**Journal of Education and Educational Research**
ISSN: 2957-9465 | Vol. 10, No. 1, 2024

# Development and Validation of a Classroom Teaching Evaluation Scale in the Era of Digital Transformation

**Rongxia Hao [1, 2, *], Amando C. Yutuc [1]**

[1] Graduate School, Angeles University Foundation, Angels, Philippines
[2] NIC, Shandong Secondary Medical University, Weifang, China

*** Corresponding author:** Rongxia Hao (Email: hao.rongxia@auf.edu.ph)

**Abstract:** With the advent of the digital era, classroom teaching models have undergone profound changes, which challenge the capacity of traditional classroom teaching evaluation systems to meet the demands of contemporary educational practices. This study aims to specifically tailor a classroom teaching evaluation scale to meet the demands of digital transformation. Incorporating scientific methods will improve the accuracy and efficacy of teaching evaluations. We conducted an extensive examination of existing literature, studied policies, and conducted expert interviews to create an initial set of evaluation indicators. Based on this initial groundwork, we used item analysis and exploratory factor analysis to improve the items, and we evaluated the scale's internal consistency using Cronbach's coefficient and split-half reliability. Additionally, we assessed the scale's validity using content validity and confirmatory factor analysis. The results show that the scale is very reliable and valid. It has an overall Cronbach's α coefficient of 0.939, split-half reliability of 0.827, KMO value of 0.934, $\chi^2$=5404.735, P<0.05; $\chi^2$/df=1.206, GFI=0.984, and RMSEA=0.025. The scale consists of six dimensions: teaching philosophy, teaching attitude, teaching content, teaching skills, teaching resources, and teaching outcomes. It includes a total of 30 items and effectively evaluates the quality of classroom teaching in the context of digital transformation in higher education. This scale's development provides a practical assessment tool for higher education, with significant implications for its application.

**Keywords:** Digital Transformation; Classroom Teaching Evaluation; Scale; Reliability and Validity.

## 1. Introduction

Classroom teaching is the main channel of talent development.With the rapid development of digital technology, especially artificial intelligence, university teaching concepts and classroom teaching models are undergoing profound changes.The digital transformation has reshaped the traditional way of teaching in the classroom, and has also brought new challenges to the quality assessment system of traditional classroom teaching [1].Existing classroom teaching evaluation tools have been developed for traditional classrooms and are not in line with the needs of the digital age [2].

In 2020, the Central Committee of the Communist Party of China and the State Council issued the "Deepening the New Era Education Evaluation Reform General Programme" emphasized the core direction of the new era education evaluation, requiring attention to the comprehensive development of students, highlighting individualized teaching, and strengthening the dual evaluation of the teaching process and results.This policy indicates the direction for the digital transformation of the education evaluation system.In 2022, the Ministry of Education launched the "Digitalization of Education Strategic Action" to further promote the comprehensive digital transformation in the field of education and put higher demands on the quality assessment of teaching in the classroom [3].

Classroom teaching quality assessment tools are key tools to guide teaching practices and have a profound impact on improving teaching methods and improving educational effectiveness.A lot of research has been carried out by scholars, but the development of teaching and teaching evaluation tools for classroom teaching in the context of digital transformation remains blank.Modern teaching design requires greater attention to the individual needs of the students, taking into account the multiple dimensions of conditions, processes and outcomes in the evaluation [4]. Therefore, the development of a set of quality assessment tools for teaching in classrooms that are scientifically effective and adapted to the digital transformation has become a matter of urgency.

Based on the developmental education evaluation theory, this study uses quantitative research methods to build a set of classroom evaluation scales that are suitable for the context of digital transformation.The scale fills the gaps in current research, provides new theoretical support and practical references for the evaluation of the quality of university teaching, and provides a scientific basis for the improvement of teaching methods.Through this study, we expect to promote an education evaluation system that better adapts to the challenges of the digital age and raises the level of scientific evaluation of the quality of university teaching.

## 2. Methods

### 2.1. Constructing the Scale Framework and Item Pool

Through extensive review of literature on classroom evaluation, digital transformation, and national policies, combined with expert guidance and based on developmental evaluation theory, the initial scale framework was determined. This framework includes 6 dimensions and 36 items, as shown in Table 1.

**Table 1.** The initial scale framework

| Dimension | Item |
|---|---|
| A.Teaching Belief | A1 Alignment of Teaching Goals with Talent Development Goals |
| | A2 Integration of Student-Centered Philosophy |
| | A3 Emphasis on Critical Thinking and Lifelong Learning Skills |
| | A4 Effective Use of Digital Thinking to Enhance Classroom Teaching |
| | A5 Integration of Moral Education and Effective Incorporation of Ideological and Political Education |
| B.Teaching Attitude | B1 Enthusiasm for Work |
| | B2 Neat and Presentable Appearance |
| | B3 Care for Student Growth and Progress |
| | B4 Timely Responses to Student Questions |
| | B5 Accurate and Thorough Analysis of Student Learning Conditions |
| | B6 Emphasis on Teaching Reflection and Student Feedback |
| C.Teaching Content | C1 Adaptation to Students' Cognitive Levels and Learning Needs |
| | C2 Integration with the Latest Research Results and Practical Cases |
| | C3 Content Richness and Logical Coherence |
| | C4 Appropriate Handling of Key and Difficult Points |
| | C5 Challenging Assignments |
| D.Teaching Skills | D1 Flexible and Diverse Teaching Methods Suitable for Content |
| | D2 Effective Use of Digital Technology for Personalized Instruction |
| | D3 Clear and Organized Language Expression |
| | D4 Reasonable Time Management |
| | D5 Reasonable Teaching Design and Organization |
| | D6 Active and Effective Teacher-Student and Student-Student Interaction |
| | D7 Timely and Accurate Student Evaluation |
| | D8 Strong Classroom Management Skills |
| | D9 Good at Activating Classroom Atmosphere |
| E.Teaching Resources | E1 Rich and Diverse Teaching Resources |
| | E2 Intelligent and Accurate Learning Resource Push |
| | E3 Timely Update of Online Learning Resources |
| | E4 User-Friendly and Functional Teaching Platform Tools |
| | E5 Advanced and Intelligent Teaching Facilities |
| F.Teaching Effectiveness | F1 Improvement in Students' Knowledge Mastery and Application Skills |
| | F2 Enhancement of Students' Thinking and Innovation Abilities |
| | F3 Improvement in Students' Communication, Collaboration, and Teamwork Skills |
| | F4 Enhancement of Students' Self-Learning Abilities |
| | F5 Improvement in Students' Character and Social Responsibility |
| | F6 Overall Improvement for Students with Different Backgrounds |

## 2.2. Participants

A random sampling method was used to choose students from four general medical colleges in Shandong Province, China, as survey participants. Ideally, the sample size should be at least 300; however, any number between five and ten times that number is fine [5]. We determined the sample size for this study by multiplying the number of scale items by 10, which led to a total of 360 participants. Students from the four medical colleges who had completed at least one semester of coursework and expressed their willingness to partake in the survey and offer comments were included. The sample selection is detailed in Table 2.

**Table 2.** Sample selection

| School | Number of Students | Sample Size | Proportion |
|---|---|---|---|
| School F | 22000 | 116 | 32% |
| School S | 14300 | 75 | 21% |
| School B | 15000 | 79 | 22% |
| School J | 17000 | 90 | 25% |
| Total | 68300 | 360 | - |

## 2.3. Data Collection

The survey was edited using the online platform Questionnaire Star, chosen for its ability to display participants' IP addresses without disclosing other personal identifiers. The questionnaire included an introduction, basic student information, and item conformity evaluations. Item conformity was assessed using a Likert five-point scale, ranging from “strongly agree” to “strongly disagree,” with values assigned from 5 to 1. All items were set as mandatory; reasonable time limits were established for each page to ensure the quality of responses; and restrictions were set to ensure each IP address could only submit the survey once. Questionnaires with identical responses were excluded. The questionnaire was forwarded to students by faculty members responsible for student affairs at each university via WeChat.

## 2.4. Item Analysis

### 2.4.1. Critical Ratio Method

The critical ratio method assesses the discriminative ability of scale items to evaluate item differentiation. The total scores of the questionnaire were sorted in descending order. The top 27% of the total sample were categorized as the high-score group, and the bottom 27% as the low-score group. Independent samples t-tests were conducted to compare differences between the two groups. Items with a critical ratio (CR) $< 3.0$ or no significant differences ($P > 0.05$) were deleted [6].

### 2.4.2. Correlation Coefficient Method

The correlation coefficient method evaluates the correlation between item scores and the total scale score, reflecting the homogeneity of the items with the overall scale. Pearson product-moment correlation coefficients (r) were calculated for each item and the total scale score. Items with $r < 0.4$ were removed [7].

## 2.5. Factor Analysis

Exploratory factor analysis was conducted to examine the dimensions and structure of the scale. The suitability of the data for exploratory factor analysis was assessed by calculating the Kaiser-Meyer-Olkin (KMO) value and performing Bartlett's test of sphericity. Data were considered suitable for exploratory factor analysis if the KMO value was greater than 0.6 and Bartlett's test was statistically significant ($P < 0.05$) [8]. Principal component analysis with varimax

rotation was used to screen items. Items with factor loadings less than 0.4 on their principal factors or items with similar factor loadings on two or more principal factors were deleted [6, 9].

## 2.6. Validity Testing

### 2.6.1. Content Validity

Content validity was evaluated using the item-level Content Validity Index (I-CVI) and the scale-level Content Validity Index (S-CVI). Eight experts in educational evaluation were invited to rate the relevance of each item using a Likert 4-point scale with values of 4, 3, 2, and 1 corresponding to "strongly relevant," "quite relevant," "somewhat relevant," and "not relevant," respectively. The I-CVI is calculated as the number of experts who rated an item as 3 or 4 divided by the total number of experts. The S-CVI/Ave is the average of all I-CVIs, and the S-CVI is the proportion of items rated 3 to 4 by all experts. A scale is considered to have good content validity if I-CVI ≥ 0.78, S-CVI ≥ 0.8, and S-CVI/Ave ≥ 0.9 [10].

### 2.6.2. Construct Validity

The construct validity was evaluated by employing confirmatory factor analysis (CFA) to evaluate whether the item relationships reflect the theoretical framework measured by the scale. Generally, a chi-square to degrees of freedom ratio < 3, a Tucker-Lewis index (TLI) ≥ 0.9, a comparative fit index (CFI) ≥ 0.9, and a root mean square error of approximation (RMSEA) < 0.05 indicate good fit validity. For convergent validity, average variance extracted (AVE) ≥ 0.5 and composite reliability (CR) > 0.7 indicate good convergent validity [9]. If the square root of AVE is greater than the absolute value of the correlations between the factor and other factors, and all factors show this conclusion, it indicates good discriminant validity. Standardized loading coefficients > 0.5 indicate strong inter-item correlations [7, 8].

## 2.7. Reliability Testing

The internal consistency of the scale was assessed using Cronbach's alpha coefficient and split-half reliability. Higher Cronbach's alpha and split-half reliability indicate greater internal consistency. A Cronbach's alpha coefficient > 0.7 and split-half reliability > 0.7 indicate good reliability; otherwise, the scale should be revised [11].

## 2.8. Statistical Methods

Data analysis was conducted using SPSS 25.0 and AMOS 22.0 software. Cronbach's alpha coefficient and split-half reliability were used to assess the scale's reliability, while content validity and construct validity were used to assess the scale's validity. A significance level of $P < 0.05$ was considered statistically significant.

# 3. Results

## 3.1. Demographic Information of the Sample

A total of 360 questionnaires were distributed, and all were returned, resulting in a 100% response rate. Of these, 34 were invalid, leaving 326 valid questionnaires, which corresponds to a validity rate of 91%. Among the respondents with valid questionnaires, 148 were male (45.40%) and 178 were female (54.60%). Regarding the academic year, 132 were sophomores (40.49%), 118 were juniors (36.20%), and 76 were seniors (23.31%). The demographic information is summarized in Table 3.

**Table 3.** Demographic information of the sample (n = 326)

| Kinds | | Number | Percentage |
|---|---|---|---|
| School | School F | 103 | 32% |
| | School S | 72 | 22% |
| | School B | 70 | 21% |
| | School J | 81 | 25% |
| Gender | Male | 148 | 45% |
| | Female | 178 | 55% |
| Year | Sophomore | 132 | 41% |
| | Junior | 118 | 36% |
| | Senior | 76 | 23% |

## 3.2. Item Analysis

**Table 4.** Item analysis results of the classroom teaching quality evaluation scale

| Items | Critical Ratio Method | | Correlation Coefficient Method | | Remark |
|---|---|---|---|---|---|
| | t | P | r | P | |
| A1 | 14.217 | .000 | .646** | .000 | |
| A2 | 15.438 | .000 | .648** | .000 | |
| A3 | 16.004 | .000 | .685** | .000 | |
| A4 | 13.348 | .000 | .642** | .000 | |
| A5 | 15.171 | .000 | .669** | .000 | |
| B1 | 13.904 | .000 | .626** | .000 | |
| B2 | 2.873 | .005 | .207** | .000 | Delete |
| B3 | 14.651 | .000 | .658** | .000 | |
| B4 | 12.963 | .000 | .587** | .000 | |
| B5 | 13.836 | .000 | .614** | .000 | |
| B6 | 13.851 | .000 | .642** | .000 | |
| C1 | 12.452 | .000 | .597** | .000 | |
| C2 | 13.181 | .000 | .612** | .000 | |
| C3 | 13.606 | .000 | .594** | .000 | |
| C4 | 14.192 | .000 | .599** | .000 | |
| C5 | 14.184 | .000 | .620** | .000 | |
| D1 | 11.627 | .000 | .555** | .000 | |
| D2 | 11.241 | .000 | .580** | .000 | |
| D3 | 12.888 | .000 | .609** | .000 | |
| D4 | 2.670 | .008 | .190** | .001 | Delete |
| D5 | 12.049 | .000 | .608** | .000 | |
| D6 | 13.053 | .000 | .649** | .000 | |
| D7 | 13.711 | .000 | .635** | .000 | |
| D8 | 12.814 | .000 | .661** | .000 | |
| D9 | 1.085 | .279 | .160** | .004 | Delete |
| E1 | 13.929 | .000 | .601** | .000 | |
| E2 | 13.154 | .000 | .559** | .000 | |
| E3 | 2.502 | .013 | .162** | .003 | Delete |
| E4 | 12.526 | .000 | .557** | .000 | |
| E5 | 1.709 | .089 | .098 | .078 | Delete |
| F1 | 10.585 | .000 | .485** | .000 | |
| F2 | 9.203 | .000 | .461** | .000 | |
| F3 | 9.931 | .000 | .523** | .000 | |
| F4 | 10.298 | .000 | .488** | .000 | |
| F5 | 7.430 | .000 | .407** | .000 | |
| F6 | 2.364 | .019 | .153** | .005 | Delete |

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

**Table 5.** KMO and Bartlett's test

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy | | 0.934 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 5404.735 |
| | df | 496 |
| | Sig. | 0.000 |

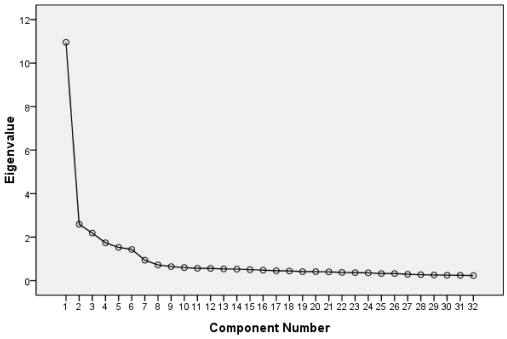


**Fig 1.** Scree plot

**Table 6.** Common factor variance of the EFA

| | Initial | Extraction |
|---|---|---|
| A1 | 1.000 | .706 |
| A2 | 1.000 | .718 |
| A3 | 1.000 | .753 |
| A4 | 1.000 | .662 |
| A5 | 1.000 | .695 |
| B1 | 1.000 | .719 |
| B3 | 1.000 | .653 |
| B4 | 1.000 | .637 |
| B5 | 1.000 | .699 |
| B6 | 1.000 | .696 |
| C1 | 1.000 | .675 |
| C2 | 1.000 | .626 |
| C3 | 1.000 | .695 |
| C4 | 1.000 | .702 |
| C5 | 1.000 | .624 |
| D1 | 1.000 | .616 |
| D2 | 1.000 | .635 |
| D3 | 1.000 | .594 |
| D5 | 1.000 | .667 |
| D6 | 1.000 | .611 |
| D7 | 1.000 | .658 |
| D8 | 1.000 | .624 |
| E1 | 1.000 | .707 |
| E2 | 1.000 | .691 |
| E4 | 1.000 | .727 |
| F1 | 1.000 | .591 |
| F2 | 1.000 | .642 |
| F3 | 1.000 | .623 |
| F4 | 1.000 | .676 |
| F5 | 1.000 | .638 |

Extraction Method: PCA.

Both the critical ratio method and the correlation coefficient method were applied to analyze the 326 valid questionnaires. The results of the item analysis are summarized in Table 4.

From the critical ratio analysis, the high-scoring group consisted of 89 samples, and the low-scoring group also consisted of 89 samples. The results showed that items B2($t<3.0$, $P<0.05$), D4($t<3.0$, $P<0.05$), D9($t<3.0$, $P>0.05$), E3($t<3.0$, $P<0.05$), E5($t<3.0$, $P>0.05$), and F6($t<3.0$, $P<0.05$). From the correlation analysis, items B2($r<0.4$, $P<0.05$), D4($r<0.4$, $P<0.05$), D9($r<0.4$, $P<0.05$), E3($r<0.4$, $P<0.05$), E5($r<0.4$, $P>0.05$), and F6($r<0.4$, $P<0.05$) were analyzed. Comparing the results of both methods, items B2, D4, D9, E3, E5, and F6 met the criteria for deletion and were thus removed.

## 3.3. Factor Analysis

The Bartlett's Test of Sphericity results indicated a KMO value of 0.934 and a p-value of less than 0.05, suggesting that the data is suitable for factor analysis (see Table 5). The scree test showed that the slope leveled off after the sixth factor and the scree plot displayed a clear inflection point (see Fig. 1), demonstrating the effectiveness of the factor analysis. Therefore, exploratory factor analysis was conducted with six factors having eigenvalues greater than 1. The results showed factor loadings ranging from 0.591 to 0.753, all above 0.4, thus retaining all 30 items in the scale. The factor loadings are detailed in Table 6. The variance explained by the six factors was 14.629%, 11.462%, 11.270%, 11.255%, 10.785%, and 7.138%, respectively, with a cumulative variance contribution rate of 66.538% (see Table 7).

## 3.4. Validity Testing

### 3.4.1. Content Validity

The content validity analysis results showed that the I-CVI ranged from 0.875 to 1.000, the S-CVI was 0.9, exceeding the threshold of 0.8, and the S-CVI/Ave was 0.988, above 0.9. These results indicate that the scale has good content validity.

### 3.4.2. Construct Validity

**Table 7.** Total variance explained

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 10.941 | 36.469 | 36.469 | 10.941 | 36.469 | 36.469 | 4.389 | 14.629 | 14.629 |
| 2 | 2.395 | 7.985 | 44.454 | 2.395 | 7.985 | 44.454 | 3.438 | 11.462 | 26.090 |
| 3 | 2.154 | 7.180 | 51.633 | 2.154 | 7.180 | 51.633 | 3.381 | 11.270 | 37.360 |
| 4 | 1.735 | 5.783 | 57.416 | 1.735 | 5.783 | 57.416 | 3.377 | 11.255 | 48.616 |
| 5 | 1.515 | 5.050 | 62.466 | 1.515 | 5.050 | 62.466 | 3.235 | 10.785 | 59.400 |
| 6 | 1.222 | 4.072 | 66.538 | 1.222 | 4.072 | 66.538 | 2.141 | 7.138 | 66.538 |
| 7 | .647 | 2.158 | 68.696 | | | | | | |
| 8 | .607 | 2.022 | 70.719 | | | | | | |
| 9 | .588 | 1.960 | 72.679 | | | | | | |
| 10 | .564 | 1.880 | 74.559 | | | | | | |
| 11 | .550 | 1.833 | 76.393 | | | | | | |
| 12 | .531 | 1.769 | 78.162 | | | | | | |
| 13 | .507 | 1.691 | 79.852 | | | | | | |
| 14 | .503 | 1.675 | 81.528 | | | | | | |
| 15 | .457 | 1.523 | 83.051 | | | | | | |
| 16 | .447 | 1.491 | 84.542 | | | | | | |
| 17 | .429 | 1.431 | 85.974 | | | | | | |
| 18 | .423 | 1.411 | 87.385 | | | | | | |
| 19 | .406 | 1.354 | 88.739 | | | | | | |
| 20 | .387 | 1.289 | 90.028 | | | | | | |
| 21 | .374 | 1.246 | 91.274 | | | | | | |
| 22 | .365 | 1.216 | 92.490 | | | | | | |
| 23 | .335 | 1.117 | 93.608 | | | | | | |
| 24 | .328 | 1.094 | 94.701 | | | | | | |
| 25 | .300 | 1.000 | 95.701 | | | | | | |
| 26 | .276 | .918 | 96.619 | | | | | | |
| 27 | .261 | .871 | 97.490 | | | | | | |
| 28 | .260 | .866 | 98.356 | | | | | | |
| 29 | .252 | .839 | 99.195 | | | | | | |
| 30 | .242 | .805 | 100.000 | | | | | | |

Extraction Method: PCA

The results of the confirmatory factor analysis showed a chi-square-to-degrees-of-freedom ratio of 1.206, meeting the

requirement of being less than 3; a TLI of 0.982, exceeding the requirement of being greater than or equal to 0.9; a CFI of 0.984, meeting the requirement of being greater than or equal to 0.9; and a RMSEA of 0.025, which is less than 0.05. These results, detailed in Table 9, indicate that the model fits well. The path analysis of the scale is shown in Fig. 2. The convergent validity and composite reliability are provided in Table 10.

**Table 8.** Rotated component matrix[a]

| | D | A | B | C | F | E |
|---|---|---|---|---|---|---|
| A1 | | .758 | | | | |
| A2 | | .765 | | | | |
| A3 | | .766 | | | | |
| A4 | | .703 | | | | |
| A5 | | .705 | | | | |
| B1 | | | .766 | | | |
| B3 | | | .676 | | | |
| B4 | | | .719 | | | |
| B5 | | | .749 | | | |
| B6 | | | .738 | | | |
| C1 | | | | .749 | | |
| C2 | | | | .705 | | |
| C3 | | | | .770 | | |
| C4 | | | | .778 | | |
| C5 | | | | .686 | | |
| D1 | .746 | | | | | |
| D2 | .741 | | | | | |
| D3 | .698 | | | | | |
| D5 | .761 | | | | | |
| D6 | .703 | | | | | |
| D7 | .716 | | | | | |
| D8 | .698 | | | | | |
| E1 | | | | | | .718 |
| E2 | | | | | | .733 |
| E4 | | | | | | .777 |
| F1 | | | | | .716 | |
| F2 | | | | | .745 | |
| F3 | | | | | .713 | |
| F4 | | | | | .781 | |
| F5 | | | | | .780 | |

Extraction method: PCA.
Rotation method: Varimax with Kaiser normalization.
a. Rotation converged in 6 iterations.

**Table 9.** Model fit results and criteria for scale path analysis

| | chi-square-to-degrees-of-freedom ratio | TLI | CFI | RMSEA |
|---|---|---|---|---|
| Results | 1.206 | 0.982 | 0.984 | 0.025 |
| Criteria | <3 | ≥0.9 | ≥0.9 | <0.05 |

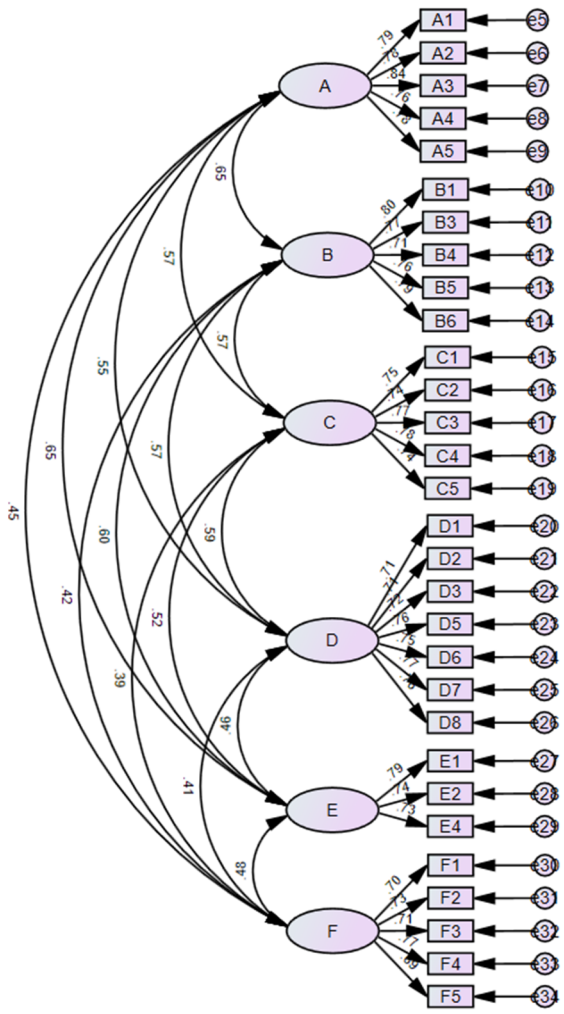


**Fig 2.** Path diagram of the classroom teaching quality evaluation model

Table 10 shows that the item loadings for each dimension are all greater than 0.5, indicating that the explanation of common factors is satisfactory. The AVE for each dimension is greater than 0.5 and the CR for each dimension is greater than 0.7, indicating good convergent validity.

### 3.5. Reliability Testing

The Cronbach's α coefficients and Guttman Split-Half coefficients for the overall scale and its six dimensions are shown in Table 11.

Table 11 shows that both the Cronbach's α coefficients and Guttman Split-Half coefficients for the total scale and for each of the six dimensions are all greater than 0.7. This indicates that the developed scale exhibits good reliability.

## 4. Discussion

### 4.1. Innovation of the Scale

In the context of digital transformation, the new classroom teaching evaluation scale demonstrates unique innovation by incorporating developmental educational evaluation theory. Unlike traditional evaluation scales, this scale emphasizes the multidimensional and dynamic nature of the teaching process, reflecting that educational evaluation is not only an assessment of the current state but also a driver for teaching improvement and student growth. which specifically focus on intelligent teaching materials, emphasise the significant importance of technology in contemporary education. This

technological use goes beyond simple tools, influencing the reorganisation of teaching models and improving teaching results [4].

**Table 10.** Convergent validity and composite reliability for CFA

| Path | | | Estimate | AVE | CR |
|---|---|---|---|---|---|
| A1 | <--- | A | 0.787 | 0.624 | 0.892 |
| A2 | <--- | A | 0.782 | | |
| A3 | <--- | A | 0.838 | | |
| A4 | <--- | A | 0.758 | | |
| A5 | <--- | A | 0.782 | | |
| B1 | <--- | B | 0.802 | 0.589 | 0.877 |
| B2 | <--- | B | 0.770 | | |
| B3 | <--- | B | 0.709 | | |
| B4 | <--- | B | 0.763 | | |
| B5 | <--- | B | 0.790 | | |
| C1 | <--- | C | 0.754 | 0.575 | 0.871 |
| C2 | <--- | C | 0.744 | | |
| C3 | <--- | C | 0.772 | | |
| C4 | <--- | C | 0.780 | | |
| C5 | <--- | C | 0.739 | | |
| D1 | <--- | D | 0.714 | 0.547 | 0.894 |
| D2 | <--- | D | 0.711 | | |
| D3 | <--- | D | 0.715 | | |
| D4 | <--- | D | 0.763 | | |
| D5 | <--- | D | 0.746 | | |
| D6 | <--- | D | 0.768 | | |
| D7 | <--- | D | 0.759 | | |
| E1 | <--- | E | 0.793 | 0.569 | 0.798 |
| E2 | <--- | E | 0.736 | | |
| E3 | <--- | E | 0.733 | | |
| F1 | <--- | F | 0.702 | 0.522 | 0.845 |
| F2 | <--- | F | 0.727 | | |
| F3 | <--- | F | 0.715 | | |
| F4 | <--- | F | 0.770 | | |
| F5 | <--- | F | 0.695 | | |

**Table 11.** Cronbach's α coefficients and split-half coefficients for the scale and its dimensions

| Dimensions | Cronbach's α Coefficients | Guttman Split-Half Coefficient |
|---|---|---|
| A | 0.892 | 0.840 |
| B | 0.877 | 0.847 |
| C | 0.870 | 0.820 |
| D | 0.894 | 0.880 |
| E | 0.798 | 0.711 |
| F | 0.845 | 0.826 |
| Overall Scale | 0.939 | 0.827 |

Developmental educational evaluation theory emphasizes the processual and interactive nature of education [11], which is thoroughly reflected in the new scale. Indicator A2, "Integration of Student-Centered Principles," specifically highlights the central role of students in the teaching process, showing that teaching evaluation should focus on students' individual development and self-regulation during learning. Indicators B4 and D5 underscore the significance of instructing contact and feedback. This interaction takes place not just between teachers and students but also inside students' own self-reflection and regulation, hence fostering holistic student development.

The new scale adds an evaluation dimension, especially the F class.It emphasizes students' access to information, as well as the development of critical thinking, innovation and social responsibility.This concept of comprehensive development is in line with the basic principles of the evaluation theory of developmental education.It emphasizes the achievement of a balanced development of cognitive, emotional and social capabilities and, ultimately, the long-term objectives of education [12].

## 4.2. Continuity of the Scale

The digital transformation has led to significant changes in the teaching methods, but the new system similarities with the previous evaluation system and follows the principles of the evolutionary education evaluation theory. This continuity is reflected in adhering to the core educational values of promoting student comprehensive development and lifelong learning. This continuity is reflected in the adherence to core educational values, namely promoting comprehensive student development and lifelong learning. Multiple dimensions in the new scale, such as teaching philosophy, teaching attitude, and teaching skills, continue to emphasize the important role of teachers in education. Moreover, it extends the teacher's responsibilities to include not just imparting knowledge but also providing guidance and assistance to students throughout their learning journey [7].

The philosophy of developmental educational evaluation supports for evaluation that promotes student growth [13]. This corresponds with the new scale's focus on the timely and innovative nature of instructional content. Teaching content must remain up-to-date and should include practical instances and research findings to foster students' abilities in critical thinking and innovation. The emphasis on the quality and relevance of information ensures that educational evaluation in the digital age preserves the stability of traditional ideals while also being adaptable to address new difficulties [1].

## 4.3. Applicability of the Scale

Within the context of developmental educational evaluation theory, the new scale highlights the importance of digital transformation, specifically in its ability to facilitate ongoing student growth. The theory of developmental educational evaluation suggests that assessment serves as both a measure of teaching efficacy and a vital instrument for continuously enhancing the quality of teaching [14]. Guided by this theory, the new scale provides clear pathways for teaching improvement and promotes continuous adjustment and optimization of teaching strategies through an emphasis on teaching reflection and feedback.

In practical applications, the new scale requires educators to master digital technologies and to be able to effectively integrate them into teaching to facilitate deep learning and

self-development of students. This transformation requires teachers to integrate digital thinking into the teaching design and maintain a sharp ability to reflect during teaching, adjusted to students' feedback [4].

And educational managers need to recognize the guiding role of evaluation indicators. In the context of digital transformation, evaluations should not only focus on the use of technological tools but also consistently aim to enhance the effectiveness of classroom teaching. This ensures that educational evaluation in the new era can support teachers' professional development and students' comprehensive growth [7].

# 5. Conclusion

This study employed a combination of literature review, policy analysis, and in-depth interviews, guided by developmental educational evaluation theory and intimately related to the context of digital transformation. A classroom teaching evaluation scale, tailored for the digital transformation period, has been created by the research team. This scale is based on a thorough analysis of both domestic and international research, and has undergone numerous rounds of talks and changes. The scale's scientific rigour and practical usefulness are ensured through rigorous item screening using project analysis and exploratory factor analysis, as well as reliability and validity testing.

In the current field of study, the classroom teaching evaluation scale specifically for the digital transformation situation is not widely found.[15] The scale developed in this study is a major innovation and supplement to the system of teaching assessment in classrooms in higher education. The results provide a solid theoretical basis and practical tools for teaching appraisal, bringing new perspectives and support for improving the quality of higher education classrooms teaching.

However, this study has certain limitations. Due to objective constraints, the sample is limited to four medical schools in Shandong Province, which affects the representativeness and generalizability of the findings. Future research will focus on gathering sample data from a wider variety of schools and performing comparison analyses in order to create a classroom teaching assessment scale that can be used in other types of higher education institutions. This will further validate and enhance the findings of this study.

# References


[1] Yang Zongkai. Digital development in higher education: Connotation, stages, and implementation paths. China Higher Education, 2023, (5): 16-20.

[2] Chen Y.L. Reconstruction of classroom teaching quality evaluation index system in higher vocational education under educational informatization background. Journal of Huanggang Vocational and Technical College, 2018, 20(4): 41-43.

[3] Cao Peijie, Wang Axi. Empowerment of educational evaluation reform by new generation digital technology. People's Education, 2023, (20): 30-32.

[4] Zhu Xumei, Pan Zhujuan. Research on digital classroom teaching evaluation based on data analysis. China Education Informatization, 2023.

[5] Field A. Discovering statistics using IBM SPSS Statistics. 4th ed. Sage Publications, 2013.

[6] Zhou Jun., Zhang Hongyu, Chen Shen, Chen Wen. Compilation and reliability and validity testing of pre-examination assessment scale for emergency trauma patients. Medical Information, 2024, 37(9): 60-64.

[7] Zheng Donghui., Ye Shengnan. Development and survey of classroom evaluation behavior questionnaire from student perspective. Prospects of Education, 2023, 52(5): 63-80.

[8] Joshua C. Watson. Establishing evidence for internal structure using exploratory factor analysis. Measurement and Evaluation in Counseling and Development, 2017, 50(4): 232-238.

[9] Hu Li-tze, Peter M. Bentler. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling, 1999, 6: 1-55.

[10] Jiang Minhui, Feng Yalin, Liu Min, et al. Compilation and reliability and validity analysis of pregnant women anxiety scale. Chinese Journal of Behavioral Medicine and Brain Science, 2019, 28(3): 275-279.

[11] ZhangWangying, Tsang Kwok Kuen. Assessing the validity and reliability of enabling structure scale in mainland China. International Journal of Educational Management, 2024, 38(5): 1524-1539.

[12] Xu Jinjie, Shen Xin. Reshaping the learner-centered educational evaluation ecosystem—A global observation based on the intelligent development of educational evaluation. Open Education Research, 2023, (03): 40-46.

[13] Yuan Binhua, Zhang Xuan. Research on characteristics of applied undergraduate education quality evaluation system: Based on developmental educational evaluation theory background. Teacher Magazine, 2021, (36): 4-7.

[14] Shen Yushun. Classroom Evaluation. Beijing Normal University Press, 2006.

[15] Ding Yan, Wan Qianyi. A comprehensive review of existing problems in university teaching quality evaluation systems. Theoretical Research and Practice of Innovation and Entrepreneurship, 2021, (8): 3-6.