Investigation of an Outcomes-Driven Assessment model in Programming Course

Ying Li[1] ShiJie Gui[1] JinCheng Qiu[1] Tianqi Liu[2]

ltq399@buaa.edu.cn

(Corresponding Author)

[1]School of Computer Science and Engineering, Beihang University, Beijing, China

[2]School of BeiJing, Beihang University

***Abstract*—Course assessment is an important quantitative evaluation for students' learning effect in certifying engineering educations. In this paper, the achievement degree based on learning outcomes is used as the basic indicator to construct a program design course evaluation model. According to the characteristics of the program design course, as well as applying the basic theories for course assessment and previous practical experiences, this study mainly discusses the scope and content of course assessment, assessment principles, assessment index system, evaluation standards and process. In the process of constructing the evaluation index system, more than 40 evaluation indicators were designed from the perspectives of both the teachers and students. The system also takes into account the weightage of evaluation indicators at all levels, determined by analytic hierarchy process, and consistency test is carried out to ensure the design of the evaluation index system and evaluation standards are strictly scientifically determined. In experiment, the controlled program design foundation course provided actual test data, and satisfactory results were obtained. This leads to conclusion that the research and practice conducted can provide reference for curriculum teaching reforms.**

Keywords—Learning Outcomes Assessment, Bloom's Taxonomy, Analytic Hierarchy Process, Fuzzy Evaluation Algorithm

I. INTRODUCTION

The Outcomes-Driven Assessment Model (ODAM) borrows the concept of learning outcomes from the Engineering Education Professional Accreditation, and follows the strategies of "backwards design" and "forward implementation" [1]. This way, the result (i.e., requirement) is both the starting and end point, thus ensuring the consistency of training objectives and teaching effectiveness to the greatest extent. In ODAM, backwards design is the core of this model. It uses students' learning output as the driving force for iterative improvement of the teaching evaluation system. Starting from the expected learning outcomes of students, the training objectives and assessment index are identified by backward design. Obviously, ODAM proposed in this paper is different from the traditional course-oriented and content-driven positive design strategy [2]. It is more suitable for solving the teaching evaluation problem with the characteristics like being "non-deterministic" and "changing dynamically". ODAM fully embodies the student-centered evaluation concept. It can be regarded as a kind of educational evaluation paradigm that optimizes the traditional teaching evaluation to student-centered evaluation rather than teacher-centered evaluation.

Unfortunately, that process of evaluating teaching quality requires tedious workload and statistics. Therefore, it is a difficult but worthwhile problem to conduct scientific and accurate evaluation of teaching quality. At present, there exist multiple problems in the evaluation model for higher education courses:

Challenge 1, there is a lack of a scientific and reasonable evaluation methods: the evaluation standards are inconsistent, and there is no unified standard guidance; the evaluation body consists of only a single party, which is usually carried out by teachers.

Challenge 2, there is a lack of a set of comprehensive and complete evaluation indicators: the current evaluation for programming courses is mainly based on performance evaluation, which fails to take into account the evaluation of process. The result is that more emphasis is placed on the evaluation of students' theoretical knowledge and programming ability, while ignoring students' thinking ability and academic literacy.

Challenge 3, there is a lack of efficient and accurate evaluation algorithms: An evaluation model can be thought as a multi-objective and multi-condition-optimization problem, which itself has high complexity and is difficult to perform quantitative analysis on. Current evaluation scoring criteria are mainly based on experts' subjective experience.

Based on the above challenges, this research focuses on designing an evaluation model driven by OBE (Outcome based Education) to tackle these three challenges. The specific research contents are as follows:

Our approach to challenge 1 i.e. improving the evaluation method: the evaluation content is reversely designed based on the degree of achievement of learning outcomes: This design method follows the usual way of teaching and realizes the unification of curriculum settings, student requirements and evaluation standards into one single entity.

Our approach to challenge 2 i.e. perfecting the evaluation indicator: an evaluation indicator system driven by the achievement of learning outcomes is designed. Under the guidance of OBE, we can decompose each graduation requirement of the subject into smaller entities and can then better arrange the teaching contents and evaluate the implementation effect of specific indicator point for graduation requirements.

Our approach to challenge 3 i.e. optimizing the evaluation algorithm: a fuzzy comprehensive evaluation algorithm is proposed to obtain the optimal degree of learning achievement. When designing the algorithm, through a combination of quantitative and qualitative methods, and according to the different importance in evaluation indicators, the goal can be decomposed into multiple goals or criteria. The implicit relationship between the learning outcomes, evaluation indicators and teaching quality can be further investigated to obtain a relatively comprehensive and objective evaluation indicator weightage.

Based on the OBE educational concept and the characteristics of the program design course, this paper focuses students as the primary subject of study, takes the results as reference, decomposes course objectives, and proposes a program design course evaluation model based on the learning outcome assessment. From the perspectives of both the teachers and students, this model effectively judges the degree of achievement of the training objectives of the course by decomposing the course objectives, refining the course assessment, and calculating the achievement of the course objectives.

II. EVALUATION OF TEACHERS' TEACHING QUALITY

The evaluation of teachers' teaching quality is a process of describing and judging a teacher's teaching activities. The evaluation takes a teacher's behavior as the item of study, and judges, analyzes and compares accordingly to the perspective of behavior in their teaching activities [3]. The main characteristics of teachers' evaluation are as follows: Firstly, the evaluation is a purposeful, planned and continuous process; secondly, the goal of the evaluation is to produce valuable judgments, make decisions based on the judgments, and pay attention to the explanation and analysis of teaching activities in the evaluation process; thirdly, the evaluator and the teacher-in-evaluation are unified in the evaluation process. In many cases, the teacher-in-evaluations are closely related to the evaluators; sometimes, they are even the evaluators themselves [4]. In order to improve the effectiveness and accuracy of teachers' evaluation, a diversified teacher evaluation method is adopted.

A. Student Evaluation

As the main body of the system, students will mainly evaluate the teachings in the classroom through a teaching quality evaluation system, as shown in Table 1. Also, students interact with a teacher's teachings the most, and we can confidently assume the students' evaluation will uphold congenital advantage in the overall evaluation compared to other factors [5]. Moreover, the stability and reliability of student evaluations have been intensively verified in many studies and are recognized by teachers. Furthermore, the student base is large, which can satisfy many statistical calculations. Students in colleges and universities are comparatively more mature and thus can differentiate between right and wrong. Therefore, they are most competent for conducting teaching quality evaluations.

Table 1 Student Evaluation Indicators

Primary Indicator	Secondary Indicator	Description
Student Evaluation	Teaching Attitude	Proficiency in lectures.
		Lectures are serious and responsible.
		Actively respond to students' questions after class.
		Review students' assignments carefully.
		Discipline attitude.
	Teaching Content	Highlight important points, appropriate difficulty.
		Fun and interactive.
		Integrating theory with practice.
		Easy to understand.
	Teaching Method	Focus on student-teacher interactions.
		Effective use of learning tools.
		Flexible and diverse teaching methods, such as online and offline integrated teaching.
		Teach students according to their aptitude, and pay attention to heuristic teaching.
	Teaching Effect	The degree to which students acquire knowledge.
		Students' ability to write programs to solve practical problems.
		Students' information literacy.
	Teaching Quality	Teaching Responsibility.
		Concise and vivid language.
		Strong professional knowledge.

B. Expert Evaluation

Experts in related fields generally have a higher degree of knowledge and a broader perspective on their field. They tend to evaluate teachers' teaching reasonably better based on their years of teaching experience [6], as shown in Table 2. Experts mainly conduct their teaching evaluation by attending lectures without prior notice. This sort of method for evaluation better reflects a teachers' teaching level, and the evaluation result is far more realistic and precise.

C. Peer Evaluation

Peer teacher hearings ensure multi-faceted participation in educational activities and evaluations. Peer evaluation only takes into consideration the evaluation of teaching quality in class, that includes the full-cycle evaluation of before, during and after class, as shown in Table 3. This way, it makes teacher evaluation more objective and complete.

Table 2 Expert Evaluation Indicators

Primary Indicator	Secondary Indicator	Description
Expert evaluation	Teaching Attitude	Showcase as a role model for others, organize lectures in an orderly manner, and abide by teaching disciplines.
		Detailed teaching plans and lecture notes, conduct teaching based according to the requirements of the syllabus.
	Teaching Content	Able to teach students in accordance with their aptitude, concise and accurate, highlight important points, clear thinking, inspiring. Lecture hours are reasonable. Students have room to recap and revise and be able to better focus on the cultivation of abilities and creative thinking.
	Teaching Method	Able to effectively use multimedia and online teaching resources, appropriate use of advanced teaching techniques.
		Able to flexibly adjust teaching strategies according to reflected teaching effects.
	Teaching Effect	Lively classroom atmosphere, good communication between teachers and students.
		Lectures are enthusiastic, full of energy, influential, and can attract students' attention.
		Students demonstrate sustained curiosity and desire to learn more knowledge.
	Teaching Quality	Have a certain academic level and ability to keep up with new achievements and information of the field related, and integrate theory with practice.
		Possess good basic teaching skills, strong teaching adaptability, as well as demonstrates the ability and skills to solve problems and deal with sudden changes.

Table 3 Pear Evaluation Indicators

Primary Indicator	Secondary Indicator	Description
Peer evaluation	Before Class	Participate in various teaching activities, such as group lecture preparation.
		Carefully prepare various teaching materials, such as writing lecture notes, lesson plans, etc.
	During Class	High-quality teaching skills.
		Lecture ideas are clearly presented, concepts are accurate, content is substantial and advanced, and theory is linked to practice.
		Use a variety of teaching methods to inspire students to participate in thinking.
		Teaching method is complete, used reasonably, and produces excellent results.
	After Class	Actively engage in discussing teaching experiences with peers.
		Summarizes and conclude problems faced during lectures.
		Participate in various teaching reforms.
		Research advance teaching methods.

III. EVALUATION OF STUDENTS' LEARNING ABILITY

The evaluation system of college students' learning ability is crucial to determining the quality of a college course. Learning ability refers to a comprehensive ability for students to acquire knowledge through external elements such as knowledge, ability, quality, and others, stemming from learning abilities such as learning efficiency, attitude, self-control, and other internal elements. It mainly includes three primary indicators: Knowledge, Ability and Quality, also known as the KAQ evaluation system. Among them, knowledge is the basis for forming ability and quality; skill is the reflection of the internalization of knowledge and quality; the improvement of quality will promote a faster grasp and expansion of knowledge and a better showcase and development of skills.

A. Evaluation of Knowledge Achievement Degree

Knowledge is the surface layer of the KAQ evaluation system. It acts as the carrier for ability and quality. Modern information processing psychology believes that knowledge can be divided into declarative, procedural, and strategic knowledge. Among them, in information processing terms, declarative knowledge mainly refers to describing the concepts and grammar when designing a program, procedural knowledge mainly refers to the process and method of writing programming codes, and strategic knowledge mainly refers the use of existing knowledge to solve new problems. ZAHID ULLAH and Adidah Lajis et al. tried to verify students' learning progress from educational goals [7]. From that, they proposed a method that applies Bloom's taxonomy to evaluate the students' programming (Java language) level.

Therefore, we use Bloom's cognitive ability classification as the basis for evaluating the mastery of programming knowledge. According to Bloom's classification theory for teaching objectives, knowledge acquisition can be divided into six dimensions: cognition, understanding, application, analysis, synthesis and evaluation. Among them, cognition and understanding are declarative knowledge; application and analysis are procedural knowledge; synthesis and evaluation are strategic knowledge, as shown in Table 4.

Table 4 Knowledge Achievement Index

Primary Indicator		Secondary Indicator	Description
Know-ledge	Declarative knowledge	Cognition	(1) Requirements: understand the contents of a subject, as it is the most basic requirement for learning. (2) Contents: Basic programming concepts (such as variables, constants, operators, expressions, statements, functions, arrays, pointers, structures, files, etc.).
		Understand-ing	(1) Requirements: understand the laws and connections of knowledge points and understand the fundamental principles and reasons behind them. (2) Contents: grammar and design methods of programming languages (such as data types, control structures, function calls, sorting, searching, file manipulation, modular programming, etc.).
	Procedural knowledge	Application	(1) Requirements: know how to use knowledge learned to solve problems based on memory and understanding. (2) Contents: use computers to describe things in the real world, represent relationships between things, solve practical problems.
		Analysis	(1) Requirements: be able to choose appropriate solutions to solve a variety of problems. (2) Contents: write high-performance programs (choose appropriate data structures, Space-Time Tradeoffs and Efficiency)
	Strategic knowledge	Synthesis	(1) Requirements: understand the position of knowledge in the subject and be able to integrate and build a systematic knowledge system. (2) Contents: Use programs to solve scientific problems in specific fields and promote multidisciplinary integration.
		Evaluation	(1) Requirements: express and judge themselves on problems in the subject of study. (2) Contents: discover and explore new issues in this discipline from program designs, and propose innovative viewpoints and solutions.

B. Evaluation of Ability Achievement Degree

Abilities are constructed through training and practice on the basis of mastering specific knowledge. It belongs to the inner layer of the KAQ evaluation system. Computer programming ability is the core ability that needs to be cultivated in computer science and software engineering disciplines. The current way of evaluating programming ability is mainly divided into the evaluation of programming code and programming behavior. However, most of the existing methods for educators to assess programming ability are summative evaluations, focusing on code evaluation in programming learning. Not much research has been done on developing an improved programming ability evaluation and a unified indicator system or evaluation method.

According to the views represented by Caspersen scholars, programming abilities should include programming knowledge such as language, syntax, and algorithms [8], as well as various programming skills such as designing algorithms, writing programs, understanding program syntax and logic, and solving problems; according to the views represented by DalBey scholars, programming abilities should be reflected in the process of constructing code and its four stages: problem understanding ability, algorithm design ability, code writing ability, and program debugging ability corresponding to understanding, design, coding and debugging [9]. Therefore, programming evaluation indicators are designed for the above four programming abilities, as shown in Table 5.

Table 5 Ability Achievement Index

Primary Indicator	Secondary Indicator	Sub- Secondary Indicator	Description
Ability	Algorithm Design Capability	Input availability	Determine input availability refers to determining the valid input conditions.
		Expected outcome	State the expected outcome refers to knowing the final output beforehand.
		Structured decomposition	Structured decomposition problem refers to decomposing a complex problem into relatively minor, less complex problems, according to the structure of the problem itself. From the sub-problems, the entire solution process is systematically recombined to form a complete solution to the problem.
		Solution optimization	Solution optimizing problem refers to designing problem-solving methods that utilize the need for lesser time, space, and operation steps, achieving optimization of problem-solving efficiency.
		Visual representation	Visual representation problem refers to the use of visual tools to represent the thought processes of solving problems, usually includes the use of natural languages, flow charts, pseudocodes, etc.
	Programming Capability	Apply IDE	Apply IDE refers to correctly applying the application provider's development environment, such as their native code editors, compilers, debuggers, and graphical user interface tools, etc.
		Apply API	Apply API refers to correctly applying API knowledge during the processing of coding, use of API interfaces.
		Apply framework	Apply framework refers to being able to familiarize yourself with the usage of frameworks, and use frameworks in the coding process.
		Comply with coding standards	Comply with coding standards means that the code written must adhere according to formal coding standards.
	Debugging Capability	Finding errors	Finding errors refer to finding coding mistakes in the program's code, which includes two parts: finding syntax errors through compilation; and through operation, finding logic errors and processing errors.
		Positioning errors	Positioning errors refer to locating the exact spot of where the error occurred, including the use of debugging tools and

C. Evaluation of Quality Achievement Degree

Quality is the foundation for ability development, is the result of knowledge accumulation and transformation, and belongs to the core layer of the KAQ evaluation system. Specifically, quality refers to the general term for a relatively stable psychology, quality and self-cultivation, formed by internalizing knowledge and ability through acquired education and training. In the learning process, the quality of students not only determines the result of knowledge cultivation and ability cultivation, but also gives a rough estimation for the degree of knowledge acquisition and ability achievement. In college teachings, strengthening the cultivation of students' thinking mindset is conducive to the better cultivation of knowledge and ability, as shown in Table 6.

IV. ESTABLISH A FUZZY EVALUATION ALGORITHM

Fuzzy comprehensive evaluation method is a highly practical evaluation method, which is very suitable for the evaluation of things that have many attributes or are affected by multiple factors during the evaluation process.

It is clear that determining the weight of the evaluation indicator is the key link for fuzzy evaluation. Because the result-driven evaluation model established in this paper is a multi-level evaluation system, it is very compatible with the Analytic Hierarchy Process (AHP). The AHP was proposed by American operations researcher

Thomas L. Saaty in the 1970s. There are some main steps to calculate the weight by AHP [10].

A. Build a Hierarchical Structure Model

After an in-depth analysis of teaching quality in the program design course, evaluation indicators with several levels are designed, as shown in Table 1-Table 6. Each indicator at the same level is a factor attached to the previous level, and itself determines the next level of factors.

Table 6 Quality Achievement Index

Primary Indicator	Secondary Indicator	Tertiary Indicator	Description
Quality	Professional Quality	Teamwork Ability	The ability to cooperate with the team for a common goal, work together to complement each other, and achieve maximum work efficiency.
		Communication Ability	Able to handle interpersonal relationships.
		Learning Ability	Able to independently study professional knowledge and update one's knowledge system.
		Competitiveness	Able to seize any opportunity and make the best use of advantages.
		Management Skills	Able to lead or cooperate with a team or group to complete a comprehensive task.
		Innovation and Creativity	Able to propose, formulate, implement, and evaluate innovative ideas and solutions, and play differently by the book.
	Occupation Quality	Collecting and acquiring knowledge and information ability	Able to acquire new knowledge and information by reading domestic and foreign materials on the related field of interest.
		Ideological consciousness	Possess a positive outlook towards the outside world, life and values.
		Moral Characteristics	Good moral qualities, such as being law-abiding, helpful, honest, trustworthy, etc.
		Sense of Responsibility	Ability to perform tasks responsibly, take responsibility for mistakes, and be responsible for one's actions.
		Market Awareness	Possess awareness to compete in the face of competition, improve oneself, and play to any advantages available.
	Physical and Mental Quality	Physical Quality	Maintains a healthy physique in order to perform tasks.
		Mental Quality	A healthy personality and psychological quality, able to withstand setbacks.

B. Create a Judgment Matrix

The value of each factor in the judgment matrix reflects people's understanding of the relative importance of each factor. Firstly, the ranking index for each indicator is calculated according to the collected expert scoring data. The calculation method for ranking index is shown in Formula 1. Among them, aj represents the score of each level, nij represents the number of people selected for each level on each indicator, and N represents the total number of experts.

$$wi = \frac{\sum a_j \times n_{ij}}{N \times \sum a_j} \quad \text{Formula 1}$$

Then, the indicators are compared pairwise according to the ranking index using the 1-9 scale method. If there are two indicators, A and B, the ranking level scale of the pairwise comparison between indicator A and indicator B is shown in Table 7. Finally, a judgment matrix is constructed by comparing the indicators of the same level. Assuming that the constructed judgment matrix is A.

$$\mathbf{A} = (a_{ij})_{n \times n}, \quad a_{ij} > 0, \quad a_{ii} = 1, \quad a_{ij} \times a_{ji} = 1, i = (1,2,\ldots,n)$$

$a_{ij} = a_i : a_j$, where : indicates to compare the importance of a_i and a_j

Table 7 Ranking Level Scale between indicator A and indicator B

Rank scale	Description
9	Comparing indicator A with indicator B, indicator A is significantly more important than indicator B.
7	Comparing indicator A with indicator B, indicator A is much more important than indicator B.
5	Comparing indicator A with indicator B, indicator A is noticeably more important than indicator B.
3	Comparing indicator A with indicator B, indicator A is slightly more important than indicator B.
1	Comparing indicator A with indicator B, indicator A is as important as indicator B.
8,6,4,2	Median value of two adjacent grades.
Reciprocal	If the ranking index between A and B is n, then the ranking index comparing indicator B with indicator A is 1/n.

(Note: the value of Rank scale in Hierarchical Structure Model is given by a statistician and has been proved reasonable in practice)

C. Determine the Weight of Sub-Level Indicators

Firstly, use the square root method to calculate the geometric mean of each row of the judgment matrix, $\overline{W}_i$.

Then, normalization is performed to obtain the eigenvector W. Finally, a consistency test is carried out to ensure that the consistency of values is within an acceptable range. In this paper, we have improved the original formula $\widetilde{W_i}$, and it innovatively introduced the impact factor ρ_i of each ai. ρ_i is calculated by using deep learning technology to uncover implicit evaluation indicators in teaching data that have an influence on teaching quality. In other words that is, the element ai in A is characterized, and ρ_i is obtained by training.

Original Formula	$\overline{W}_i = \left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}} \ (i,j = 1,2,\ldots,n)$	$W = \frac{\overline{W}_i}{\sum_{j=1}^{n} \overline{W}_j} \ (i,j = 1,2,\ldots,n)$
Improved Formula	$\widetilde{W_i} = (\prod_{j=1}^{n} a_{ij} * \rho_i)^{\frac{1}{n}}$	$W = \frac{\overline{W}_i}{\sum_{j=1}^{n} \overline{W}_j} \ (i,j = 1,2,\ldots,n)$

Usually, the CR (Consistency Ratio) random consistency ratio is used to test the consistency of the judgment matrix, CR = CI/RI, where CI represents the consistency ratio of the judgment matrix ($CI = \frac{(\lambda_{max} - 1)}{n-1}$, where λ_{max} is the largest eigenvalue), RI represents the average random consistency index of the judgment matrix. Different RI values are used for different matrix orders, as shown in Table 8.

If CR<0.1, it means that the consistency of the judgment matrix A is ideal, and its eigenvectors can be used as an indicator weight for the evaluation model; otherwise, the values of the elements in the judgment matrix A need to be re-evaluated and adjusted, until A meets the consistency requirement.

Table 8 RI values under different orders of matrix A

Order	1	2	3	4	5	6	7	8	9
RI	0.00	0.00	0.50	0.90	1.12	1.26	1.36	1.41	1.46

(Note: Ri will be affected by the order of matric A; RI and the order have an increasing relationship)

V. EXPERIMENT

The experiment takes students' learning evaluation as an example to illustrate the practical data for the result-driven evaluation model in a program design course. As an example, students in a certain program design course for the academic year 2020-2021 are selected as the evaluation subjects. There are a total of 90 students in this particular class. The teaching course adopts a mixed teaching mode of online and in-class. Teachers carry out teaching activities through in-class teaching while also having online experiment guidance. Students obtain relevant knowledge points through attending in-class lectures and online experiments.

A. Construct a Learning Evaluation Scale

Before a model is constructed, it is necessary to clarify the evaluation criteria, that is, to form an evaluation scale. Taking the secondary indicators for the learning evaluation as the evaluation standard, we can divide it into five levels: 5, 4, 3, 2, and 1. Of which, 5 means 90 points and above, 4 means 80~90 points, 3 means 70~80 points, and 2 means 60~70 points, 1 means below passing grade. The evaluation scale constructed is shown in Table 9 (excerpted knowledge table).

Table 9 Design and evaluation of a learning

Primary Indicator	Secondary Indicator	Tertiary Indicator	5	4	3	2	1
Knowledge	Contents of Table 1						
Ability	Contents of Table 2						
Quality	Contents of Table 3						

B. Evaluate Collected Data

The evaluation of students' learning behavior mainly adopts two methods: automatic code evaluation based on OJ platform and self-evaluation based on questionnaire surveys. The students' learning situation is evaluated from the two aspects: on one hand, the evaluation is based on the data recorded on the OJ platform, such as the number of homework submissions, the correctness, the number of times students participated in discussions, etc; on the other hand, students evaluate their own in-class performance and learning situations through various evaluation questionnaires. These questionnaires are mainly filled in anonymously by students, thus ensuring the authenticity and validity of the evaluation data.

C. Leaning Evaluation Data Analytics

The learning evaluation data of 90 students is sorted and counted according to the learning evaluation scale. The statistical results are shown in Table 10. The larger the Average Score (Scoring Rate), the better the students have the corresponding indicators. Further, the more the items with higher scores, the better the results of the teaching evaluation of this course. A higher teaching evaluation result translates into a better quality course.

Table 10 Learning statistics

Primary		Secondary	5	4	3	2	1	Average Score	Scoring Rate
Knowledge	Declarative knowledge	Cognition	14	12	10	4	2	3.7619	75.24%
		Understanding	18	23	20	12	8	3.3827	67.65%
	Procedural knowledge	Application	25	28	24	15	10	3.4216	68.43%
		Analysis	22	20	18	13	9	3.4024	68.05%
	Strategic knowledge	Synthesis	13	16	10	7	4	3.54	70.8%
		Evaluation	11	13	8	6	3	3.5610	71.22%

(Note: the number 1-5 represents the grade indicators, they are excellent, good, normal, bad and awful. Their value indicates the number of students counted at each grade; Average Score = $\sum grade\ value * grade\ number$/Total number of students; Scoring Rate = Average Score/ 5)

VI. CONCLUSION

This paper proposed an evaluation model for programming course based on the learning outcome. It mainly studied three aspects: to improve the evaluation method, to perfect the evaluation system and to optimize the evaluation algorithm. The innovations are as follows:

(1) First, unlike most teaching evaluations that analyze the evaluation indicators directly from teaching content and teaching objectives [11], this paper proposes an outcomes-driven backwards design evaluation method. And our method can effectively help navigate the design of evaluation indicators, and convert traditional theoretical research into practical models proposed in [1] that can be used. The evaluation method tries to better reflect the student-centered teaching concept, and ensure the consistency of teaching objectives with the students' and social requirements.

(2) Second, deeply integrate between the teaching evaluation and objective law of teaching. Based on the evaluation system based on Bloom's taxonomy propose in [7], we further investigate students' cognitive process and the characteristics of the programming course, and divide the evaluation content into three aspects: knowledge, ability, and quality (see Table 4, 5 and 6). We then use quantitative and qualitative means to extract more than 40 evaluation indicators, which will have high reference significance and practical value for the formulation of teaching standards.

(3) Third, we innovatively introduced a dynamic evaluation algorithm using an impact factor pi (in part C of section IV) based on the original formula with static arguments [12]. It used deep learning technology to uncover implicit evaluation indicators in big teaching data that have an influence on teaching quality, that is, it can dynamically adjust the weight of the teaching indicator, making the evaluation results much more accurate.

It can be seen the Outcomes-Driven Assessment Model proposed in this paper is applicable to programming course. In the future, we will apply this method to other courses and study a general evaluation framework.

ACKNOWLEDGMENTS

This work was supported in part by NSFC (No.62107002) and National Key R&D Program of China (2021ZD0110703).

REFERENCES

[1] Graff N. "An effective and agonizing way to learn": backwards design and new teachers' preparation for planning curriculum[J]. Teacher Education Quarterly, 2011, 38(3): 151-168.

[2] Tanggaard L. Content-Driven pedagogy: on passion, absorption and immersion as dynamic drivers of creativity[M]//Dynamic Perspectives on Creativity. Springer, Cham, 2019: 165-177.

[3] Darling-Hammond L. Doing what matters most: Investing in quality teaching[M]. National Commission on Teaching & America's Future, Kutztown Distribution Center, 15076 Kutztown Road, PO Box 326, Kutztown, PA 19530-0326, 1997.

[4] Day C, Gu Q. The new lives of teachers[M]. Routledge, 2010.

[5] Spooren P, Brockx B, Mortelmans D. On the validity of student evaluation of teaching: The state of the art[J]. Review of Educational Research, 2013, 83(4): 598-642.

[6] Marouf A, Yousef M K A, Mukhaimer M N, et al. An Intelligent tutoring system for learning introduction to computer science[J]. International Journal of Academic Multidisciplinary Research (IJAMR), 2018, 2(2).

[7] Ullah Z, Lajis A, Jamjoom M, et al. Bloom's taxonomy: A beneficial tool for learning and assessing students' competency levels in computer programming using empirical analysis[J]. Computer Applications in Engineering Education, 2020, 28(6): 1628-1640.

[8] Bennedssen J, Caspersen M E. Abstraction ability as an indicator of success for learning computing science?[C]//Proceedings of the fourth international workshop on computing education research. 2008: 15-26.

[9] Dalbey, J., Linn, M.C.: Cognitive consequences of programming: augmentation to BASIC instruction. J. Educ. Comput. Res. 2, 55–72, 1986.

[10] Vaidya O S, Kumar S. Analytic hierarchy process: An overview of applications[J]. European Journal of operational research, 2006, 169(1): 1-29.

[11] Ozdamli F, Ozdal H. Developing an instructional design for the design of infographics and the evaluation of infographic usage in teaching based on teacher and student opinions[J]. EURASIA Journal of Mathematics, Science and Technology Education, 2018, 14(4): 1197-1219.

[12] Supriyanto A, Hartini S, Syamsudin S, et al. Indicators of professional competencies in research of Guidance and Counseling Teachers[J]. Counsellia: Jurnal Bimbingan dan Konseling, 2019, 9(1): 53-64.