On Deeper Learning Evaluation Model for College Students Based on SOLO Taxonomy

Li Qian, Fanglin Chen and Ningjiang Chen *

School of Computer and Electronic Information, Guangxi University, Nanning, 530004, Guangxi, China
chnj@gxu.edu.cn.

Abstract. With the development of information technology and education concepts, shallow learning in the traditional classroom can no longer meet the social demand for high-quality talents. Deeper learning emphasises learners' thinking ability and knowledge transfer, and plays a key role in problem solving. However, the existing research mainly focuses on technical support and disciplinary application. Empirical research and evaluation system construction of deeper learning is slightly insufficient. This results in a vague awareness of the evaluation of deeper learning in the whole teaching system. The SOLO taxonomy is characterized by its ability to measure student understanding and evaluate performance. It can be applied to the evaluation of deeper learning. This paper combines the theory to construct a deeper learning assessment tool and revise the scoring criteria through the Delphi method to make it more suitable for the characteristics of deeper learning of college students. Empirical research and experimental validation of the deeper learning evaluation model were carried out by preparing questionnaires and analysing online classroom cases. The results showed that the accuracy of the evaluation was more than 93 per cent.

Keywords: SOLO taxonomy, College students, Deeper learning evaluation model.

1 Introduction

Higher education has been advocating deeper learning, which enables students to solve real-world problems using critical thinking, self-directed learning and mutual collaboration. At present, deeper learning has become a hot issue in the fields of educational technology, learning sciences, curriculum and teaching reform [1]. In the face of ever-changing technology and complex information, people are no longer satisfied with low-order thinking activities such as simple extraction, mechanical memory and shallow understanding. Learners need to extend their learning beyond the classroom. At the same time, they focus on the breadth and depth of knowledge to promote the connection of existing knowledge and experience, carry out migration in new problem situations, and thus creatively solve practical problems and similar issues. The evaluation of the educational and teaching activities is a key link in educational and teaching activities. Proper evaluation can not only determine the real ability level of learners and provide

H. M. Briel (ed.), *Proceedings of the 2024 SSEME Workshop on Social Sciences and Education (SSEME-SSE 2024)*, Advances in Social Science, Education and Humanities Research 864,
https://doi.org/10.2991/978-2-38476-289-7_5

timely feedback to learners, teachers and education managers, but also provide guidance for teaching optimization and regulate the teaching design and process. Compared with general learning, deeper learning pays more attention to learners' higher-order thinking ability, reflection and innovation ability, problem solving ability, transfer and connection ability, etc. Therefore, the evaluation for deeper learning must be more thorough, comprehensive, and specific. However, current evaluation methods have some flaws. They cannot effectively reflect learners' deeper learning level or promote deeper learning.

Currently deeper learning evaluation focuses on how to judge whether deeper learning occurs and how to facilitate deeper learning to occur. There are two main tendencies regarding how to judge whether deeper learning occurs or not: learning mode and learning outcomes. In the early stage of research, deeper learning is mainly defined as a learning mode, and those who hold this view mainly include Marton, Ramsden [2] and Biggs, and the Study Process Questionnaire (SPQ) constructed by Biggs based on the 3P learning process theoretical model was used to judge the deeper learning status of the students, and due to the high reliability and validity, the questionnaire has become a typical way to judge the evaluation of the deep learning process of the students.[3, 4]. Deeper learning evaluation research focuses on learning outcomes, including students' higher-order thinking such as criticism, reflection and innovation, among which the typical one is the systematic assessment system developed by Smarter Balanced Assessment Consortium (SBAC), which evaluates a variety of key competencies such as creative thinking, knowledge innovation and problem solving ability as mentioned in the U.S. Common Core Standards [3, 5]. Biggs et al. proposed the evaluation method of the Structure of Observed Learning Outcomes (SOLO) aiming at the complexity of the thinking structure, which inferred the thinking structure state of students through explicit behaviors, and then assessed the deeper learning situation of students at that time.

SOLO is a qualitative assessment method characterised by hierarchical descriptions [6], aiming to judge the level of knowledge understanding of individuals through the quality of learning outcomes. The evaluation of deeper learning must reflect its essence and features, aligning with its value orientation. Key to this is dividing levels of understanding. The SOLO taxonomy fulfills this need. It divides understanding levels in a measurable and evaluable manner. These levels reveal changes in students' learning quality and are grounded in specific learning activities.

Based on the five levels of thinking of SOLO taxonomy, this paper constructs an evaluation index system with reference to the mature evaluation framework, then revises it through the Delphi method, and finally determines the assessment tool for evaluating students' learning performance. Develop a questionnaire and test its reliability and validity through a pre-survey, and use the questionnaire to supplement data for evaluation. Furthermore, an evaluation model is constructed, encompassing the entire learning and evaluation process. To assess its effectiveness, we test it using real-world online course cases. Lastly, the fuzzy comprehensive evaluation method is employed to analyze the questionnaire, while assessment tools are used to evaluate classroom observations. The performance of the evaluation model is verified by analyzing key indicators, such as accuracy and precision.

2 Construction of Deeper Learning Evaluation Model

Understanding the deeper learning status of students usually uses survey methods such as questionnaires or scales. Online questionnaires break through the limitations, survey multiple subjects, and are easy to study quantitatively. However, the design is fixed and cannot be adapted to complex situations, and the authenticity is uncertain. Interviews can provide in-depth information, but the skills required are demanding and time-consuming. Classroom observation method is direct, but cannot understand the inner activities. After weighing the advantages and disadvantages of each method, it was decided to use a combination of the classroom observation method to understand the students' task completion, the interview method to supplement the data, and then use the scales to investigate the students' deeper learning.

2.1 Deeper Learning Evaluation Index System Design

In this paper, the university English programme is taken as an example, and the research object is the first-year students. Based on the SOLO taxonomy, the teaching requirements of university English courses, the international assessment programme PISA, the compatibility framework, the NSSE [7, 8], and relevant studies in the literature [9, 10], and combined with our knowledge of the various levels of understanding of the SOLO taxonomy, a deeper learning evaluation index system in the context of university English courses is formed, as described below:

①Pre-structural level. The student does not answer or responds to content unrelated to the question, or states the teacher's question in a different way; basically does not have the appropriate knowledge to solve the problem.

②Single-point structure level. Students are not detached from the content of the textbook and are unable to put forward their own views on specific content and issues; they can only answer one point of the learning task based on the learning material and lack systematic thinking, such as simple identification and substitution of phrases to complete sentences; they understand the issue superficially and draw conclusions that are one-sided and incomplete.

③Multi-point structure level. Students are able to answer multiple points of a learning task but do not link the points together, e.g., give examples, categorise, etc.; draw different conclusions about the same material.

④Correlation structure level. Students can grasp clues and information about multiple aspects of a problem and summarise and integrate them, e.g. they can compare and analyse them in relation to relevant theories and explain cause and effect, but they lack the involvement of self-experience.

⑤Abstraction and extension of structural level. Students are able to integrate existing materials and external information and put forward their own views and insights, such as knowledge transfer and reasoning; abstractly summarise original problems and generalise about unexperienced situations; they are able to appreciate the significance behind the problem, and apply relevant knowledge when encountering practical problems; and their conclusions are open-ended.

These five first-level indicators represent the gradual deepening of learners' knowledge understanding. The first three levels are basically shallow learning, and the last two levels are mainly deep learning. We identified specific indicators under each dimension and revised the evaluation indicators using Delphi method. The expert group of evaluation indicators is composed of 5 doctoral students, postgraduates and teachers in universities who have rich learning or teaching experience in SOLO classification theory and deep learning evaluation. After two reviews, the expert group unanimously approved the revised assessment tool. The average score of 16 evaluation indicators was greater than or equal to 4, and the standard deviation was less than 0.8, indicating that the members of the expert group had unanimous opinions and could be used for deep learning assessment of college students.

2.2 Deeper Learning Questionnaire Development for Undergraduates

A questionnaire was designed in order to understand the level of deeper learning of the students, especially to make up for the shortcomings of the classroom observation method. The questionnaire is based on four dimensions: motivation, strategy, commitment, and outcome, and specific questions are set to improve the reliability and validity. We used Questionnaire Star online to create the questionnaire and conducted a pre-survey. The valid questionnaires were analysed for reliability. The results showed that the reliability coefficient of the motivation dimension was 0.827, the strategy dimension was 0.886, the input dimension was 0.822, and the outcome dimension was 0.884. The overall reliability coefficient of the questionnaire was greater than 0.9, which indicated that the reliability was of high quality, and could be used for further analysis. In the structural validity analysis, we used the validated factor analysis method, and the model met the standard in all fitting indicators, had a good fit, and could be used for further investigation.

2.3 Deeper Learning Evaluation Model for College Students Based on Solo Taxonomy

The evaluation of deeper learning is not only to consider what aspects to evaluate, how to implement the evaluation, how to analyse the results, etc., deeper learning and evaluation are both a complete process, and the results of evaluation are closely related to every aspect of the process. Therefore, based on the connotation of deeper learning, this study incorporates the instructional design into the evaluation process and designs the whole process of implementing deeper learning evaluation, the specific model is shown in Fig. 1.

The evaluation model is divided into three phases: pre-preparation, evaluation implementation and summary reflection. Pre-preparation includes tool preparation and understanding the instructional design. The tools used are classroom observations, interviews, questionnaires and assessment tools. Classroom observation is the key source of data, recording student behaviour and outcomes during specific learning tasks. The data collected then determines the quality of deep student learning based on assessment tools. If it was not possible to understand student learning outcomes through classroom observations, data were supplemented through post-classroom interviews. Questionnaires completed by students were used to understand students' inner activities, to improve the accuracy of the assessment, and to promote student reflection. In the implementation phase, evaluators observe student behaviour and determine the level of deeper learning, students fill in questionnaires, and then the data obtained are analysed. Finally, summarize the data processing results and reflect on the entire process. Feed back the results to the first two stages and provide targeted improvement suggestions. At the same time, the results are fed back to teachers and students to promote teachers to improve teaching and students to improve deeper learning levels.

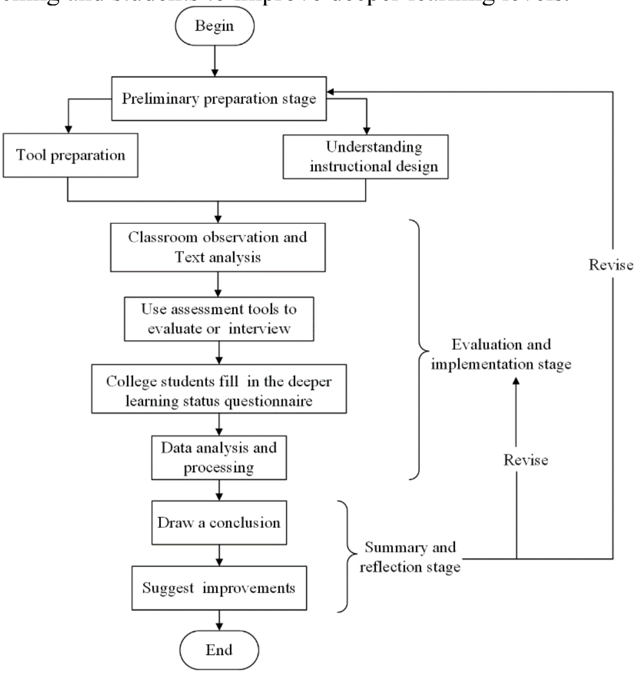

Fig. 1. Deeper Learning evaluation model for college students based on SOLO taxonomy.

3 Empirical Evaluation

In this study, the data of 59 course playbacks were collected from a freshman year College English course of a major in science at a university. Select students who answered questions in class and five representative students with varying learning profiles as observation subjects. Additionally, their practice problems, test papers, homework books, and notes were collected for analysis. The data coding was grouped according to the course progress, with the learning tasks as clues, noted as T_1 to T_n, respectively. Similarly, the students' learning outcomes were X_1-O_1......O_n, X_2-O_1......O_n, X_n-O_1... ...O_n. Through a deeper learning assessment tool, five researchers scored the student learning outcomes and averaged them to ensure objectivity.

3.1 Analysis of Results Based on Scale Evaluation

In this study, five graduate students from the fields of deeper learning and educational technology were invited to form an evaluation team to evaluate the data collected from course playback, classroom observation, and text in real time. Before the formal evaluation, the evaluators gained a deeper understanding of the deeper learning evaluation model for college students based on the SOLO taxonomy through online discussions and formed a unified evaluation standard.

Overall Student Deeper Learning. In the lesson "Smart Technology, Smart Life", the teacher set three learning objectives and carried out 38 learning tasks, 21 of which involved comprehension learning. Students produced a total of 49 learning outcomes, of which 22 were comprehension learning outcomes.

Table 1. Overall student understanding based on response results.

learning goal	Learning task	Comprehension task	Level 0 Answer	Level 1 Aswer	Level 2 Answers	Level 3 Answers	Level 4 Answers
Objective 1: What makes smart cities smart?	20	9	0	5	3	2	0
Objective 2: How to make cities smart?	12	8	1	6	1	1	0
Objective 3: To grasp the main idea of the article, analyse the writing techniques and apply them to writing	6	4	0	3	0	0	0
Total	38	21	1	14	4	3	0

Based on the learning outcomes of the responding students: As shown in Table 1, out of 22 comprehension learning tasks, 14 learning outcomes were at level 1, accounting for more than 63 per cent of the total; the highest level was 3, with only 3 items, accounting for no more than 14 per cent. This suggests that students' comprehension is mostly at a superficial stage, with a strong reliance on textbook content.

Table 2. Based on the level of understanding of the sample students.

Sample students	Comprehension Learning Outcomes	Level 0 answer	Level 1 answer	Level 2 answer	Level 3 answer	Level 4 answer
Student A	19	3	13	1	2	0
Student B	20	0	14	2	3	1
Student C	18	9	7	2	0	0
Student D	20	2	14	1	3	0
Student E	18	6	10	2	0	0
Total	95	20	58	8	8	1

Based on the learning outcomes of the sample students: As shown in Table 2, the five sample students produced a total of 95 learning outcomes, of which 86 were at the shallow learning level, accounting for about 91 per cent; only 9 items were at the deeper learning level. Detailed analyses showed that students' understanding of the problem was mainly focused on Level 1, with fewer cases reaching the deeper learning level.

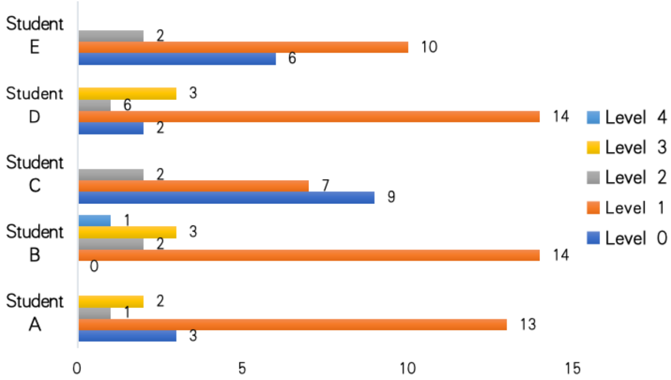

Fig. 2. Differences in the level of comprehension of students from different academic backgrounds.

Learning outcomes for Student A, Student B and Student D were mainly concentrated at Level 1, with a high proportion of about 68 per cent or more. A small number of learning outcomes at Level 3 also appeared in Student B and Student D, accounting for about 15 per cent, while Level 0 accounted for about 16 per cent of Student A's learning outcomes. Among them, Student B showed some critical thinking and was able to analyse and relate multiple knowledge points from English problems, while Student A and Student D were more average. Students C and E showed a shallow level of learning, indicating a superficial understanding of the problem. Overall, students' learning outcomes were mainly at Level 1. This shows that they mainly answer questions from a single perspective, mostly based on life experience rather than knowledge of English. It shows the trend of shallow learning. This may be because the teacher's task setting tends to be shallow. However, there are very few students with deep learning level, and their proportion is very small.

Deeper Learning with Differences in Academic Conditions. According to the teachers' and classmates' evaluations of the five sample students, Student A and Student D have a good learning foundation, Student B is at an intermediate level, while Student C and Student E are struggling students with a relatively poor level of English. Fig. 2 shows the differences in the comprehension levels of the students in different academic situations.

Based on the above analysis of the deeper learning of the five sample students, there was no significant difference in deeper learning between students of different academic backgrounds, but students with poorer fundamentals were more inclined to show a lack of understanding of the learning tasks.

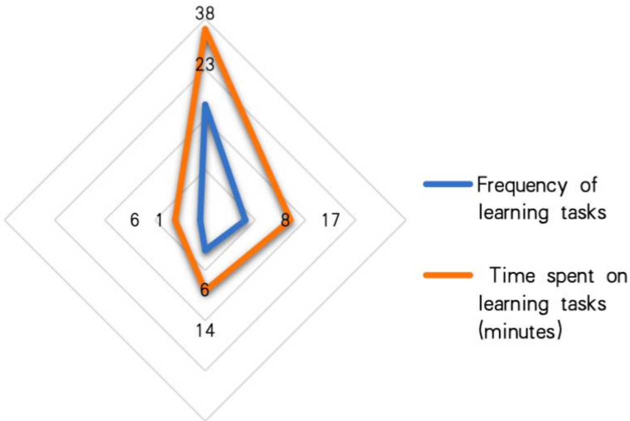

Fig. 3. Frequency and time spent on various learning tasks in the classroom.

Deeper Learning for Different Learning Tasks. Teachers design four types of learning tasks: "What", "Why", "How" and "Have a talk". Fig. 3 shows the frequency and time spent on each type of learning task.

In the classroom, 23 "what" type learning tasks were completed, taking about 38 minutes, or about 42 per cent of the time. This was followed by 8 "why" type learning tasks, which took about 17 minutes, or about 19 per cent of the time. Next came the "how" type of learning tasks, with a total of 6 items, taking about 14 minutes, or an average of about 2 minutes per task. Lastly, there was only one open-ended task, "Tell me about it", which took about 5 minutes and required students to reflect on what they had learnt and their life experiences.

3.2 Analysis of Student Questionnaire Results

After the end of the course, the Questionnaire on the Status of Deeper Learning among College Students was made into an online version through Questionnaire Star, and then sent to 44 students in the research class to fill in at a unified time with the help of the teacher and classmates, to learn about the situation of deeper learning from the perspective of the students, and at the same time, it is also a supplement to the classroom observation data. At the end of completion, 44 questionnaires were returned, out of which 1 questionnaire became invalid due to improper completion, i.e., 100 per cent of questionnaires were returned with an effective rate of about 97.73%. Among the 43 valid questionnaires, boys filled out 20, accounting for about 46.51%, and girls filled out 23, accounting for about 53.49%.

The questionnaire is divided into four dimensions: motivation, strategy, commitment and outcome, with specific questions set under each dimension, each of which is followed by a corresponding five-point Likert scale. Currently, the commonly used method is to use SPSS statistical software for analysis, while this study tries to quantify the subjective evaluation indexes in the questionnaire by using the fuzzy comprehensive evaluation method. The fuzzy comprehensive evaluation method is based on fuzzy mathematics, firstly, the factors that are not easy to quantify are quantified, and then the vector product is used to calculate the weights of the indexes and the grade scoring, and finally, the comprehensive evaluation results of all the evaluation objects are sorted to choose the best, and also the values on the fuzzy evaluation set can be used to evaluate the grade to which the research object belongs according to the principle of the maximum degree of affiliation [11, 12, 13].

Five sample students are analysed here:

①Indicator set: U = (u1, u2, u3, u4) = (motivation to learn, learning inputs, learning strategies, learning outcomes);

② Set of comments: $w = (w_1, w_2, w_3, w_4, w_5)$ = (fully compliant, compliant, fair, not compliant, not compliant at all);

③ Weight set: A = (0.2, 0.2, 0.2, 0.4), as assessed by expert experience;

④ Since the evaluation team consists of experts, $K = (1, 0)$;

⑤ Fraction set: $B = (100, 75, 50, 25, 0)^T$

⑥ The fuzzy relationship from U to W, denoted by the fuzzy evaluation matrix R.

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{bmatrix} \tag{1}$$

where (i = 1, 2, ..., m; j = 1, 2, ..., n) denotes the degree of affiliation of the j-th level of comment made on the i-th evaluation indicator.

⑦ According to the evaluation of the expert group, for the i-th evaluation indicator of a certain evaluated subject, there are w_{i1} w_1-level comments, w_{i2} w_2-level comments, ..., w_{in} w_n-level comments. Then:

$$r_{ij} = w_{ij} / \sum_{i=1}^{n} w_{i1} \qquad j = 1, 2, \cdots, n \tag{2}$$

⑧ Integrated evaluation model: $P = A \circ R = (P_1, P_2, ..., P_n)$

⑨ Finally, the final evaluation result is obtained: $C = P \circ B$

Then the evaluation results for the five evaluation subjects are in order: 79, 81.5, 71.75, 76.25, 64. It can be seen that the second student has the highest level of deeper learning but only reaches 81.5, and the last student is even below 65, which indicates that most of the students are in the state of shallow learning, which is not very much different from the above analysis of the scale.

3.3 Model Validation

After completing the construction of the model, the effectiveness of the model needs to be evaluated so that the model can continue to be adjusted until it achieves good enough results. Determining whether a student is at a shallow or deeper learning level is a dichotomous problem, and the statistics that can be used to measure the performance of their evaluation model include Accuracy, Precision, Recall, F1-score, AUC, and ROC curves, etc., and they are all related to each other, but just focusing on different aspects. different aspects [14].

The class selected for this study consisted of 44 students, and the expert panel evaluated the students' classroom observation, text collection, and questionnaire data through the assessment tool. The data obtained are shown in Table 3, where deeper learning is used as a positive sample (P) and shallow learning is used as a negative sample (N), and the prediction result is T for correct and F for wrong.

Table 3. Results of deep versus shallow learning judgements.

Projected actual	deeper learning	shallow learning
deeper learning	3 (TP)	2 (FP)
shallow learning	1 (FN)	38 (TN)

That is, 38 students were predicted to be at a shallow level of learning while only 3 were predicted to be at a deep level of learning; the predictions did not match the actual situation: 1 student was actually at deeper learning but predicted to be at a shallow level of learning, and 2 students were actually at shallow learning but predicted to be at deeper learning. Calculate the accuracy according to the following formula:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

The judgement of students' knowledge comprehension is not exactly the same as the general dichotomous model, the former does not involve the issue of correctness and error, so the judgement here is mainly based on the accuracy rate, and other indicators are used as auxiliary observations. From the calculation formula, the accuracy rate is about 93.18%, and the judgement ability of the model is better.

3.4 Discussion

Current research on evaluating deep learning is relatively scarce. Existing studies focus more on theoretical aspects, such as evaluation strategies and the importance of various theories in deep learning assessment. However, empirical research is somewhat lacking, and there is still a lack of a unified standard evaluation paradigm. Reference [15] developed a classroom assessment tool based on a fundamental indicator to evaluate deep learning but lacked experimental research. Reference [16] discussed how to create evaluation systems to support deep learning but remained in the theoretical research stage. This study not only identified evaluation indicators but also designed a comprehensive process model for the entire teaching design and evaluation process, validated through experimentation.

In this study, the college students' course was selected as a teaching case for experimentation, collects data through various methods such as classroom observation, text collection, questionnaires, and interview feedback, and conducts detailed organization and analysis of these data. Research results show that the overall understanding level of college students is low and the degree of in-depth learning is insufficient. Most students stay at the single-point or multi-point structure level and have not yet formed their own knowledge system structure. At the same time, the study also found that students with different academic backgrounds did not show significant differences in their understanding levels, and there was no obvious correlation between students' understanding levels and basic knowledge. In addition, students' learning results are greatly affected by the learning goals and tasks set by teachers. In the 90-minute class, the teacher carried out a total of 38 learning tasks. More than half of the time is spent completing "exploratory" learning tasks. To complete these tasks will inevitably lead to too many "teachers ask questions and students answer" situations. Students have little time and little opportunity to engage in in-depth thinking, collaborative discussions, reflection and improvement, and other activities. This in turn leads to superficial responses from students. It is worth mentioning that the accuracy of the deep learning evaluation model is as high as 93.18%, showing good judgment ability.

4 Conclusion

In the paper, we combine the evaluation index system constructed based on SOLO taxonomy with the questionnaire on deeper learning of college students to evaluate the deeper learning level of college students, which is more in line with the characteristics of deeper learning of college students. In addition, based on the teaching design and evaluation process, we have constructed a complete evaluation model from early preparation to summary and reflection. This model makes deeper learning evaluation more operable and applicable, and provides a new evaluation paradigm for evaluating students' deeper learning. Future research can further verify the model effect by expanding sample data, strengthening the professionalism of raters, and strengthening the application of computer technology.

Acknowledgments. Funded projects: Innovation Project of Guangxi Graduate Education (JGY2022020)

References

1. Linda Darling-Hammond, Jeannie Oakes, Steve Wojcikiewicz, et al: Preparing teachers for deeper learning. Harvard Education Press, Cambridge (2021).
2. Marton, F., Saljo, R: On Qualitative Differences in Learning: I—Outcome and Process. British Journal of Educational Psychology 46(1), 4–11 (1976).
3. Fatemah Alahmari, Amani Basudan, Marwa Shaheen: Assessment of Learning Approaches of Saudi Dental Students Using the Revised Two-Factor Study Process Questionnaire: A Cross-Sectional Study. The Open Dentistry Journal 17 (2023).
4. Yusuf F. Zakariya, Barattucci Massimiliano: Short form of revised two-factor study process questionnaire: Development, validation, and cross-validation in two European countries. Studies in Educational Evaluation 75 (2022).

5. Herman Joan, Linn Robert: On the Road to Assessing Deeper Learning: The status of Smarter Balanced and PARCC Assessment Consortia. pp. 10–12. University of California, CRESST, Los Angeles, CA (2013).
6. John Biggs, Kevin Collis: Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome). Academic Press, New York (2014).
7. Janita Rawls, Stacy A. Hammons, Andrew V. Stephenson: Gaining Practical Competence in Higher Education. Journal of Financial Education 45(2), 206–224 (2019).
8. Jillian Kinzie, Samantha Silberstein, Dajanae Palmer: Elevating Student Voice in Assessment: Approaches to Using NSSE's Student Comments. Assessment Update 33(2), 1–16 (2021).
9. Gunnel Svensäter, Madeleine Rohlin: Assessment model blending formative and summative assessments using the SOLO taxonomy. European Journal of Dental Education 27(1), 149–157 (2023).
10. Blain Patterson: Analyzing student understanding of cryptography using the SOLO taxonomy. Cryptologia 45(5), 439–449 (2021).
11. Jie Cao, Bowen He, Nan Qu, et al: Benefits Evaluation Method of an Integrated Energy System Based on a Fuzzy Comprehensive Evaluation Method. Symmetry 15(1), 84–84 (2022).
12. N. Pagliarecci, F. Zimmer, A. Birouche, et al: Test methodology for the vehicle-tire handling performance evaluation: objectification of driver’s subjective assessment. IFAC PapersOnLine 53(2), 14394–14400 (2020).
13. Honghe Li, Nannan Jin, Limin Zhang: Construction of Health Education Competency Model for Nurses Based on Fuzzy Comprehensive Evaluation Method and Fuzzy Cluster Analysis Method. Mathematical Problems in Engineering 2022 (2022).
14. Ngwira, B., Gobin-Rahimbux, B., Sahib, N. G.: Deep-Learning Framework for Analysing Students’ Review in Higher Education[J]. Computational Intelligence and Neuroscience, 2023 (2023).
15. Christopher, J. Harris, Joseph, S. Krajcik, James, W. Pellegrino, et al. Designing knowledge-in-use assessments to promote deeper learning[J]. Educational measurement: issues and practice 38(2), 53-67(2019).
16. Conley, David T., Linda Darling-Hammond: Creating systems of assessment for deeper learning[J]. (2013).