Construction and Exploration of an Intelligent Evaluation System for Educational APP through Artificial Intelligence Technology

Juan Tang
HuNan Mechanical & Electrical Polytechnic, Changsha, China

Longli Hai (✉)
Children's Hospital Medical Center, Cincinnati, USA
`babyblue5555@sina.com`

Abstract—To improve the evaluation accuracy of educational applications (APPs), the evaluation methods of educational APPs under artificial intelligence (AI) technology are explored. First, based on the principles of establishing evaluation indexes, the evaluation indexes for educational APPs are established. Second, an index evaluation system for educational APPs is constructed, and weights are assigned to the established evaluation indexes of educational APPs with the aid of analytic hierarchy process (AHP). Finally, the availability and effectiveness of the established evaluation system are investigated through empirical analysis. The results show: Five first-level indexes and 20 second-level indexes have been established through the existing index establishment principles, and a framework for intelligent evaluation of educational APPs has been successfully constructed; AHP is utilized to calculate the weight of each index; among the first-level indexes, the weight ratios of the educational and scientific indexes of the educational APPs are larger, whose proportion exceeds 60%; among the second-level indexes, the educational objective, educational principle, and knowledge systematization account for the highest proportion; therefore, the intelligent evaluation system of education APPs is obtained; finally, the empirical analysis has revealed that the score given by the intelligent evaluation system and the actual score of users have a high consistency, which proves that the proposed intelligent evaluation system is feasible and effective. The proposed intelligent evaluation system can be used as the basis for the design of educational APPs to improve the values of educational APPs.

Keywords—Artificial intelligence; educational APP; intelligent evaluation system

1 Introduction

Artificial intelligence (AI) is an influential branch of computer science. The major focus of AI is to simulate the thinking and intelligent behaviors of human beings

through computers. The fields of AI include natural language information processing, expert system construction, image language and other information recognition, and robot research [1]. With the continuous development of information technology, the combination of education and information technology has led to the development and continuous updating of educational applications (APPs). The emergence of educational APPs is also a critical research result under AI. The emergence of educational APPs has brought convenience to learning. At present, there are many classifications of educational APPs, which can be divided into categories of teaching-assistance, management-assistance, and learning-assistance [2]. However, currently, the educational APPs have problems such as uneven quality, and it is difficult for learners to find high-quality and suitable learning software from these APPs. Therefore, the intelligent evaluation of educational APPs has very important significance [3]. Educational APPs are countless, and their quality is uneven. At present, there is neither a unified concept for the quality of educational APPs nor a unified method of judgment.

Different scholars have different definitions of the evaluation of educational APPs. The reason is that various experts and scholars often have subjectivity and bias when giving a concept to the evaluation of educational APPs. Through the research by authoritative scholars and referring to the definition of general evaluation, the evaluation of educational APPs can be defined as follows: According to a given goal, the value of educational APPs is judged through feasible operation techniques and means, which is a basis for the improvement of projects of educational APPs [4]. Currently, there are three types of evaluation methods for educational APPs. The first type focuses on the evaluation of software product functions, which evaluates the educational APPs with the evaluation methods of software products and focuses on the perspectives such as software usability, interface design, and human-computer interaction; the second type focuses on the teaching effect of the educational APPs during the evaluation, which requires a long-term follow-up observation on the objects using the APPs and is easily affected by interference factors, making it not conducive to the experiment; the third type simplifies the evaluation method and evaluates the educational APPs from the perspectives of usability, content, entertainment, and social interaction, which is more in line with the results of scientific evaluation and the expected evaluation standard [5].

Based on the previous research results, first, the evaluation indexes of educational APPs are established according to the principle of evaluation index establishment. Second, the weights are assigned to the established evaluation indexes with the aid of analytic hierarchy process (AHP). Finally, the availability and effectiveness of the established evaluation system are investigated through empirical analysis.

2 Methods

2.1 Educational APPs

With the continuous development of information technology, more APPs that can be used on mobile devices have emerged and developed in large numbers. The mobile software is referred to as APPs, which help mobile devices achieve information carrying and interactive functions; in addition, the behaviors of humans have also changed correspondingly with the appearance of various mobile APPs [6]. The classification and functions of APPs are shown in Figure 1 below.
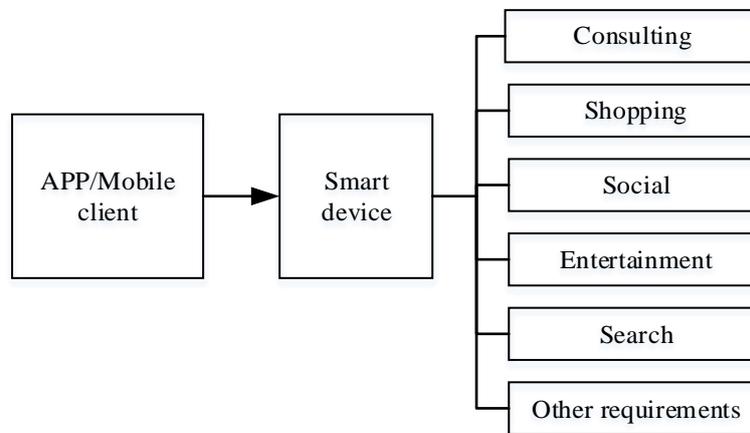
Fig. 1. Functional structure division of APP

Educational APP refers to the application programs of intelligent mobile terminals that provide educational learning resources for various learners, which mainly carries educational information. The emergence of educational APPs makes learning more convenient and flexible, with certain interactivity, personalization, and entertainment, which meets the learning needs of contemporary people to a great extent [7]. Educational APPs have been popularized in all stages of education, including early childhood education, elementary education, higher education, and adult education. According to the functional nature of educational APPs, they can be divided into categories such as early childhood education enlightenment, primary and secondary education assistance, language learning, vocational education and exam counseling, educational administration, and educational games; each category of educational APPs has its specific service objects and contents [8].

2.2 The construction principles of intelligent evaluation indexes for educational APPs

While constructing intelligent evaluation indexes for educational APPs, the principles of scientificity, comprehensiveness, independence, essentiality, operability, and quantitativeness-quantitativeness combination should be considered [9]. Principle of scientificity: The determination of the evaluation indexes of educational APPs should show that the contents of the upper and lower indexes of the index system should be consistent. There is some logic between the indexes. The evaluation indexes and weight coefficients of the index system should be quantified according to the scientific methods and steps. Principle of comprehensiveness: It refers to that the evaluation index of educational APPs can fully indicate the situation of all aspects of educational APPs, evaluate the various attributes of these APPs from multiple perspectives, and assess the educational APPs from different points of view. Principle of independence: It refers to that of all the evaluation indexes, the content of each index is independent and does not overlap or repeat, thereby avoiding the phenomenon of redundancy; to some extent, this principle also helps reduce the complexity of computer algorithms. Principle of essentiality: The established indexes should indicate the most basic aspects of educational APPs, express all aspects of APPs concisely and clearly, and show the status of educational APPs from the perspective of most fundamental attributes. Principle of operability: The established indexes should be operable in the evaluation, and the evaluation method is feasible and complies with the practicality of the evaluation indexes. Principle of quantitativeness-quantitativeness combination: The qualitative indexes are more in line with the ideal evaluation, while the quantitative indexes are more scientific, more accurate, and more reasonable. The combination of the two makes the evaluation results more realistic and comprehensive. While establishing the evaluation indexes for educational APPs, this principle can make the evaluation result conform to the actual situation and have scientific rationality at the same time [10].

2.3 Determination of intelligent evaluation indexes for educational APPs

The previous research results and Delphi method are utilized to determine the factors influencing the quality of educational APPs. By referring to the literature, it is found that most of the current evaluation index system for educational APPs are composed of first-level indexes and second-level indexes. Therefore, the two-level system is also adopted here to construct the evaluation index system for educational APPs [11]. Through searching, it is found that the current evaluation standards for educational APPs are shown in Figure 2 below.
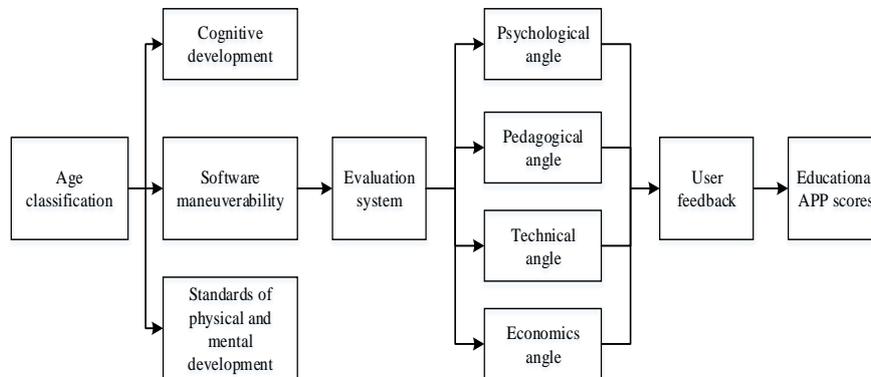
Fig. 2. Evaluation standard for educational APPs

The first-level evaluation indexes of educational APPs are classified into five types: education, functionality, scientificity, artistry, and practicality [12]. The determination of the second-level indexes is based on the content of the five indexes and the principles of index establishment, with the help of the Delphi method, 20 second-level indexes are selected. The design of the first-level indexes is as follows. Education: The application degree and educational value of the educational APPs in the process of teaching and learning of the students are measured according to the theories and principles of instructional design. Its secondary indexes are designed mainly from the perspectives of educational subject and correlation. The subject refers to the core elements included in the teaching design and teaching process, and the correlation refers to the elements correlated to the teaching design or teaching process.

Functionality: Functionality refers to the stability, ease of use, and usability of the APP system. Whether the functional services of the APPs can meet the needs of users in learning is evaluated. Its secondary indexes are determined from the perspectives of inherent attributes and extended attributes of the APPs. The inherent attributes refer to the technical requirements and indexes of APPs during development and design; the extended attributes refer to the extended functions and services while utilizing the APPs.

Scientificity: The functionality of education APPs is elaborated through the content information contained in the APPs. The functionality of educational APPs is mainly evaluated from the perspective of teaching content organization and utilization. Its second-level indexes are designed mainly from the internal and external perspectives. The internal perspective refers to the inherent characteristics of the content, while the external perspective refers to features and characteristics associated with the content of the subject and the learning environment.

Artistry: The artistry of APPs analyses the feelings that APP brings to users during utilization, which mainly includes the interface design of the APPs, and whether the visual and auditory effects can bring users a comfortable and pleasant experience. Therefore, its second-level indexes are designed from the perspectives of artistry and

sense of experience. Artistry refers to the aesthetics and comfort of APP design. Experience mainly refers to the experience of using the APPs.

Practicality: Practicality refers to the benefits brought to users. These benefits refer to the degree of knowledge acquisition, followed by the convenience of the APPs to users. Its secondary indexes include whether the APPs are easy to operate, and whether the relevant advertisements and paid contents meet the needs and consumption of users.

2.4 AHP

After obtaining the evaluation index system of educational APPs, the weights of these indexes need to be calculated. The weight is the quantification of the importance of each index. It converts the importance of each evaluation index into a value and elaborates the specific contribution or importance of each index to the entire evaluation index system through a numerical form [13]. Only after the weight of each index is determined can the evaluation object be evaluated correctly and objectively. Therefore, the weight of the evaluation index needs to be calculated and determined by a scientific and reasonable method, and each index is given a weight value of corresponding importance. Here, the AHP method is utilized to determine the weight of the index; the weight values of the first-level indexes and the second-level indexes in the evaluation index system are obtained through the AHP software [14]. The architecture of index AHP method is shown in Figure 3 below.
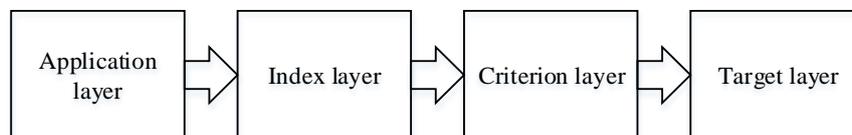
Fig. 3. The architecture of index AHP method

The target layer in the index AHP architecture is the target of this AHP; the criterion layer is the process of implementing the intermediate links of evaluation according to various measures; the index layer is utilized to select the research means, techniques, and methods to be applied; the application layer is the bottom layer of the AHP, which mainly contains various methods for evaluation and research using this index system [15].

The weight of each index in the index system is determined by the AHP method. The specific process includes the construction of AHP model, the construction of the judgment matrix, the single hierarchical arrangement and consistency test, and the total hierarchical arrangement and consistency test. However, in the actual application process, the final total hierarchical arrangement and consistency test step can be omitted [16].

Construction of AHP model: The evaluation of educational APPs is utilized as the target layer of the model, the first-level indexes are utilized as the criterion layer of the model, and the second-level indexes are utilized as the sub-criterion layer of the

model. The constructed evaluation index system for educational APPs is shown in Figure 4 below.
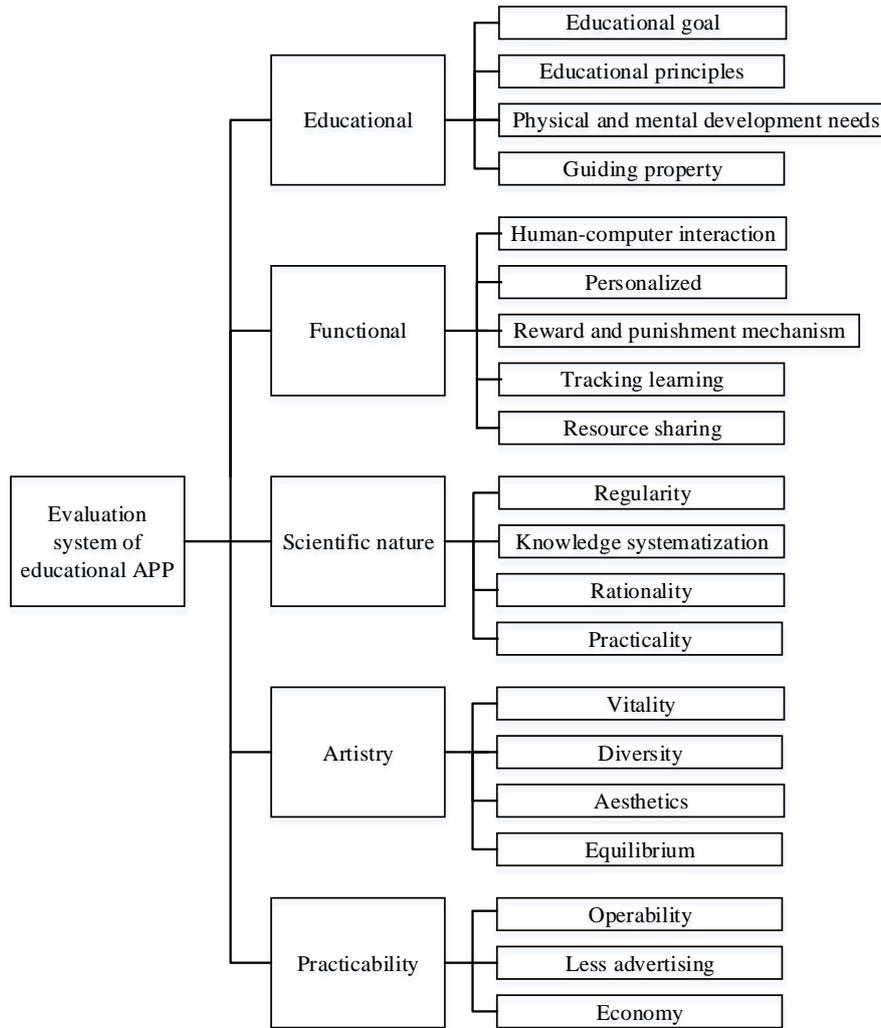
Fig. 4. The architecture of evaluation index system for educational APPs

Construction of the judgment matrix: The average value of the evaluation results of all levels of indexes is calculated by a mathematical calculation method to obtain the final result of index weight [17]. The first-level indexes, education, scientificity, functionality, artistry, and practicality, are represented by A1, A2, A3, A4, and A5, respectively. The constructed judgment matrix model for the first-level indexes is shown in Table 1 below.

Table 1. Judgment matrix of first-level indexes

T	Education (A1)	Scientificity (A2)	Functionality (A3)	Artistry (A4)	Practicality (A5)
Education (A1)	1	a12	a13	a14	a15
Scientificity (A2)	a21	1	a23	a24	a25
Functionality (A3)	a31	a32	1	a34	a35
Functionality (A4)	a41	a42	a43	1	a45
Practicality (A5)	a51	a52	a53	a12	1

Since the decision-makers have certain differences in the weighting ratio of each index, here, the 9-scale method is utilized to determine each calculation result [18].

This judgment matrix A =(aij)n×n is expressed as follows:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = (a_{ij})_{n \times n} \tag{1}$$

Where: i and j represent evaluation indexes, respectively, n is the number of indexes, that is, the order of the judgment matrix. Besides, aij >1, aji =1/aij, aii =1. The above judgment matrix has symmetry.

Single hierarchical arrangement: The weight of each first-level index and the largest characteristic root of the judgment matrix are calculated [19]. First, each column of the judgment matrix A needs to be normalized. The calculation method is as follows.

$$\overline{a_{ij}} = \frac{a_{ij}}{\sum_{j=1}^{n} a_{ij}}, (i, j = 1, 2, \ldots, n) \tag{2}$$

Then, it is necessary to sum the judgment matrix W by row. The calculation method is as follows.

$$\overline{w_t} = \sum_{j=1}^{n} \overline{a_{ij}}, (i = 1, 2, \ldots, n) \tag{3}$$

Next, the vector of the judgment matrix is transposed as follows.

$$\overline{w_t} = \left[\overline{w_1}, \overline{w_2}, \cdots \overline{w_n} \right]^T \tag{4}$$

It is also necessary to convert the transposed vector to obtain the feature vector. The conversion method is as follows.

$$W = \frac{\overline{W_i}}{\sum_{j=1}^{n} \overline{W_j}}, (i = 1, 2, \cdots, n)$$

(5)

The feature vector w is also the weight vector of each indication, and the weight value of each indication can be obtained.

Finally, the maximum characteristic root λmax of the judgment matrix needs to be calculated, as shown below.

$$\lambda_{max} = \frac{1}{n} \sum_{i=1}^{n} \frac{(AW)_i}{W_i}$$

(6)

Where: $(AW)_i$ represents the i-th component of the vector W in the matrix A. If the judgment matrix has complete consistency, the maximum characteristic root will be equal to n.

Consistency test: During single hierarchical arrangement, the principle of data consistency should be followed. Therefore, it is necessary to perform a consistency test on the indexes. The calculation method of the consistency index (CI) of the judgment matrix A is shown below.

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

(7)

If the value of CI is 0, the judgment matrix will have complete consistency; the larger the value of CI is, the worse the consistency of the judgment matrix will be. Whether the matrix conforms to the consistency test result can be determined by judging the ratio of the consistency of the matrix and the random consistency index (RI). The calculation method of the consistency ratio (CR) of the judgment matrix is as follows.

$$CR = \frac{CI}{RI}$$

(8)

Here, the order of the judgment matrix is 5; then, the value of RI can be obtained through the average random consistency index table, which is 1.12 [20].

3   Results and Analysis

3.1   Index weight determination results

The AHP method is utilized to obtain the weights of the first-level indexes and the second-level indexes. The calculation results of the weights of evaluation indexes for educational APPs are shown in Table 2, Table 3, and Figure 5 below.

Table 2. Calculation results of the weights of first-level indexes for educational APPs

T	Education (A1)	Scientificity (A2)	Functionality (A3)	Artistry (A4)	Practicality (A5)
Education (A1)	0.41	0.44	0.38	0.42	0.31
Scientificity (A2)	0.20	0.22	0.19	0.32	0.31
Functionality (A3)	0.21	0.22	0.19	0.11	0.19
Artistry (A4)	0.10	0.08	0.18	0.11	0.13
Practicality (A5)	0.08	0.04	0.06	0.04	0.06
W	0.392	0.248	0.181	0.119	0.060

Table 3. Calculation results of the weights of second-level indexes for educational APPs

Education		Functionality		Scientificity		Artistry		Practicality
Human-computer interaction	0.03	Regularity	0.03	Vividness	0.02	Operability		0.02
Individuation	0.05	Knowledge systematization	0.11	Diversification	0.03	Less advertisements		0.03
Incentive mechanisms	0.01	Rationality	0.07	Aesthetics	0.02	Economy		0.01
Follow-up learning	0.06	Actuality	0.03	Balance	0.02			
Resource sharing	0.02							
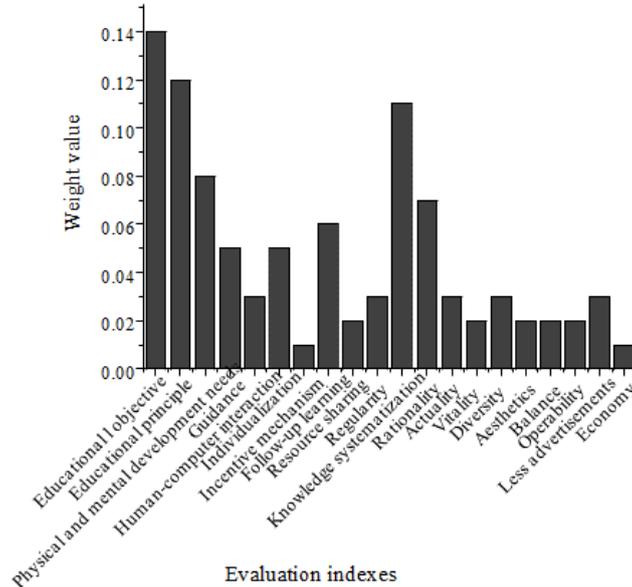
Fig. 5. Analysis of the weights of second-level indexes for educational APPs

As shown in Table 2, Table 3, and Figure 5, among the first-level indexes, the weight ratios of educational and scientific indexes for educational APPs are larger, whose proportion exceeds 60%. Therefore, while designing educational APPs, it is necessary to value the educational and scientific nature of the educational APPs. Among the second-level indexes, the educational objective, education principle, and knowledge systematization account for the highest proportion; therefore, APP designers can put more emphasis on the educational objective, educational principle, and knowledge systematization of educational APPs to improve the value and selection probability of educational APPs.

3.2 Performance analysis of intelligent evaluation system

To verify the application effectiveness and evaluation reliability of the proposed intelligent evaluation system, several common educational APPs are selected to compare the evaluation results of the proposed intelligent evaluation system and the actual evaluation of the users (questionnaire survey). The selected educational APPs are represented by 1, 2, 3, 4, and 5, respectively, and the scores obtained are shown in Table 4 below.

Table 4. Statistical evaluation of educational APPs by the intelligent evaluation system

Educational APPs	1	2	3	4	5
Scores	3.63	3.44	3.08	3.42	2.93

The scores and feedback from 20 users are obtained through questionnaire survey. The scoring grades are divided into four categories: good, fair, qualified, and poor, and their corresponding scores are 4 points, 3 points, 2 points, and 1 point, respectively. The scores obtained are shown in Table 5 below.

Table 5. Evaluation of educational APPs by users

Educational APPs	1	2	3	4	5
Scores	3.61	3.47	3.12	3.39	3.03

The comparative analysis of the scores of educational APPs by the two methods is shown in Figure 6 below.
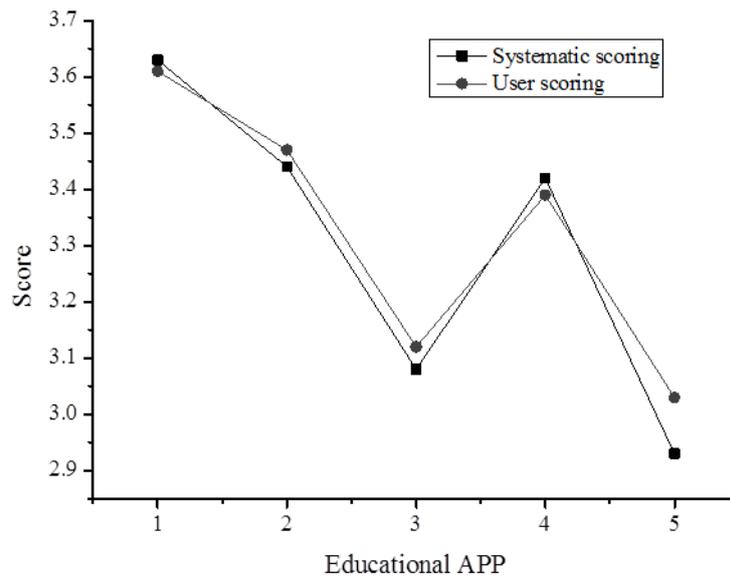
Fig. 6. Comparison of the scores of educational
APPs scored by the two methods

As shown in Tables 4 and 5 above, the ranking order of the included educational APPs by the two evaluation methods is the same, which is 1, 2, 4, 3, 5; therefore, it proves that the two evaluation methods are consistent, the results obtained by the intelligent evaluation system are true and reliable, and the adjustments of the index system is unnecessary. Figure 6 also shows that the scores of the two evaluation methods are close, indicating that the proposed intelligent evaluation system has high reliability and practicality.

4 Conclusion

The evaluation methods of educational APPs under AI technology are explored. First, based on the principles of evaluation indexes, the evaluation indexes for educational APPs are established. Five first-level indexes and 20 second-level indexes are established according to the existing index establishment principles, and a framework for intelligent evaluation of educational APPs has been successfully constructed. Second, the weights are assigned to the established evaluation indexes of educational APPs with the aid of AHP method. The AHP software is utilized to calculate the weights of first-level and second-level indexes. The calculation results show that among the first-level indexes, the weight ratios of the educational and scientific indexes of the educational APPs are larger, whose proportion exceeds 60%; among the second-level indexes, the educational objective, educational principle, and knowledge systematization account for the highest proportion. Finally, the availability and effectiveness of the established evaluation system are investigated through empirical analysis. The empirical analysis has found that the ranking and evaluation results by the users and the proposed intelligent evaluation system are consistent; besides, the scores obtained have a high consistency, indicating that the proposed intelligent evaluation system is feasible and effective. During the actual development process of educational APPs, the proposed intelligent evaluation system can be used as the basis for the design of educational APPs to improve the values and selection probability of educational APPs.

Due to some objective limitations, while calculating the weight of each index through the AHP method, limited by the personnel conditions, most of the opinions come from the previous research results, which lacks the actual scoring by relevant professionals. Also, during the empirical analysis and comparison, because of the time limitations, the experience and scoring of only 20 users are obtained, which lacks the support of massive data. It is hoped that in the follow-up works, the above deficiencies can be improved, and the professionalism can be enhanced; at the same time, more data can be obtained to support the results. In addition, with the continuous development of science and technology, it is hoped that more intelligent means can be used for the evaluation of such APPs to improve the scientificity while reducing the manual works.

5 References

[1] Lu, H. Li, Y. Chen, M. Kim, H. Serikawa, S. (2018). Brain intelligence: go beyond artificial intelligence. Mobile networks & applications, 23(2): 368-375. https://doi.org/10.1007/s11036-017-0932-8

[2] Papadakis, S. Kalogiannakis, M. Zaranis, N. (2017). Designing and creating an educational app rubric for preschool teachers. Education and information technologies, 22(6): 3147-3165. https://doi.org/10.1007/s10639-017-9579-0

[3] Oliveira, M. L. D. Galembeck, E. (2015). Using a web analysis tool to evaluate an educational app usage. International Journal of Mobile Learning & Organisation, 9(2): 146. https://doi.org/10.1504/ijmlo.2015.070705

[4] Shahidi, N. Tossan, V. Cacho-Elizondo, S. (2018). Assessment of a mobile educational coaching app. International Journal of Technology & Human Interaction, 14(1): 22-43. https://doi.org/10.4018/ijthi.2018010102

[5] Papadakis, S. J. Michail, K. Vassileios, O. Nikolaos, Z. (2017). The appropriateness of scratch and app inventor as educational environments for teaching introductory programming in primary and secondary education. International Journal of Web-Based Learning and Teaching Technologies, 12(4): 58-77. https://doi.org/10.4018/ijwltt.2017100106

[6] Song, D. Oh, E. Y. (2016). A participatory design approach for a mobile app-based personal response system. Journal of Educational Technology Systems, 44(3): 346-361. https://doi.org/10.1177/0047239515618465

[7] Cherner, T. Lee, C. (2016). A detailed rubric for assessing the quality of teacher resource apps. Journal of Information Technology Education: Innovations in Practice, 15(1): 117-143. https://doi.org/10.28945/3527

[8] Tishman, S. Clapp, E. P. (2017). Building students' sense. Educational Leadership Journal of the Department of Supervision & Curriculum Development N.e.a, 75(2): 58-62.

[9] Xinkai, D. Yuan, T. (2015). The definition of strong province in ip and the principles and methods of construction of evaluation index. Science & Technology Progress and Policy, 398(4): 765-770.

[10] Chen, S. (2016). Construction of evaluation index system of innovation and entrepreneurship in local colleges and universities. journal of biological chemistry, 275(27): 20748-53.

[11] Weng, P. L. (2015). Developing an app evaluation rubric for practitioners in special education. Journal of Special Education Technology, 30(1): 43-58. https://doi.org/10.1177/016264341503000104

[12] Mahmoud, M. H. Abo El-Hamayed, S. H. (2016). An intelligent tutoring system for teaching the grammar of the arabic language. Journal of Electrical Systems and Information Technology, 3(2): 282-294. https://doi.org/10.1016/j.jesit.2016.04.001

[13] Shaw, C. M. Tan S A. (2015). Integration of Mobile Technology in Educational Materials Improves Participation: Creation of a Novel Smartphone Application for Resident Education[J]. Journal of Surgical Education, 72(4):670-673. https://doi.org/10.1016/j.jsurg.2015.01.015

[14] Al Qubaisi, A. Badri, M. Mohaidat, J. Al Dhaheri, H. Yang, G. Al Rashedi, A. et al. (2016). An analytic hierarchy process for school quality and inspection. International Journal of Educational Management, 30(3): 437-459. https://doi.org/10.1108/ijem-09-2014-0123

[15] Ooi, J. Promentilla, M. A. B. Tan, R. R. Ng, D. K. S. Chemmangattuvalappil, N. G. (2018). Integration of fuzzy analytic hierarchy process into multi-objective computer aided molecular design. Computers & Chemical Engineering, 109(JAN.4): 191-202. https://doi.org/10.1016/j.compchemeng.2017.11.015

[16] Siew, L. W. Singh, R. S. A. B. Hoe, L. W. (2017). An empirical study on the mold machine-tool selection in semiconductor industry with analytic hierarchy process model. Journal of Computational and Theoretical Nanoence, 23(9): 8286-8289. https://doi.org/10.1166/asl.2017.9878

[17] Yusoff, A. M. Salam, S. Mohamad, S. N. M. Daud, R. (2017). Gamification element through massive open online courses in tvet: an analysis using analytic hierarchy process. Journal of Computational & Theoretical Nanoscience, 23(9): 8713-8717. https://doi.org/10.1166/asl.2017.9956

[18] Ding, X. Luo, Y. Yi, Y. Jie, B. Wang, T. Bian, W. (2016). Orthogonal design for scale invariant feature transform optimization. Journal of electronic imaging, 25(5): 053030.1-053030.12. https://doi.org/10.1117/1.jei.25.5.053030

[19] Cui, C. Q. Wang, B. Zhao, Y. X. Wang, Q. Sun, Z. M. (2019). China's regional sustainability assessment on mineral resources: results from an improved analytic hierarchy process-based normal cloud model. Journal of Cleaner Production, 210(FEB.10): 105-120. https://doi.org/10.1016/j.jclepro.2018.10.324

[20] Zhang, D. Shen, J. Liu, P. Zhang, Q. Sun, F. (2020). Use of fuzzy analytic hierarchy process and environmental gini coefficient for allocation of regional flood drainage rights. International Journal of Environmental Research and Public Health, 17(6): 2063. https://doi.org/10.3390/ijerph17062063

6 Authors

Juan Tang, Master of Economics, Associate Professor. Graduated from the HuNan University of Technology and Business in 2004.Worked in HuNan Mechanical & Electrical Polytechnic. Her research interests include economic development and enterprise management. 405055022@qq.com

Longli Hai Li received Ph.D. degree in Computer Science from University of Cincinnati, Cincinnati, United States. Dr. Hai is a Research Associate in the Division of Neonatology, Cincinnati Children's Hospital Medical Center.