used in combination with a 'oblimin' rotation</i> |  |  |  |  |       |       |       |

*Note.* Author's results.

The scree plot (Figure 1) provides visual confirmation of the six-factor solution, showing a clear elbow after the sixth eigenvalue that supports retaining six factors for rotation. This eigenvalue pattern reinforces the theoretical expectation of six distinct but related skill domains in the measurement model.

**Figure 1.**  
*Factor Loadings*



*Note.* Author's results.

### *Explained Variance*

The variance explanation results (Table 10) obtained from factor analysis reveal that the six-factor structure explains a significant portion of the total variance. The first factor is positioned as the strongest factor, explaining 14.27% of the total variance with a factor score loading of 5.28. The second factor has a variance explanation rate of 13.68% with a loading value of 5.06, and the first two factors together explain 27.9% of the total variance. The third and fourth factors show almost equal performance, having variance explanation rates of 11.26% and 11.24% with factor score loadings of 4.16, respectively. These four factors together explain 50.4% of the total variance, forming the main structure of the scale. The fifth factor contributes 9.66% variance with a loading value of 3.58, while the sixth factor shows the lowest contribution by explaining 8.81% variance with a loading value of 3.26. The total variance explanation rate of the six factors was calculated as 68.9%. This rate is at a quite satisfactory level for factor analysis studies in the field of social sciences and proves that the structural validity of the scale representing 21st-century skills is strong. The relatively balanced variance distribution among factors shows that no factor is excessively dominant and each dimension occupies a meaningful place in the scale.

**Table 10.**  
*Variance Loadings*

| Factor | SS Loadings | % of Variance | Cumulative % |
|--------|-------------|---------------|--------------|
| 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         |

*Note.* Author's results.

### *Distribution and Loading Values of Factor Structure*

**Collaboration Factor (COL):** This factor contains the most comprehensive item set and is represented by seven items. In this dimension, led by item COL8 with a factor loading of 0.872, factor loadings range between 0.616-0.872. Although item COL4's relatively low

factor loading of 0.616 draws attention, it is at an acceptable level. The uniqueness values remaining in the 0.212-0.354 band indicate that the common factor strongly explains the items.

**Communication Factor (COM):** In this factor operationalized with six items, item COM7 has the highest value in the entire scale with a factor loading of 0.898. The distribution of factor loadings between 0.562-0.898 reflects the measurement consistency of communication skills. Item COM9 having the lowest loading of 0.562 indicates that this item's factorial representation is limited.

**Critical Thinking Factor (CT):** In this factor, represented by six items, item CT7 stands out with a factor loading of 0.708. The distribution of factor loadings in the range of 0.595-0.708 indicates homogeneous measurement of critical thinking skills. Item CT5 having the lowest loading of 0.595 shows that it is the weakest representative in this dimension.

**Character Factor (CH):** In this five-item factor, CH6's dominant position with a factor loading of 0.876 draws attention. The variation of factor loadings between 0.694-0.876 proves that this factor is one of the most consistent structures. All items showing loadings above 0.694 reveals the strong measurement characteristics of the character dimension.

**Citizenship Factor (C):** In this factor operationalized with six items, C5's factor loading of 0.790 stands out. Considering the distribution of factor loadings in the range of 0.550-0.790, it is observed that item C3 is at a borderline value with 0.550 but remains at an acceptable level.

**Creativity Factor (CR):** In this factor represented by five items, it is observed that CR3 exhibits strong representation with a factor loading of 0.752. Item CR8 having the lowest value in the entire scale with a factor loading of 0.532 indicates that this item's factorial explanatory power is limited.

### *Psychometric Reliability Indicators*

The use of oblimin rotation allows for inter-factor correlations, reflecting the relational nature of 21st-century skills in real life. Generally, uniqueness values remaining mostly below 0.50 proves that the factorial explanatory power of the items is at an adequate level.

These findings reveal that the six-factor theoretical structure of the 21st-century skills scale has been empirically validated and can be used as a valid measurement tool in educational assessments.

**Table 11.**  
*Model Fit Measures*

| RMSEA  | RMSEA 90% CI |        | TLI   | BIC   | Model Test |     |       |
|--------|--------------|--------|-------|-------|------------|-----|-------|
|        | Lower        | Upper  |       |       | $\chi^2$   | df  | p     |
| 0.0645 | 0.0594       | 0.0699 | 0.911 | -1585 | 1037       | 459 | <.001 |

*Note.* Author's results.

Model fit indices (Table 11) generally provide acceptable level results. The RMSEA value is 0.0645, remaining below the 0.08 threshold and indicating an acceptable fit level. The TLI value was calculated as 0.911 and remained above the 0.90 threshold value, revealing that model fit is adequate. The chi-square test was statistically significant, but this is an expected situation in large sample sizes, and other fit indices provide more reliable information.

**Table 12.**  
*Inter-Factor Correlations*

|     | 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  |     |       |       |       |       | —     |

*Note.* Author's results. COL=Collaboration, COM=Communication, CT=Critical Thinking, CH=Character, C=Citizenship, CR=Creativity.

Inter-factor correlations (Table 12) range between 0.351 and 0.694. The highest correlation is observed between Collaboration and Communication factors ( $r=0.694$ ), while the lowest correlation is seen between Critical Thinking and Creativity factors ( $r=0.351$ ). These correlations show that there are meaningful relationships between factors but discriminant validity is maintained.

**Table 13.**  
*Overall and Individual KMO Values*

| Factor                 | Items     | KMO Range (Min-Max) |
|------------------------|-----------|---------------------|
| Overall KMO            | All Items | 0.960               |
| COL (Collaboration)    | COL items | 0.951 - 0.973       |
| COM (Communication)    | COM items | 0.959 - 0.976       |
| CT (Critical Thinking) | CT items  | 0.955 - 0.970       |
| CH (Character)         | CH items  | 0.947 - 0.971       |
| C (Citizenship)        | C items   | 0.939 - 0.971       |
| CR (Creativity)        | CR items  | 0.939 - 0.962       |

*Note.* Author's results.

Assumption tests support the appropriateness of factor analysis. The KMO value (Table 13) was calculated as 0.960, indicating that the sample size is at an excellent level. Individual MSA values range between 0.934 and 0.977, revealing that sample adequacy is achieved for all items. Bartlett's sphericity test was statistically significant and confirmed that there is a correlation structure suitable for factor analysis among the variables.

#### *Confirmatory Factor Analysis*

For confirmatory factor analysis, data obtained from 319 participants was used. The sample size (319) is at an adequate level for CFA. The Maximum Likelihood estimation method was used and robust standard errors were calculated. The model converged in 116 iterations, indicating that the estimation process was successful.

**Table 14.**  
*Model Fit Indices*

| Fit Index         | Value   | Acceptance Criteria                      | Result      |
|-------------------|---------|------------------------------------------|-------------|
| $\chi^2$ (Scaled) | 926.00  | -                                        | -           |
| Df                | 611     | -                                        | -           |
| p-value           | < 0.001 | > 0.05                                   | Significant |
| RMSEA (Scaled)    | 0.040   | $\leq 0.08$ acceptable, $\leq 0.06$ good | Good fit    |
| RMSEA Lower Bound | 0.036   | -                                        | -           |

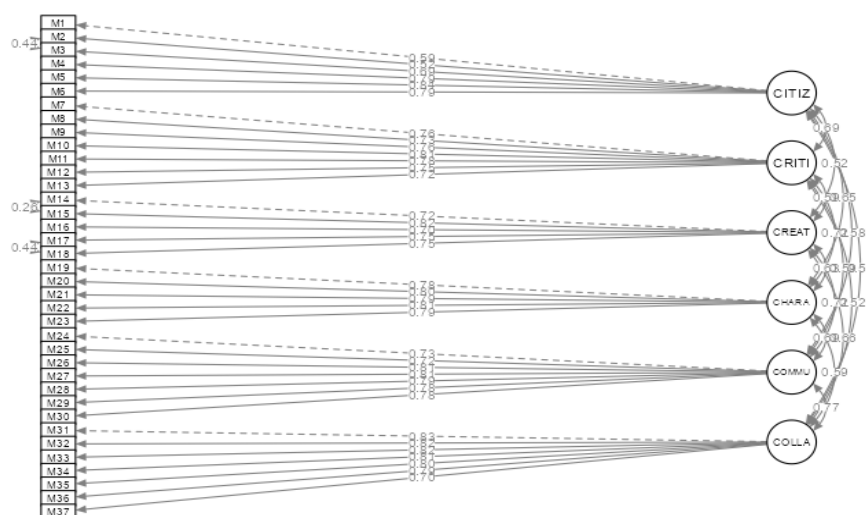
|                   |       |                                               |               |
|-------------------|-------|-----------------------------------------------|---------------|
| RMSEA Upper Bound | 0.044 | -                                             | -             |
| RMSEA p-value     | 1.000 | > 0.05                                        | Good fit      |
| CFI (Scaled)      | 0.959 | $\geq 0.95$ excellent, $\geq 0.90$ acceptable | Excellent fit |
| TLI (Scaled)      | 0.955 | $\geq 0.95$ excellent, $\geq 0.90$ acceptable | Excellent fit |
| SRMR              | 0.050 | $\leq 0.08$ good fit                          | Good fit      |
| GFI               | 0.886 | $\geq 0.90$ acceptable                        | Borderline    |

*Note.* Author's results.

Model fit indices (Table 14) evaluate how well the proposed model fits the observed data. The RMSEA (0.040), CFI (0.959), TLI (0.955), and SRMR (0.050) values reveal that the model shows excellent fit. The significant chi-square test result ( $p < 0.001$ ) is a normal situation in large sample sizes and does not require model rejection by itself.

## Path Diagram

**Figure 2.**  
*21st-century Skills Scale Confirmatory Factor Analysis Path Diagram*



Note. Author's results.

The path diagram (Figure 2) visually presents the 6-factor structure of the 37-item 21st-century skills scale. The diagram shows the CITIZENSHIP (IT1-IT6), CRITICAL THINKING (IT7-IT13), CREATIVITY (IT14-IT18), CHARACTER (IT19-IT23), COMMUNICATION (IT24-IT30), and COLLABORATION (IT31-IT37) factors and the factor loadings of items belonging to these factors. Standardized factor loadings range between 0.525 and 0.844, with 35 out of 37 items having strong factor loadings above 0.70. Inter-factor correlations range between 0.521-0.769, with the highest correlation observed between COMMUNICATION and COLLABORATION factors ( $r=0.769$ ), and the lowest correlation between CRITICAL THINKING and COLLABORATION factors ( $r=0.521$ ).

These results reveal that the theoretical structure of the six-factor model is strongly supported by empirical data. The majority of factor loadings being above 0.70 shows that each factor adequately explains its own items. The moderate-high level correlations between factors (below 0.85) confirm that 21st-century skills are related but distinguishable structures. Three error covariance modifications made in the model ( $IT2 \leftrightarrow IT3$ ,  $IT14 \leftrightarrow IT15$ ,  $IT17 \leftrightarrow IT18$ )

improved model fit, and excellent fit indices were obtained (RMSEA=0.040, CFI=0.959, TLI=0.955). These modifications were theoretically justified as they involved consecutive items with similar content characteristics. These findings prove that the scale can measure 21st-century skills validly and reliably. All factor loadings are statistically significant at the  $p < 0.001$  level. Standardized coefficients are shown.

### *Factor Loadings and Significance Levels*

Table 15 displays the factor loadings and associated statistical parameters for all items within their respective factors. The standardized factor loadings ( $\beta$ ) represent the strength of association between each item and its underlying construct, accompanied by standard errors, z-statistics, and significance levels that confirm the validity of the factor structure.

#### *CITIZENSHIP*

**Table 15.**  
*Factor Loadings*

| Factor Name       | Item | Factor Loading ( $\beta$ ) | Standard Error | z-value | p-value |
|-------------------|------|----------------------------|----------------|---------|---------|
| CITIZENSHIP       | 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 |
| CRITICAL THINKING | 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 |
| CREATIVITY        | 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 |
| CHARACTER         | 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 |
| COMMUNICATION | 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 |
|               | IT30 | 0.776 | 0.072 | 15.15 | < 0.001 |
| COLLABORATION | 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 |

*Note.* Author's results.

All items in the CITIZENSHIP factor are statistically significant ( $p < 0.001$ ). Factor loadings range from 0.525-0.844. IT1 is fixed at 1.00 as it is the reference item. IT5 has the highest factor loading (0.844), while IT2 has the lowest factor loading (0.525). All factor loadings are above 0.50, indicating they are acceptable.

The CRITICAL THINKING factor has strong factor loadings (ranging from 0.724-0.814). All items are highly significant ( $z > 14$ ). IT10 has the highest factor loading (0.814), while IT13 has the lowest factor loading (0.724). This factor exhibits a homogeneous structure.

Factor loadings in the CREATIVITY factor range from 0.700 to 0.822. IT15 has the highest factor loading (0.822) and the highest z-value (17.75). A modification was made between IT17 and IT18 (covariance=1.327). All factor loadings are above the acceptable level.

The CHARACTER factor has very consistent factor loadings (ranging from 0.781-0.806). Factor loadings are very close to each other, indicating a homogeneous factor structure. IT22 has the highest loading factor (0.806). All z-values are above 15.

COMMUNICATION is the largest factor with 7 items. Factor loadings range from 0.723-0.811. IT26 and IT27 have the highest loadings (0.810 and 0.811, respectively). This factor exhibits strong internal consistency. All items are highly significant.

The COLLABORATION factor has the most substantial factor loadings (ranging from 0.700 to 0.826). M31, M32, and M33 have particularly high factor loadings ( $>0.80$ ). This factor has the highest z-values (ranging from 17.55 to 21.57), reflecting a very strong factor structure. Although M37 has the lowest factor loading, it is still at an acceptable level of 0.70.

### Reliability Analysis

**Table 16.**  
*Reliability Analysis*

| Factor            | Cronbach's $\alpha$ | Composite Reliability ( $\omega$ ) | AVE   |
|-------------------|---------------------|------------------------------------|-------|
| CITIZENSHIP       | 0.866               | 0.833                              | 0.503 |
| CRITICAL THINKING | 0.905               | 0.905                              | 0.577 |
| CREATIVITY        | 0.876               | 0.835                              | 0.563 |
| CHARACTER         | 0.895               | 0.895                              | 0.630 |
| COMMUNICATION     | 0.913               | 0.913                              | 0.600 |
| COLLABORATION     | 0.923               | 0.923                              | 0.633 |

Note. Author's results.

Reliability analysis results (Table 16) show that all factors have high internal consistency. Cronbach's  $\alpha$  values range from 0.866 to 0.923, all above the 0.70 criterion. Composite reliability ( $\omega$ ) values are similarly high. AVE (Average Variance Extracted) values are above the 0.50 criterion, supporting convergent validity. COLLABORATION has the highest reliability, while CITIZENSHIP has an AVE value at the threshold (0.503).

### Inter-Factor Correlations

**Table 17.**  
*Inter-factor Correlations*

|                   | Citizenship | Critical Thinking | Creativity | Character | Communication | Collaboration |
|-------------------|-------------|-------------------|------------|-----------|---------------|---------------|
| Citizenship       | 1.000       | 0.692**           | 0.522**    | 0.649**   | 0.577**       | 0.540**       |
| Critical Thinking | 0.692**     | 1.000             | 0.590**    | 0.718**   | 0.590**       | 0.521**       |
| Creativity        | 0.522**     | 0.590**           | 1.000      | 0.630**   | 0.716**       | 0.656**       |
| Character         | 0.649**     | 0.718**           | 0.630**    | 1.000     | 0.694**       | 0.593**       |
| Communication     | 0.577**     | 0.590**           | 0.716**    | 0.694**   | 1.000         | 0.769**       |

|                      |         |         |         |         |         |       |
|----------------------|---------|---------|---------|---------|---------|-------|
| <b>Collaboration</b> | 0.540** | 0.521** | 0.656** | 0.593** | 0.769** | 1.000 |
|----------------------|---------|---------|---------|---------|---------|-------|

Note. Author's results. \*\*  $p < 0.001$

The inter-factor correlation matrix (Table 17) shows the relationships among the six factors. Correlations range from 0.521 to 0.769. The highest correlation is between COMMUNICATION and COLLABORATION ( $r = 0.769$ ). The lowest correlation is between CRITICAL THINKING and COLLABORATION ( $r = 0.521$ ). All correlations are statistically significant ( $p < 0.001$ ). Correlations below 0.85 indicate that the factors are distinguishable.

#### *HTMT (Heterotrait-Monotrait) Ratio Matrix*

**Table 18.**

#### *HTMT (Heterotrait-Monotrait) Ratio Matrix*

|                          | <b>Citizenship</b> | <b>Critical Thinking</b> | <b>Creativity</b> | <b>Character</b> | <b>Communication</b> | <b>Collaboration</b> |
|--------------------------|--------------------|--------------------------|-------------------|------------------|----------------------|----------------------|
| <b>Citizenship</b>       | -                  | 0.680                    | 0.558             | 0.651            | 0.607                | 0.580                |
| <b>Critical Thinking</b> | 0.680              | -                        | 0.575             | 0.720            | 0.582                | 0.522                |
| <b>Creativity</b>        | 0.558              | 0.575                    | -                 | 0.615            | 0.705                | 0.639                |
| <b>Character</b>         | 0.651              | 0.720                    | 0.615             | -                | 0.687                | 0.589                |
| <b>Communication</b>     | 0.607              | 0.582                    | 0.705             | 0.687            | -                    | 0.773                |
| <b>Collaboration</b>     | 0.580              | 0.522                    | 0.639             | 0.589            | 0.773                | -                    |

Note. HTMT values should be below 0.90. All values meet this criterion.

The HTMT (Heterotrait-Monotrait) ratio matrix (Table 18) is used to assess the discriminant validity of the factors. All HTMT values are below the 0.90 criterion, indicating that discriminant validity is achieved between factors. The highest HTMT value is between COMMUNICATION and COLLABORATION (0.773). The lowest HTMT value is between CRITICAL THINKING and COLLABORATION (0.522). These results support that the factors are distinguishable constructs. These results indicate that the 6-factor model is a

statistically acceptable and reliable structure. This model can be used as a valid and reliable instrument for measuring 21st-century skills.

### *Second-Order Confirmatory Factor Analysis (Second-Order CFA) Results*

The analysis tested whether the six first-order factors (F1-F6) can be aggregated under a single higher-order factor (SIXCS).

### *Model Fit Indices*

The second-order CFA model demonstrates excellent fit (Table 19). The RMSEA (0.042), CFI (0.954), TLI (0.951), and SRMR (0.059) values reveal that the model achieves very good fit with the data. These results support the claim that the first six-order factors can be meaningfully aggregated into a single higher-order factor. The fit indices are very similar to those of the first-order CFA, indicating that the hierarchical structure is appropriate.

**Table 19.**  
*Model Fit Indices*

| Fit Index         | Value   | Acceptance Criteria                           | Result        |
|-------------------|---------|-----------------------------------------------|---------------|
| $\chi^2$ (Scaled) | 969.00  | -                                             | -             |
| df                | 620     | -                                             | -             |
| p-value           | < 0.001 | > 0.05                                        | Significant   |
| RMSEA (Scaled)    | 0.042   | $\leq 0.08$ acceptable, $\leq 0.06$ good      | Good fit      |
| RMSEA Lower Limit | 0.038   | -                                             | -             |
| RMSEA Upper Limit | 0.046   | -                                             | -             |
| RMSEA p-value     | 0.999   | > 0.05                                        | Good fit      |
| CFI (Scaled)      | 0.954   | $\geq 0.95$ excellent, $\geq 0.90$ acceptable | Excellent fit |
| TLI (Scaled)      | 0.951   | $\geq 0.95$ excellent, $\geq 0.90$ acceptable | Excellent fit |
| SRMR              | 0.059   | $\leq 0.08$ good fit                          | Good fit      |
| GFI               | 0.880   | $\geq 0.90$ acceptable                        | At threshold  |

*Note.* Author's results.

*Second-Order Factor Loadings***Table 20.***Second-Order Loadings (SIXCS → First-Order Factors)*

| <b>Factor</b>                  | <b>Factor Loading (β)</b> | <b>Standard Error</b> | <b>Lower Limit</b> | <b>Upper Limit</b> | <b>z-value</b> | <b>p-value</b> |
|--------------------------------|---------------------------|-----------------------|--------------------|--------------------|----------------|----------------|
| SIXCS → F1 (CITIZENSHIP)       | 0.734                     | -                     | 0.661              | 0.807              | -              | -              |
| SIXCS → F2 (CRITICAL THINKING) | 0.765                     | 0.152                 | 0.698              | 0.832              | 8.79           | < 0.001        |
| SIXCS → F3 (CREATIVITY)        | 0.791                     | 0.158                 | 0.712              | 0.869              | 8.53           | < 0.001        |
| SIXCS → F4 (CHARACTER)         | 0.828                     | 0.186                 | 0.770              | 0.885              | 8.44           | < 0.001        |
| SIXCS → F5 (COMMUNICATION)     | 0.861                     | 0.175                 | 0.794              | 0.928              | 8.42           | < 0.001        |
| SIXCS → F6 (COLLABORATION)     | 0.785                     | 0.180                 | 0.712              | 0.859              | 8.57           | < 0.001        |

*Note.* Author's results.

Table 20 shows the factor loadings of the higher-order SIXCS factor on the six first-order factors. All factor loadings are above 0.70 and statistically highly significant ( $p < 0.001$ ). COMMUNICATION has the highest factor loading (0.861), while CITIZENSHIP has the lowest factor loading (0.734). These high factor loadings strongly support the view that the six skill domains are indeed components of a single higher-order “21st-century skills” construct. F1 serves as the reference factor; therefore, no z-value is available.

### Reliability Analysis

**Table 21.**  
*Reliability Analysis*

| Factor                 | Cronbach's $\alpha$ | Composite Reliability ( $\omega$ ) | AVE   | Result                             |
|------------------------|---------------------|------------------------------------|-------|------------------------------------|
| F1 (CITIZENSHIP)       | 0.866               | 0.836                              | 0.506 | Good reliability, AVE at threshold |
| F2 (CRITICAL THINKING) | 0.905               | 0.905                              | 0.578 | Very good reliability              |
| F3 (CREATIVITY)        | 0.876               | 0.835                              | 0.563 | Very good reliability              |
| F4 (CHARACTER)         | 0.895               | 0.895                              | 0.629 | Very good reliability              |
| F5 (COMMUNICATION)     | 0.913               | 0.913                              | 0.599 | Excellent reliability              |
| F6 (COLLABORATION)     | 0.923               | 0.923                              | 0.633 | Excellent reliability              |

*Note.* Author's results.

The reliability analysis results (Table 21) are identical to those of the first-order CFA and confirm that all factors demonstrate high internal consistency. All Cronbach's  $\alpha$  values are above 0.86, indicating excellent reliability levels. AVE values are above the 0.50 criterion, supporting convergent validity.

### Model Comparison: First-Order vs Second-Order

**Table 22.**  
*First-Order vs Second-Order*

| Fit Index         | First-Order CFA | Second-Order CFA | Difference | Evaluation       |
|-------------------|-----------------|------------------|------------|------------------|
| $\chi^2$ (Scaled) | 926             | 969              | +43        | Small increase   |
| df                | 611             | 620              | +9         | More constraints |
| RMSEA (Scaled)    | 0.040           | 0.042            | +0.002     | Nearly identical |
| CFI (Scaled)      | 0.959           | 0.954            | -0.005     | Minimal decrease |

|              |       |       |        |                  |
|--------------|-------|-------|--------|------------------|
| TLI (Scaled) | 0.955 | 0.951 | -0.004 | Minimal decrease |
| SRMR         | 0.050 | 0.059 | +0.009 | Small increase   |

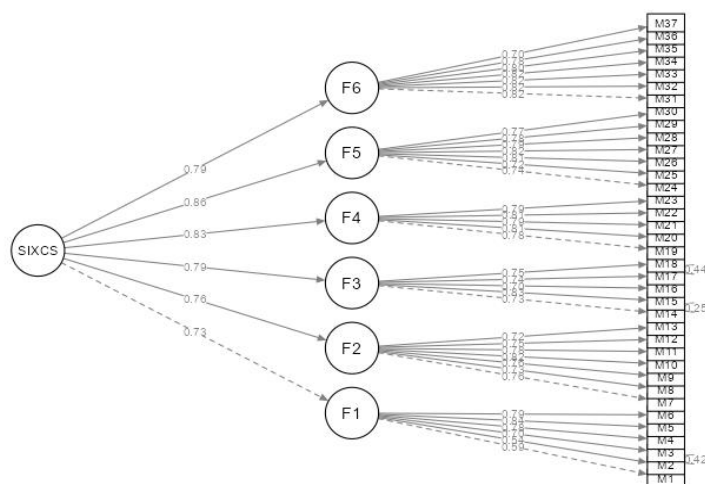
Note. Author's results.

The model comparison reveals that the second-order CFA shows a very similar fit to the first-order CFA (Table 22). The minimal differences in fit indices ( $\Delta\text{RMSEA} = 0.002$ ,  $\Delta\text{CFI} = 0.005$ ) support the appropriateness of the hierarchical structure for the data. These results indicate that the simpler second-order model can be preferred according to the parsimony principle. A simpler, theoretically more meaningful model was obtained with minimal loss of fit.

### Path Diagram

**Figure 3.**

*Second-Order CFA Path Diagram*



Note. Author's results.

The second-order CFA path diagram (Figure 3) clearly visualizes the hierarchical factor structure. The SIXCS higher-order factor is located on the left side of the diagram, while the first six order factors (F1-F6) are positioned on the right side. The arrows drawn from the SIXCS factor toward the F1-F6 factors contain the factor loadings (ranging from 0.734 to 0.861)

that indicate the extent to which the higher-order factor explains the sub-factors. The arrows extending from each first-order factor to its respective items show the item-factor relationships.

This hierarchical structure supports that 21st-century skills are sub-dimensions that unite under a single general construct while maintaining their unique characteristics. The COMMUNICATION (0.861) and CHARACTER (0.828) factors represent the general structure most strongly, while the CITIZENSHIP (0.734) factor provides the lowest but still strong contribution. The model presents a theoretically cleaner structure, free of direct correlations between first-order factors, and demonstrates that a single higher-order structure can explain 21st-century skills. All factor loadings are statistically significant at the  $p < 0.001$  level. Model fit indices:  $\chi^2=969$ ,  $df=620$ ,  $RMSEA=0.042$ ,  $CFI=0.954$ ,  $TLI=0.951$ ,  $SRMR=0.059$ .

### *Scoring Approach*

The second-order CFA model offers distinct advantages over the first-order model in terms of scoring the 21st-century skills scale. This hierarchical structure allows for separate scoring of the six sub-dimensions (CITIZENSHIP, CRITICAL THINKING, CREATIVITY, CHARACTER, COMMUNICATION, COLLABORATION) and the calculation of a general SIXCS score encompassing all items. In individual assessments, sub-dimension scores (calculated by taking the average of each factor's items) are used for detailed profile analysis, while the general SIXCS score (average of all 37 items) summarizes individuals' overall 21st-century skills level in a single value. This dual-level scoring approach provides flexibility for both identifying developmental areas and making intra-group rankings, while weighted scoring based on factor loadings (using factor loadings ranging from 0.734-0.861) offers more precise general assessment. The model fit indices being nearly identical to the first-order CFA ( $\Delta CFI=0.005$ ) support the psychometric validity of this scoring approach and endorse the preference for the second-order model in practical applications.

### *Scale Scoring*

The scoring of the 21st-century skills scale (Table 23) can be implemented through two different approaches based on second-order confirmatory factor analysis results. In the factor-based scoring approach, each of the six sub-dimensions (citizenship, critical thinking, creativity, character, communication, collaboration) is calculated separately by taking the

arithmetic mean of its respective items. This approach provides detailed profile analysis to determine which skill areas individuals are strong or weak in, and it provides a score range of 1.00-9.00 for each factor. The number of items varies across factors: the citizenship factor contains six items, the critical thinking and communication factors each contain seven items, the creativity and character factors each contain 5 items, and the collaboration factor contains seven items.

**Table 23.**

*21st-century Skills Scale Scoring Ranges*

| <b>Factor</b>          | <b>Number of Items</b> | <b>Item Numbers</b> | <b>Total Score Range</b> | <b>Average Score Range</b> | <b>Factor Loading</b> |
|------------------------|------------------------|---------------------|--------------------------|----------------------------|-----------------------|
| Citizenship (F1)       | 6                      | IT1-IT6             | 6-54                     | 1.00-9.00                  | 0.734                 |
| Critical Thinking (F2) | 7                      | IT7-IT13            | 7-63                     | 1.00-9.00                  | 0.765                 |
| Creativity (F3)        | 5                      | IT14-IT18           | 5-45                     | 1.00-9.00                  | 0.791                 |
| Character (F4)         | 5                      | IT19-IT23           | 5-45                     | 1.00-9.00                  | 0.828                 |
| Communication (F5)     | 7                      | IT24-IT30           | 7-63                     | 1.00-9.00                  | 0.861                 |
| Collaboration (F6)     | 7                      | IT31-IT37           | 7-63                     | 1.00-9.00                  | 0.785                 |
| SIXCS Overall          | 37                     | IT1-IT37            | 37-333                   | 1.00-9.00                  | -                     |

*Note.* Author's results.

The second-order general scoring approach involves calculating the SIXCS factor score as the average score of all 37 items, summarizing the individual's overall 21st-century skills level in a single value. This general score can be calculated using the simple arithmetic mean (sum of all items divided by 37) or a weighted average based on factor loadings ( $F1 \times 0.734 + F2 \times 0.765 + F3 \times 0.791 + F4 \times 0.828 + F5 \times 0.861 + F6 \times 0.785$ ). The weighted scoring method provides psychometrically more sensitive results as it considers the contributions of factors to the higher-order structure and increases the theoretical validity of the general score. Both scoring approaches can be interpreted within the level ranges of 1.00-3.00 (low), 3.01-5.00 (moderate), 5.01-7.00 (good), and 7.01-9.00 (high) (Table 24). The final scale form is given in the Appendix.

**Table 24.***Score Interpretation Criteria*

| Score Range | Level    | Description                             |
|-------------|----------|-----------------------------------------|
| 1.00-3.00   | Low      | High developmental need in skill area   |
| 3.01-5.00   | Moderate | Moderate level competency in skill area |
| 5.01-7.00   | Good     | Good level competency in skill area     |
| 7.01-9.00   | High     | High level competency in skill area     |

Note. Author's results.

*Reliability Analysis*

The reliability analysis results (Table 25) indicate that the 21st-century skills scale demonstrates high internal consistency at both the general and factor levels. The overall scale's Cronbach's  $\alpha$  was 0.962, indicating excellent reliability. McDonald's  $\omega$  value was also found to be 0.962, confirming the  $\alpha$  coefficient. When examined at the factor level, the reliability coefficients of the six sub-dimensions range from 0.866 to 0.923, with the lowest reliability observed in the Citizenship factor ( $\alpha=0.866$ ) and the highest reliability in the Collaboration factor ( $\alpha=0.923$ ). While the Critical Thinking ( $\alpha=0.905$ ), Communication ( $\alpha=0.913$ ), and Collaboration ( $\alpha=0.923$ ) factors have excellent reliability, the Citizenship ( $\alpha=0.866$ ), Creativity ( $\alpha=0.876$ ), and Character ( $\alpha=0.895$ ) factors have good reliability.

The item-level analysis results show that all items contribute to the scale and none need to be removed. Item-rest correlation values range from 0.501 to 0.780, with all items above the critical threshold value of 0.30. Although item IT2 (0.501) has the lowest item-rest correlation in the overall scale (Table X), this value is still at an acceptable level. The items with the highest item-rest correlations were identified as IT29 (0.706), IT27 (0.691), and IT31 (0.696). The change in Cronbach's  $\alpha$  when items are removed remains minimal (ranging from 0.960 to 0.962), and no item is recommended for removal. These results support that all items in the scale function appropriately according to the theoretical structure and that the scale is a reliable instrument in terms of psychometric properties.

**Table 25.***21st-century Skills Scale Reliability Analysis Results*

| Factor                       | Number of Items | Mean        | Standard Deviation | Cronbach's $\alpha$ | McDonald's $\omega$ | Item-Rest Correlation Range |              |
|------------------------------|-----------------|-------------|--------------------|---------------------|---------------------|-----------------------------|--------------|
|                              |                 |             |                    |                     |                     | Lowest                      | Highest      |
| <b>Overall Scale (SIXCS)</b> | <b>37</b>       | <b>5.90</b> | <b>1.66</b>        | <b>0.962</b>        | <b>0.962</b>        | <b>0.501</b>                | <b>0.706</b> |
| Citizenship (F1)             | 6               | 5.64        | 1.97               | 0.866               | 0.868               | 0.587 (IT2)                 | 0.730 (IT3)  |
| Critical Thinking (F2)       | 7               | 6.33        | 1.91               | 0.905               | 0.905               | 0.685 (IT13)                | 0.769 (IT10) |
| Creativity (F3)              | 5               | 5.63        | 2.15               | 0.876               | 0.877               | 0.646 (IT16)                | 0.767 (IT15) |
| Character (F4)               | 5               | 6.20        | 2.12               | 0.895               | 0.895               | 0.722 (IT19)                | 0.752 (IT20) |
| Communication (F5)           | 7               | 5.87        | 2.08               | 0.913               | 0.913               | 0.693 (IT25)                | 0.769 (IT27) |
| Collaboration (F6)           | 7               | 5.71        | 2.19               | 0.923               | 0.924               | 0.674 (IT37)                | 0.780 (IT31) |

*Note.* Author's results. Item-rest correlation threshold was accepted as 0.30. All items are above this threshold. Reliability evaluation criteria are interpreted as follows: alpha values of 0.90 or higher indicate excellent reliability, values between 0.80 and 0.89 indicate good reliability, values between 0.70 and 0.79 indicate acceptable reliability, and values below 0.70 indicate low reliability.

**Teachers' 21st-century Skills Implementation Competency Evaluations by Student Gender**

Analyses conducted by student gender (Table 26) reveal that female and male students evaluate their teachers' 21st-century skills implementation competencies at similar levels. Independent sample t-test results show that there are no statistically significant differences between groups in five of the six skill dimensions and in the overall total score ( $p > .05$ ). The only notable finding is a marginally significant difference observed in the critical thinking dimension ( $p = .064$ , Cohen's  $d = 0.210$ ); in this dimension, female students evaluate their teachers higher than male students (45.54 vs 42.74, mean difference = 2.80). The largest difference in other dimensions is observed in the creativity area, where male students evaluate their teachers slightly higher than female students (28.50 vs 27.92). However, this difference is not statistically significant ( $p = .631$ ).

**Table 26.***Teachers' 21st-century Skills Implementation Competency Evaluations by Student Gender*

| Skill Area        | Student Gender | N   | Mean   | SD    | df  | t-value | p-value | Cohen's d | Effect Size | Result                 |
|-------------------|----------------|-----|--------|-------|-----|---------|---------|-----------|-------------|------------------------|
| 6C SKILLS (ALL)   | Female         | 179 | 220.22 | 64.24 | 317 | 0.611   | 0.541   | 0.069     | Very small  | Not significant        |
|                   | Male           | 140 | 216.00 | 57.59 |     |         |         |           |             |                        |
| Citizenship       | Female         | 179 | 33.93  | 12.20 | 317 | 0.205   | 0.837   | 0.023     | Very small  | Not significant        |
|                   | Male           | 140 | 33.73  | 11.35 |     |         |         |           |             |                        |
| Critical Thinking | Female         | 179 | 45.54  | 13.60 | 317 | 1.861   | 0.064   | 0.210     | Small       | Marginally significant |
|                   | Male           | 140 | 42.74  | 13.04 |     |         |         |           |             |                        |
| Creativity        | Female         | 179 | 27.92  | 11.30 |     |         | 0.631   | -0.054    | Very small  |                        |

|                           |        |         |       |           |             |                |           |        |               |                        |
|---------------------------|--------|---------|-------|-----------|-------------|----------------|-----------|--------|---------------|------------------------|
|                           | Male   | 14<br>0 | 28.50 | 9.9<br>3  | 3<br>1<br>7 | -<br>0.48<br>1 |           |        |               | Not<br>significa<br>nt |
| <b>Character</b>          | Female | 17<br>9 | 31.43 | 11.<br>00 | 3<br>1<br>7 | 0.74<br>3      | 0.45<br>8 | 0.084  | Very<br>small | Not<br>significa<br>nt |
|                           | Male   | 14<br>0 | 30.53 | 10.<br>00 |             |                |           |        |               |                        |
| <b>Communic<br/>ation</b> | Female | 17<br>9 | 41.54 | 15.<br>20 | 3<br>1<br>7 | 0.57<br>0      | 0.56<br>9 | 0.064  | Very<br>small | Not<br>significa<br>nt |
|                           | Male   | 14<br>0 | 40.61 | 13.<br>76 |             |                |           |        |               |                        |
| <b>Collaborati<br/>on</b> | Female | 17<br>9 | 39.93 | 16.<br>20 | 3<br>1<br>7 | -<br>0.04<br>8 | 0.96<br>1 | -0.005 | Very<br>small | Not<br>significa<br>nt |
|                           | Male   | 14<br>0 | 40.01 | 14.<br>24 |             |                |           |        |               |                        |

*Note.* Author's results. Effect size interpretation criteria (Cohen's d):  $d < 0.20$  = Very small effect;  $0.20 \leq d < 0.50$  = Small effect;  $0.50 \leq d < 0.80$  = Medium effect;  $d \geq 0.80$  = Large effect.

These findings offer important educational implications regarding teacher-student interaction. The fact that female and male students perceive their teachers' pedagogical competencies similarly in general indicates that teachers exhibit gender-neutral approaches in classroom management and instructional practices. The marginally significant difference in the critical thinking dimension may indicate that female students tend to evaluate their teachers' practices more positively in this skill area. However, the very small effect sizes across all dimensions (Cohen's  $d < 0.20$ ) show that these differences are practically insignificant. The imbalance in sample sizes ( $N_1=179$ ,  $N_2=140$ ) and the violation of variance homogeneity assumption in some dimensions are factors that should be considered when interpreting the results. In general evaluation, these results support that teachers' competencies in implementing 6C skills are perceived independently of student gender and provide a positive indicator in terms of gender equality in educational environments.

### *Teachers' 21st-century Skills Implementation Competency Evaluations by Grade Level*

One-Way ANOVA analysis results (Table 27) by grade level show that there are no statistically significant differences in 5th-11th grade students' evaluations of their teachers' 21st-century skills implementation competencies. According to Welch ANOVA test results, p-values exceeded 0.05 in all six skill dimensions (citizenship, critical thinking, creativity, character, communication, collaboration) and in the overall total score ( $p = 0.254 - 0.717$  range). These findings reveal that students at different grade levels perceive their teachers' pedagogical competencies at similar levels. Tukey post-hoc tests also detected no significant differences in any pairwise comparisons, supporting that differences between grade levels are random.

**Table 27.**

*Teachers' 21st-century Skills Implementation Competency Evaluations by Grade Level*

| Skill Area                | <i>F</i> | <i>df</i> | <i>p</i> |
|---------------------------|----------|-----------|----------|
| Overall Total (6C Skills) | 1.020    | 6, 109    | .416     |
| Citizenship               | 1.320    | 6, 111    | .254     |
| Critical Thinking         | 1.172    | 6, 109    | .326     |
| Creativity                | 1.131    | 6, 109    | .349     |
| Character                 | 1.044    | 6, 110    | .401     |
| Communication             | 0.616    | 6, 109    | .717     |
| Collaboration             | 0.973    | 6, 109    | .447     |

*Note.* Author's results. Welch's ANOVA was used due to unequal group sizes. All p-values > .05, indicating no significant differences between grade levels.

When descriptive statistics (Table 28) are examined, it is observed that 6th grade students generally evaluated their teachers with the highest scores (overall total: 234.3, critical thinking: 47.6, character: 33.7, communication: 44.5), while 11th grade students made the lowest evaluations in most dimensions (overall total: 205.9, critical thinking: 41.1, character: 29.1, collaboration: 36.1). This trend can be interpreted as teacher expectations rising with increasing student age and evaluation criteria becoming more stringent. It is noteworthy that 5th grade students gave the highest scores in creativity and citizenship dimensions. From an educational perspective, these results show that teachers exhibit consistent pedagogical performance at different grade levels and implement effective practices in 21st-century skills instruction regardless of age group.

**Table 28.***Descriptive Statistics for 21st-century Skills by Grade Level*

| Grade Level | n  | Overall Total    | Citizenship     | Critical Thinking | Creativity      | Character       | Communication | Collaboration |
|-------------|----|------------------|-----------------|-------------------|-----------------|-----------------|---------------|---------------|
| 5th         | 18 | 226.7<br>(60.70) | 36.6 (9.67)     | 43.7<br>(13.31)   | 31.3 (9.64)     | 31.9<br>(9.55)  | 39.8 (13.85)  | 43.5 (14.05)  |
| 6th         | 57 | 234.3<br>(70.29) | 36.0<br>(12.56) | 47.6<br>(12.76)   | 29.7<br>(12.41) | 33.7<br>(11.05) | 44.5 (15.26)  | 42.8 (16.35)  |
| 7th         | 72 | 215.9<br>(52.42) | 34.8<br>(11.47) | 44.0<br>(12.71)   | 26.8 (9.65)     | 29.9<br>(10.07) | 40.8 (12.62)  | 39.6 (13.29)  |
| 8th         | 38 | 223.3<br>(56.28) | 34.2 (9.17)     | 45.4<br>(12.67)   | 30.0<br>(10.54) | 32.2<br>(10.32) | 39.8 (14.99)  | 41.7 (15.39)  |
| 9th         | 54 | 211.1<br>(60.02) | 31.2<br>(12.08) | 43.4<br>(13.68)   | 26.8 (9.78)     | 30.0<br>(11.18) | 40.3 (15.27)  | 39.4 (13.99)  |
| 10th        | 38 | 214.2<br>(67.00) | 33.6<br>(13.82) | 43.9<br>(16.27)   | 26.7<br>(10.68) | 31.1<br>(9.43)  | 41.1 (14.98)  | 37.9 (15.81)  |
| 11th        | 42 | 205.9<br>(62.34) | 31.3<br>(11.69) | 41.1<br>(12.66)   | 28.4<br>(11.77) | 29.1<br>(11.52) | 39.9 (15.60)  | 36.1 (18.42)  |

*Note.* Author's results. Values represent M (SD). Total N = 319. No significant differences were found between grade levels on any measure using Tukey's HSD post hoc comparisons.

The findings support earlier research suggesting that 6C skills require active teacher involvement and intentional instructional design. Students reported higher competence in areas where teachers used collaborative and communication-focused tasks, consistent with Fullan's deep learning model. The validated scale provides new evidence that students can offer meaningful judgments about teachers' 6C implementation, complementing teacher self-reports found in previous studies.

## DISCUSSION

The primary objective of this study was to develop and validate a comprehensive scale for evaluating teachers' competence in implementing and promoting 6C skills from the student perspective. The systematic five-phase development process resulted in a psychometrically robust 37-item instrument that demonstrates excellent reliability and validity characteristics, addressing a significant gap in educational assessment.

### *Psychometric Properties and Theoretical Validation*

The scale development process yielded compelling evidence for both the theoretical coherence and empirical validity of the 6C framework. The successful extraction of six distinct factors corresponding to the theoretical dimensions of Citizenship, Critical Thinking, Creativity, Character, Communication, and Collaboration provides strong empirical support for conceptualizing these as separate but related competencies. This finding is particularly significant as it demonstrates that these theoretical constructs can be observed and distinguished in actual classroom practice.

The excellent overall reliability coefficient ( $\alpha = 0.962$ ) and robust factor-level reliabilities (ranging from 0.866 to 0.923) substantially exceed conventional standards for educational measurement instruments. These results indicate that the scale provides consistent and dependable measurement across the six competency domains, with the Collaboration factor demonstrating the highest internal consistency ( $\alpha = 0.923$ ) and the Citizenship factor showing the lowest, though still acceptable, reliability ( $\alpha = 0.866$ ).

The confirmatory factor analysis results provide compelling evidence for the structural validity of the instrument. The excellent fit indices (RMSEA = 0.040, CFI = 0.959, TLI = 0.955, SRMR = 0.050) demonstrate that the six-factor model adequately represents the underlying structure of teacher competencies in implementing 6C skills. Furthermore, the successful validation of the second-order model with minimal deterioration in fit indices ( $\Delta$ RMSEA = 0.002,  $\Delta$ CFI = 0.005) supports the proposition that the six competency areas can be conceptualized as manifestations of a higher-order construct representing overall 21st-century pedagogical competence.

### *Factor Structure and Inter-relationships*

The factor loading patterns reveal important insights into how students perceive different aspects of 21st-century teaching competencies. The Communication factor demonstrated the strongest relationship with the higher-order factor (loading = 0.861), suggesting that students view communication skills as central to effective 21st-century pedagogy. This is followed closely by Character (loading = 0.828), indicating that students highly value teachers who model ethical and moral behaviors.

The inter-factor correlations ranging from 0.351 to 0.694 reveal interesting patterns of relationships among the 6C skills. The particularly strong correlation between Collaboration and Communication ( $r = 0.694$ ) aligns with expectations, as these skills are inherently interconnected in classroom practice. Conversely, the relatively weaker correlation between Critical Thinking and Creativity ( $r = 0.351$ ) suggests that these represent more distinct pedagogical approaches, supporting their separate conceptualization in the framework.

### *Student Perceptions and Demographic Invariance*

The absence of significant differences across gender and grade levels provides important insights into the universality of student perceptions regarding effective 21st-century pedagogy. These findings suggest that effective teaching practices in implementing 6C skills transcend demographic boundaries and are recognized consistently across diverse student populations. This invariance across demographics strengthens the instrument's utility across different student groups and educational contexts.

The marginal significance observed in the Critical Thinking dimension between female and male students ( $p = 0.064$ , Cohen's  $d = 0.210$ ) warrants careful interpretation. While this finding approaches statistical significance, the small effect size suggests limited practical importance. This pattern may reflect subtle differences in how students of different genders perceive or value critical thinking instruction, though such interpretations should be made cautiously given the borderline significance level.

The stability of perceptions across grade levels (5th through 11th grade) is particularly noteworthy, as it suggests that the 6C competencies are valued consistently by students regardless of their developmental stage or academic level. This finding supports the universal applicability of the framework across secondary education levels.

### ***Item Performance and Scale Refinement***

The rigorous item elimination process, which reduced the initial 160 items to the final 37 items, demonstrates the importance of systematic scale refinement. The removal of 23 items during the exploratory factor analysis phase was based on clear psychometric criteria, including low factor loadings, cross-loading problems, and insufficient communality values. This process resulted in a more parsimonious and psychometrically sound instrument.

The final distribution of items across factors (Citizenship: 6 items; Critical Thinking: 7 items; Creativity: 5 items; Character: 5 items; Communication: 7 items; Collaboration: 7 items) reflects both theoretical considerations and empirical performance. The relatively equal distribution suggests that each competency domain is adequately represented while maintaining measurement efficiency.

### ***Practical Implications for Educational Assessment***

The development of this scale offers significant practical benefits for educational practice and teacher professional development. The instrument provides a structured framework for gathering systematic feedback on teacher implementation of 21st-century skills, which can inform both individual professional growth and institutional improvement initiatives. The student perspective adds unique value, as learners are the ultimate recipients of these pedagogical approaches.

The dual-level scoring approach enabled by the second-order model offers particular utility for educational practitioners. The availability of both factor-specific scores and an overall 6C competency score allows for nuanced diagnostic assessment while also providing a comprehensive evaluation metric. This flexibility supports differentiated professional development approaches, enabling targeted intervention in specific competency areas while maintaining awareness of overall pedagogical effectiveness.

### ***Methodological Strengths and Innovations***

Several methodological strengths distinguish this study from previous scale development efforts. The use of separate samples for exploratory and confirmatory factor analyses provides robust evidence of the factor structure's stability and generalizability. The

systematic five-phase development process, beginning with extensive literature review and qualitative data collection, ensures strong content validity and theoretical grounding.

The focus on student perceptions as evaluators of teacher competency represents an innovative approach in the field. Rather than relying solely on teacher self-assessments or external observations, this study captures the authentic classroom experience from the learner perspective, providing insights into how pedagogical approaches are actually experienced by students.

### *Limitations and Methodological Considerations*

Several limitations should be acknowledged when interpreting these findings. The sample was drawn exclusively from Turkish secondary schools in a specific geographic region, which may limit the generalizability of the results to other cultural and educational contexts. The validity of the instrument across different educational systems and cultural frameworks requires further investigation.

The reliance on student perceptions as the sole source of data represents both a strength and a limitation. While student perspectives provide valuable insights into classroom practices, they may not capture all aspects of teacher competency in implementing 21st-century skills. Students may not always recognize or accurately evaluate certain pedagogical strategies, particularly those that are more subtle or long-term in their effects.

This study developed and validated a reliable scale for measuring students' perceptions of teachers' competence in implementing 6C skills. The scale contributes to the literature by offering a student-based evaluation tool that aligns with major 21st-century frameworks. Its strong psychometric properties suggest that it can support educational improvement, teacher development, and future research on competency-based instruction.

The cross-sectional design of the study also limits conclusions about the stability of the factor structure over time and the instrument's sensitivity to changes in teacher practice. Additionally, the study does not establish relationships between teacher competencies as measured by this scale and actual student learning outcomes in 21st-century skills.

### **FINAL CONSIDERATIONS**

This study successfully developed and validated a comprehensive instrument for evaluating teachers' competence in implementing and promoting 6C skills from the student

perspective. The resulting 37-item scale demonstrates excellent psychometric properties and provides strong empirical support for the theoretical framework of 21st-century learning competencies. The instrument's robust factor structure, high reliability, and demographic invariance establish it as a valuable tool for both research and practice in contemporary education.

### ***Theoretical Contributions***

The study's findings contribute significantly to the theoretical understanding of 21st-century skills by providing the first empirical validation of the 6C framework through student observations of teacher practice. The successful extraction of six distinct factors corresponding to the theoretical dimensions, combined with the validation of a second-order hierarchical model, offers compelling evidence for both the distinctiveness and unity of these competencies in educational practice.

The hierarchical structure revealed by the second-order confirmatory factor analysis suggests that while the six competencies represent distinct pedagogical domains, they also reflect a unified approach to 21st-century education. This finding has important implications for understanding how these skills function in practice and how they should be developed and assessed.

### ***Practical Applications and Educational Impact***

From a practical standpoint, the scale offers multiple applications for educational improvement. It can serve as a diagnostic tool for identifying professional development needs, a formative assessment instrument for monitoring teacher growth, and an evaluation metric for institutional effectiveness in implementing 21st-century pedagogies. The excellent reliability characteristics ensure consistent and dependable measurement across diverse contexts.

The instrument's design as a student evaluation tool is particularly valuable, as it captures the authentic classroom experience from the learner perspective. This approach aligns with contemporary educational philosophies that emphasize student voice and agency in educational evaluation, while offering practical insights into the effectiveness of pedagogical approaches.

The scoring framework provides flexibility for different assessment purposes. Educational leaders can use factor-specific scores to identify areas for targeted professional development, while the overall 6C score provides a comprehensive measure of pedagogical

effectiveness. This dual-level approach supports both diagnostic and summative evaluation purposes.

### **Implications for Teacher Development**

The scale provides a concrete framework for conceptualizing and assessing 21st-century teaching competencies. By making these abstract concepts observable and measurable through student perceptions, the instrument can guide teacher reflection and professional growth. The specific behavioral indicators embedded in the scale items offer clear targets for pedagogical improvement.

The finding that student perceptions remain consistent across demographic groups suggests that effective implementation of 6C skills is universally recognizable, regardless of student characteristics. This supports the development of standardized professional development programs while maintaining confidence that such programs will be relevant across diverse student populations.

### **Future Research Directions**

Several avenues for future research emerge from this study. Cross-cultural validation studies are essential to establish the instrument's applicability across different educational systems and cultural contexts. The framework's universal applicability cannot be assumed without empirical verification in diverse settings.

Longitudinal research examining the temporal stability of teacher competencies and the instrument's sensitivity to professional development interventions would enhance its utility for program evaluation purposes. Such studies could establish whether the scale can effectively measure growth over time and respond to targeted improvement efforts.

Investigation of the relationship between student evaluations using this scale and objective measures of student learning outcomes would provide valuable evidence for the predictive validity of the instrument. Establishing empirical links between teacher implementation of 6C skills and student achievement in both traditional and 21st-century competency domains would strengthen the theoretical foundation of the framework.

## REFERENCES

- 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*

---

- ☐ **Acknowledgments:** This study was conducted as part of a BAP (Scientific Research Project).
  - ☐ **Funding:** This study was by İstanbul Kültür University.
  - ☐ **Conflicts of interest:** There are no conflicts of interest?
  - ☐ **Ethical approval:** Ethical approval was obtained from the Ethical Review Board of İstanbul Kültür University.
  - ☐ **Data and materials availability:** The data and materials used in this study are available upon request from the corresponding author.
  - ☐ **Author contributions:** All authors contributed substantially to this work. They were involved in conceptualization, research design, data collection, analysis, and interpretation. All authors participated in drafting and revising the manuscript and approved the final version for publication.
- 

**Processing and editing: Editora Ibero-Americana de Educação**  
Review, formatting, standardization, and Translation

