IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 9, 2020, accepted September 28, 2020, date of publication October 9, 2020, date of current version October 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3029947

Analysis of Learning Patterns and Performance–A Case Study of 3-D Modeling Lessons in the K-12 Classrooms

YI-CHIEH WU AND WEN-HUNG LIAO, (Member, IEEE)
Department of Computer Science, National Chengchi University, Taipei 11605, Taiwan
Corresponding author: Wen-Hung Liao (whliao@nccu.edu.tw)

ABSTRACT As more schools incorporate 3D printing into their curriculum to stimulate the creativity of K-12 students with a learning-by-doing approach, it becomes crucial to understand how users work with 3D modeling tools and to evaluate integrated lesson plans in the STEAM (Science, Technology, Engineering, Arts and Math) educational framework. Our work consists of two stages: an investigation of usage patterns of modeling, and an evaluation of the usability of Qmodel Creator, in collaboration with Lanyu Primary and Junior High School, Sanchong High School, and Affiliated Experimental Elementary School of National Chengchi University. Participants operation logs, screen recordings, and finished work for respective 3D modeling software were recorded and analyzed. Two types of indicators have been developed. One is concerned with the quantification of learning behavior, including Effective Operating Period (EOP), Trial-and-Error Period (TEP), and Implementation Period (IP). The other has to do with the evaluation of learning outcome, i.e., the complexity of 3D models, including the Degree of Detail (DoD), shape (C_f), partition (C_p), and block-ratio (C_r) complexity. Based on the proposed features, we are able to identify the key factors that affect students' learning experiences and performance in terms of learning patterns and model completeness. Through these indicators, instructors can gain better insights into student's learning status of 3D modeling software.

INDEX TERMS Learning patterns, 3D modeling software, STEAM lesson plan, Qmodel creator, performance evaluation, K-12 education.

I. INTRODUCTION

Learning behavior begins for various reasons and in quite different styles. One can keep obtaining fresh knowledge and engaging with novel technologies, either for solving problems or just for fun [1]. Moreover, a material for learning may inspire one person but not another [2]. Researchers have proposed a variety of theories and models in an attempt to identify the key factors affecting our learning process. Motivation is an intuitive attribute [3], [4], having both intrinsic and extrinsic forms, and it can change over time [5]. Although learning motivation is an interesting aspect, there can be various types of motivation to explain the same learning behavior. It is thus a challenging task to select quantitative indicators for accurate measurement of learning motivation. In addition, individual differences are a critical issue. For educators who design different lessons for students according to their aptitudes, how can the performance of their lesson plans be assessed fairly? Is it attainable to observe and analyze their learning patterns objectively?

The K-12 classrooms serve as ideal test-beds for investigating learning behavior because multiple stages in learning development can be observed, explored, and examined [6]. It is also noted in [6] that children have the capability to plan, organize, and integrate the knowledge to be learned as the cognitive development matures in all aspects. Traditional questionnaire-based approach for assessing learning process and outcome has its limitations due to lack of objective measure. Moreover, choosing suitable materials for children in lesson plans is imperative. In this work, an integrated approach encompassing quantitative and qualitative evidence is adopted to study learning patterns in the K-12 classrooms.

A. BACKGROUND OF THE CASE STUDY

To enhance the competitiveness of the future workforce, the fundamental concepts and practices in science, technology,

The associate editor coordinating the review of this manuscript and approving it for publication was Juan A. Lara.

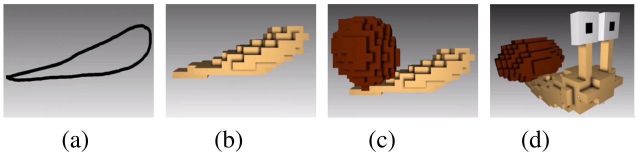


FIGURE 1. Demonstration on how to create a snail using Qmodel Creator [15]. (a) Sketch a contour, (b) System generates a rough model with thickness adjustment, (c) Sketch another part, and then combine all components into one, (d) Add details to the model.

engineering, art, and mathematics (STEAM) have been redesigned and developed as regular curriculum of K-12 education over the past decade [7]. Due to advances in 3D printing technology and the affordability of 3D printers, many teachers have introduced 3D printers [8], [9] and relevant lesson plans into K-12 classrooms [10] to build an integrated framework of STEAM education [11].

The common objective is to stimulate the creativity of K-12 students with a learning-by-doing approach [12]. However, studying with computers and 3D printing technology is not the same as tutoring the lessons of the original K-12 education, and it is still challenging in practice [13]. One of the potential issues is that the learning curve of new technologies is always high for both students and teachers. Moreover, applying the learning-by-doing method in lesson plans involves drawing connections between essential STEAM disciplines and results in a specific task. Multiple strategies might solve the above problem, depending on the level of domain knowledge [14]. This variability leads to another question: how to evaluate the effectiveness of the STEAM lesson plan and student's performance?

This research examines learning patterns when working with 3D modeling software. The process of 3D manufacturing usually includes the following stages: (1) constructing models using software or scanning tools, (2) editing and refining, and (3) 3D printing. In [13], [14] for example, the usage of 3D modeling and editing software is probably too complicated, so certain K-12 lecture cases focus more on the printing phase of the process. To ease the learning curve of modeling for K-12 students, we adopted Qmodel Creator [15], a cubic style modeling software with an intuitive user interface that quickly converts 2D drawings into 3D models. Fig. 1 shows how a 3D model of a snail can be generated using very simple sketches. It is also critical to provide teachers and students a platform for online communication, idea exchange, and peer evaluation. In response to the needs of the course, Lin *et al.* developed the 3D Model Co-learning Space (3DMCLS) [16], a web interface for students uploading their finalized work. It can also be achieved through the in-app function of Qmodel Creator if students create models using this application. The file format, Standard Template Library (STL), is supported by 3DMCLS, so teachers can easily utilize this platform by choosing any modeling software that can export STL files. Fig. 2 shows models on 3DMCLS created in different modeling programs, including Tinkercad [17], Qmodel Creator and 123D Design. Fig. 2 reveals an additional concern: the instructor has to assess heterogeneous models to evaluate students' learning status.

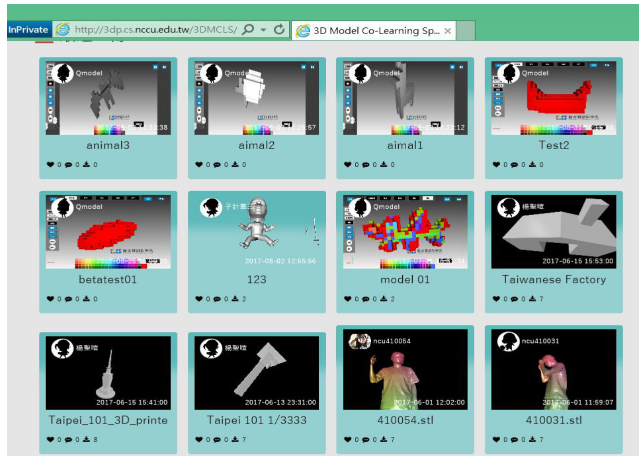


FIGURE 2. Models created in various 3D modeling software programs on 3DMCLS [16].

B. MAIN CONTRIBUTIONS

Our investigation centers on two perspectives: (1) student's learning pattern, and (2) the outcome after learning. The learning pattern informs us about individual's behavior during the learning progress, representing the skill level currently. Therefore learning patterns should be task-independent. Meanwhile, the outcome presents the production based on the domain knowledge after one's learning. Hence, the evaluation for an outcome is expected to be highly dependent to the task. The main contributions of this work are listed as follows:

1) IDENTIFYING FEATURES FOR LEARNING PATTERNS

We collected and derived the indicators from operation logs, including: Effective Operating Period (EOP), Trial-and-Error Period (TEP), Implementation Period (IP), mean and standard deviation of TEP and IP step. The operation logs of the participants were recorded not only as a reference to understand the learning behaviors, but also as a field observation (data-logging) of the usability test [18] of Qmodel Creator. Based on the results in this paper, we believed that the analysis of the modeling operating pattern is essential.

Although we could easily obtain the log file of Qmodel Creator, we still need to examine students' operating procedures when other commercial modeling applications are employed. Therefore, we adopted a video segmentation approach to make it easier to retrieve and label user operation indicators so that the same features of user behavior could be observed and compared when different software was employed.

2) DEFINING THE DEGREE OF COMPLETION FOR 3D MODELS

In order to evaluate the learning outcome, the learning task should be specified firstly. That is, we need to develop quantitative indicators to better characterize the constructed 3D model. We took multiple features of a model into consideration, including the Degree of Detail (*DoD*), the number of connected components, the largest component ratio (*LCR*) of a model, and the shape (C_f), partition (C_p), and block-ratio (C_r) complexity, which will be described in the subsequent section. In terms of multiple analytic results, these model-related features, especially C_r, are appropriate for representing the completion degree of 3D models.

3) DEVISING AN INTEGRATED QUANTITATIVE AND QUALITATIVE APPROACH

In addition to the above quantitative methods for learning patterns and the degree of completion of outcome, we also carried out qualitative investigation and analysis. Firstly, different from questionnaires which are more suitable for assessing adult learning patterns, we adopted observation approach to identify the issues students encounter when learning the modeling software. Besides, to further explore students' learning background and attitudes, we conducted face-to-face interview, which includes a simple drawing test, to improve for the efficacy of traditional questionnaires. From these qualitative data, we can profile the student's knowledge and ability (i.e. computer, art, etc.), and compare the qualitative analysis with the above quantitative results to examine the learning process from a broader perspective.

The items of qualitative data collection and analysis are adjusted according to the objectives of different stages. Our two-stage research evaluated the usability of Qmodel Creator and investigated the learning patterns of modeling in K-12 classrooms. In the first stage, we hosted two workshops to introduce Qmodel Creator to K-12 students and collected user logs. The conclusion of the stage was that there were detectable distinctions in the details of the finished models built by users from different skill level groups. Furthermore, our analysis indicated that students with different backgrounds had respective preferences on particular functions of Qmodel Creator. Finally, we examined the operation logs and concluded that Qmodel Creator was a suitable 3D modeling software for all ages of K-12 students.

In the second stage, we reviewed the performance of the previous experimental design and developed a lesson plan for teaching multiple modeling software and 3D printing in an elementary school. According to the operation patterns, we roughly divided students into the following four categories: Novice with little interest, Novice with caution, Intermediate, and Advanced. The proposed classification and quantitative indicators for models will help teachers to track students' progress and identify aspects that require further improvement.

The remainder of this paper is organized as follows. Section II reviews the referenced literature, including measures for both learning behavior and the degree of completion of 3D models. Section III elucidates the two categories of quantitative indicators. In Section IV, we present the experimental design of our two stage research. In Sections V and VI, the experimental results of the two stages are illustrated and discussed respectively. Section VIII concludes this work.

II. LITERATURE REVIEW

In the section, we will review literature in the two categories: learning behavior and 3D model complexity, which are fundamental in measuring learning pattern and learning outcome, respectively.

A. STUDIES OF LEARNING BEHAVIOR: LEARNING PATTERN

Learning behavior involves an interactive process of taking knowledge, responding, and presenting outcomes. The definition of learning pattern, however, varies slightly in the literature.

Vermunt *et al.* summarized the learning pattern [19], which consists of the four elements: cognitive processing strategies, meta-cognitive regulation strategies, (metacognitive) conceptions of learning, and learning motivations or orientations, and therefore can be identified as the following qualitative learning patterns: reproduction-directed, meaning-directed, application-directed, and undirected. The authors derived the "learning style" from [20] to "learning pattern" [21], which eventually defined the discerned behavior "as a coherent whole learning activities that learners usually employ, their beliefs about learning and their learning motivation", and constructed the cognitive framework based on the empirical evidence. Moreover, the patterns are interdependent with the personal factors (age, gender, culture), contextual factors (teaching style, collaborative type, lesson plan), and learning outcomes (score or performance), usually measured by questionnaires or Likert scale (in [22], the instrument is Inventory of Learning Styles, as known as ILS). However, the dedicated scales should be designed for the target students, as well as the sufficient number of data collection, ensuring the reliability and validity to support the analytic results of **the cognitive factors**. That is, the learning questionnaires for students of higher education may not be so appropriate in elementary schools.

What other measurement approach for learning pattern can be applied, especially introducing a brand new lecture or for detecting students who feel stuck in their learning process? Nemiro *et al.* have conducted a STEM program in K-12 classrooms over three years. In [23], the data collection methods for learning behavior are mainly observation (by Psychology undergraduate students as external experts) and journals (written by part of participants). Their interests lie in students' creative process and creativity techniques, and the two types of behavior are the essential quantitative indicators. Either external observation or self-journal is a kind of data logging approach. However, the logging frequency is weekly, and the quantitative results are not only representative of learning process, but also for the outcome. Also, lots of discussion is necessary to meet the consensus of the event coding. The degree of experts' mutual agreement is the key to the quality of the quantitative indicators.

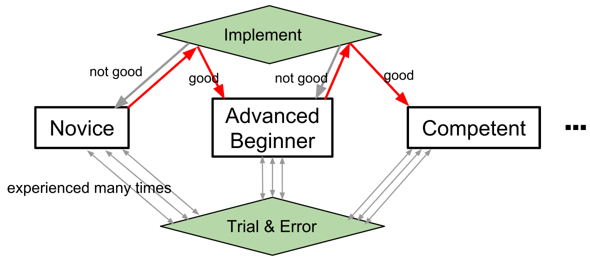


FIGURE 3. The assumed skill acquisition model derived from Dreyfus *et al.*'s work [28] and trial-and-error theory.

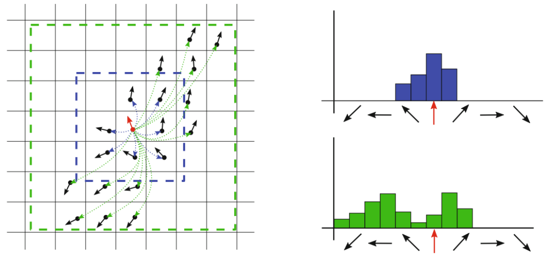


FIGURE 4. Shape descriptor extraction using Fiolka *et al.*'s method [35].

The importance of trial-and-error period has been pointed out in [23], "Once students had successfully completed programming a robotic challenge task, the long and frustrating iterative trial-and-error period ended with a feeling of relief, a sense of pride in one's accomplishment, and often behavioral displays of excitement accompanied this pride." Learning by trial-and-error is well observed as one of the primary methods in the development of human beings [24], [25]. Studies on the trial-and-error approach applied in education are abundant and have a historical context. On the other side, the effect of frustration reaction should be still considered in the progress of one's trial [26], [27]. We speculate whether the element of trial-and-error conduces to human's learning. Can the factor be defined or labeled explicitly?

Furthermore, can we generalize the human behavior of learning process as the evolving progress or status of skill acquisition? Dreyfus *et al.* [28] presented a phenomenology of skill acquisition by humans and offered a theoretical explanation for it. Based on subjective and objective results, they presume that the learning process goes through five stages: novice, advanced beginner, competent, proficient, and expert. The process from novice to expert is universal and applicable in many fields, including aircrew emergency decision [29], ethical action [30], and even evaluation of an AI system [31]. Since the learning activities exist in the interactive relationship between humans and their daily life, if we can record the details of how people learn as the actual evidence, we may find the distinguishing/unconscious patterns between skill levels, for example, from novice to advanced beginner.

Based on these theories, we assume a model that formulates the process of skill acquisition from novice to advanced beginner and then competent user, as depicted in Fig. 3. By analyzing the features of learning behavior in the trial-and-error period, the degree of mastery of the skill could be better understood.

B. STUDIES FOR MEASURING OUTCOME: 3-D MODEL COMPLEXITY

The means to evaluate the effectiveness of learning depend on the context of learning or tasks. For learners who have not matured to construct 3D models, what we are interested in is whether they will pay attention to an unfamiliar task, as well as whether the quality of the finished outcome reflects the degree of the author's passion. For more inquiring students, will their 3D models have more degree of details?

In order to obtain information from 3D models for further analysis, we explore some relevant literature in quantifying the complexity of an entity. Valentan *et al.* proposed several basic indicators of shape complexity to evaluate if the manufacturing procedure is optimal [32]. These geometry-based features inspire us to consider the shape complexity of models. Morphological and topological complexity have been discussed in [33], [34], and applied to a wide range of categories. Such algorithms work well on point clouds or more complicated models. 3D objects created by students are too simple to be properly characterized by these features.

Fiolka *et al.* [35] proposed 'SURE' descriptor in 3D point clouds based on entropy. The shape descriptor is extracted by computing histograms of surfel pairs of a local interest point. Fig. 4 illustrates the process of extracting the descriptor.

Liao and Chen also calculate the complexity of image based on entropy using information-theoretic modeling for logo design analysis [36]. Three different indices of complexity of the image can be computed: (1) partition-based complexity (C_p), (2) homogeneity-based complexity (C_f), and (3) area-ratio complexity (C_r). Fig. 5 shows the results of two logo analysis using their proposed measures. In this paper, we will employ the entropy of normal vectors to measure the complexity of a 3D model. We will also compare the difference between the entropy of normal vectors in 3D with the entropy of pixel intensity in 2D when used to gauge model complexity.

III. THE PROPOSED METHODOLOGY

In this section, we propose two classes of quantitative indicators, namely, usage pattern related features and 3D model related features, to address the performance evaluation of a general 3D modeling task.

A. USAGE PATTERN RELATED FEATURES

For extracting the features of a usage pattern, we adopted a field observation method. That is, the indicators: Trial-and-Error Period (*TEP*), Implementation Period (*IP*) and Effective Operating Period (*EOP*), were extracted from the user logs recorded using Qmodel Creator and the screen recordings of users operating the commercial modeling application. The flowchart we used to analyze usage patterns is presented in Fig. 6.

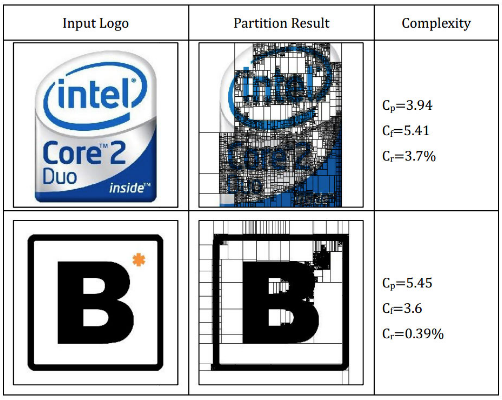


FIGURE 5. Results of logo complexity based on partition, feature, and ratio entropy [36].

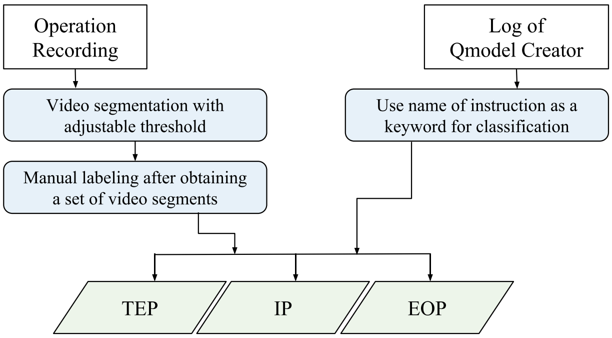


FIGURE 6. Workflow of usage pattern analysis.

1) FROM THE LOG OF QMODEL CREATOR

Fig. 7(a) depicts the Qmodel Creator user interface. For all participants, we recorded the operation along with the corresponding timestamp for subsequent calculation of the operation time span. We stored the information in a log file as follows: The first field was the timestamp, and the second field was the operation event. Fig. 7(b) presents a snapshot of the Qmodel Creator log file.

In Qmodel Creator, traditional 3D modeling functions such as adding and deleting voxels require more steps than using intuitive modeling. Certain function keys are shared (e.g., Clean/Undo/Redo). The following list contains operations that are classified as "intuitive". The rest are regarded as traditional 3D operations.

- Change mode to "Draw Simple"
- Change mode to "Draw Symmetry"
- Start drawing
- End drawing
- Click OK

As shown in Fig. 7-(b), drawing operations come in pairs, with "Click OK" as the operation logged after "Adjust Thickness" and "Click Confirm".

2) FROM THE OPERATION RECORDING

Fully automatic processing of operation videos for an arbitrary commercial modeling application is undoubtedly challenging and still unattainable. Instead, we implemented a video segmentation system using OpenCV [37] with an adjustable threshold so that the video could be separated hierarchically. Fig. 8 illustrates the simplified workflow of the usage pattern extraction from the operation recording. Through comparison with the finalized model, a labeler can focus on a video segment to tell whether the user is in a trial (Action 1, 3, and 4-3), implementation (Action 4-2 and 6-1), or ineffective operation (Action 2 and 5) mode.

More detailed examples reveal all the steps of the hierarchical segmentation further, as shown in Fig. 9. Fig. 9-(a) is a series of significantly segments using a larger threshold, representing an obvious change between frames, such as expanding a sub-menu or switching to another window. Fig. 9-(b),(c) and (d) are segmented results using decreasing thresholds respectively. Of course, the labeler can choose either top-down or bottom-up way by using decreasing or increasing value, depending on the labeler's need. We adopt and recommend the top-down approach, because coarser blocks will be extracted in the beginning so that the labeler can discern which segments are effective, related to outcome, or just a trial. If he/she just wants to retrieve the effective steps to check the smoothness of operation without labeling the function, the labeler can set a smaller value (default being half of previous threshold) to further process the blocks.

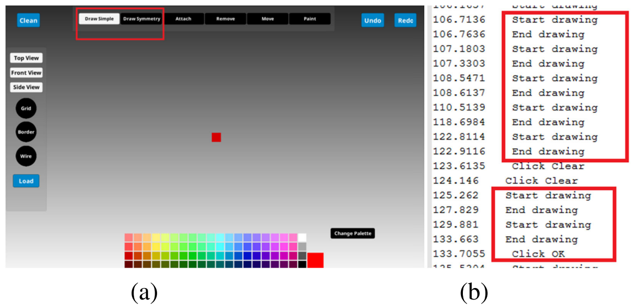


FIGURE 7. User interface and log of Qmodel Creator. (a) Interface; red square indicates two intuitive modeling functions: Draw Simple and Draw Symmetry. (b) Snapshot of Qmodel Creator log file. Red frame indicates intuitive operations.

3) INDICATORS OF USER BEHAVIOR

We propose the following indicators derived from operation logs or recordings.

a: MEAN AND STANDARD DEVIATION OF STEP PERIOD

In the log files or videos, each operation has a timestamp. Therefore, we can compute the period from the previous operation to the next, thus fully encoding the student's situation. If the student leaves and causes interruption in his/her operation, the mean and standard deviation will be larger.

b: EFFECTIVE OPERATING PERIOD (EOP)

We estimate an effective operating period that excludes daze, idle, or disturbed periods, with the threshold set at 5 seconds.

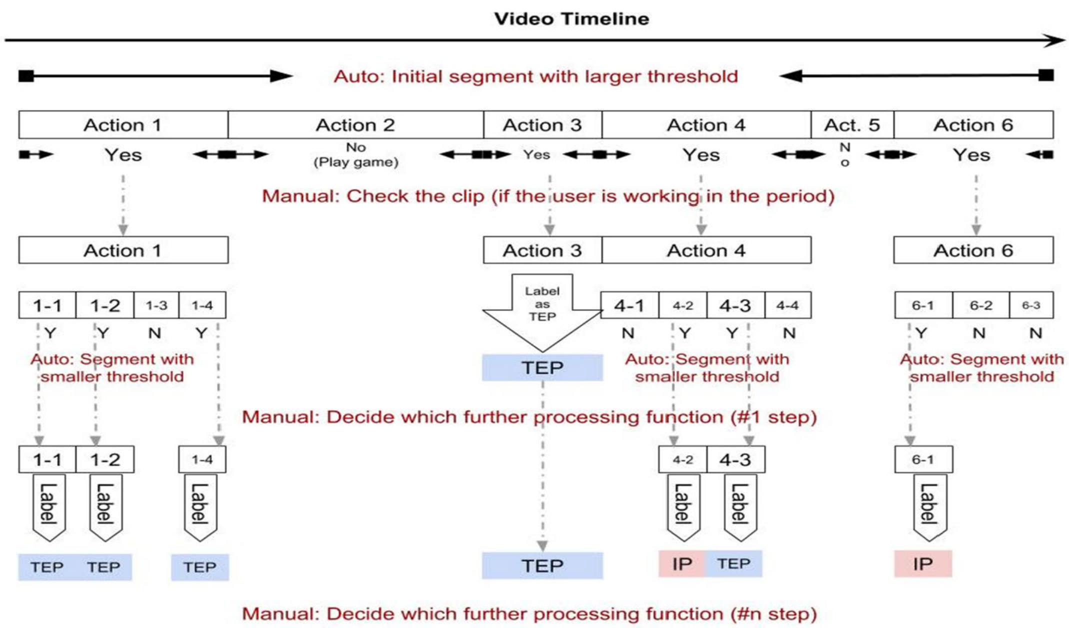


FIGURE 8. Semi-automatic procedure of usage pattern extraction from operation recordings. In the manual steps, the labeler should decide whether to process the clip further or just to label it.

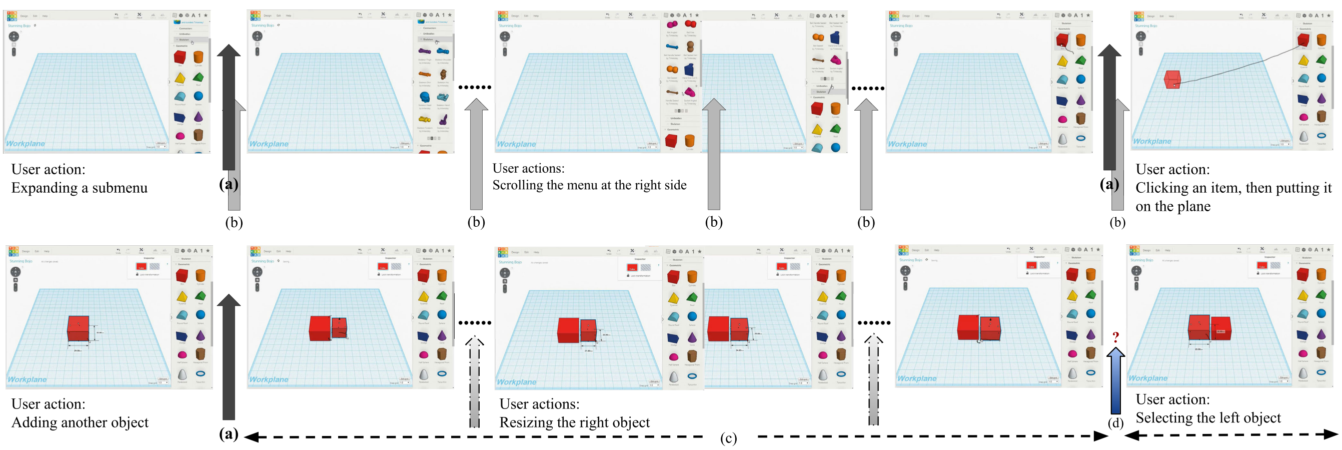


FIGURE 9. Extracting usage patterns of Tinkercad from screen recordings. (a) Segmentation with a larger threshold (b) Segmentation with a smaller threshold (c) A sequence of the same actions, which can be further decomposed by adopting a smaller value than b). (d) There is a slight difference between frames switching from c) to d). The labeler can confirm the new action extracted or not, by just decrementing the threshold to be lower than c).

c: Trial-and-Error Period (TEP):

Operations in this period do not affect the final outcome. In the log files, we check for the operator "Click Clear" or "Click Undo" commands, and then we label the related operators and perform the calculation.

d: IMPLEMENTATION PERIOD (IP)

We define the duration of a set of operations that results in the creation of a model as the Implementation Period. Actually, we can define the Trial-and-Error Period, Implementation Period, and Effective Operating Period relations as Eq. 1.

$$EOP = TEP + IP + \Delta t \tag{1}$$

where Δt is the switching cost between *TEP* and *IP*.

e: MEAN AND STANDARD DEVIATION OF TEP STEP

This represents an overview of a student's operating speed during the *TEP*.

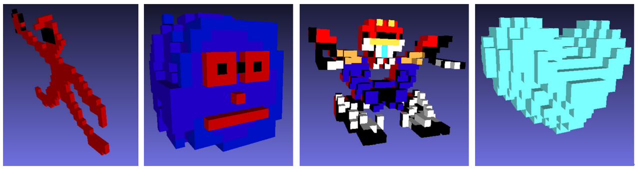
f: MEAN AND STANDARD DEVIATION OF IP STEP

This represents an overview of a student's operating speed during the *IP*.

Since we focus on the operating patterns of modeling mostly, these indicators not only provide the task interval, but also the speed and the stability of the students' work, representing the different perspectives of the learning behaviors.

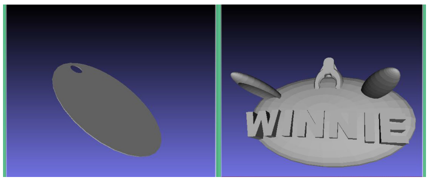
B. 3D MODEL RELATED FEATURES

Due to various conditions in K-12 classrooms, such as limited resources or vague criteria, the evaluation of modeling



(a) (b) (c) (d)

FIGURE 10. **Models created by Qmodel Creator with different degrees of detail. (a) 9.81 (b) 1.46 (c) 6.61 (d) 1.89.**



(a) (b)

FIGURE 11. ***DoD* from Tinkercad models. (a) 37.14 (unreasonable value) (b) 3.51 (reasonable value).**

production is volatile. The main objective for developing 3D model related features is to quantify the complexity of the 3D model automatically for consistent analysis. For comparison, we also collected expert evaluations of the models, as will be described in the following section.

1) DEGREE OF DETAIL (DoD)

First, we adopt a simple, geometric method inspired by [32] to measure the complexity of 3D models. The complexity of the model should be easily verified by intuition or rules. For example, the larger value indicates the more degree of complexity, or more degree of complexity represents more details with the model. Different from [32], we discovered that if the surface to volume ratio is larger, and thus a higher surface area percentage, then the model has more details. Moreover, to exclude the size factor, we multiply an approximate side length of the model. Thus, we compute the value named *Degree of Detail* according to the following formula:

$$Degree\ of\ Detail = \frac{Surface}{Volume} \times \frac{SL_{avg}}{6} \tag{2}$$

where SL_{avg} is the average side length of the bounding box of the model. Using Eq. 2, a cube model with size N will have a *DoD* equal to 1. Cube is an initial geometry without decorating any details, and its value is the lower limit of *DoD*. This result is in line with our expectations.

$$Cube\ DoD = (6 \times N^2/N^3) \times (N/6) = 1 \tag{3}$$

Fig. 10 presents some models created using Qmodel Creator and their corresponding *DoD*s. It works reasonably well on Qmodel Creator models. However, we detected an absurd condition with models created in other software, such as Tinkercad [17]. A flat, simple model such as the one shown in Fig. 11(a) will have a high *DoD* because the average side length is used as one of the parameters.

To extract more reliable structural information from 3D models, we combine [35] and [36] on quantifying the complexity of an entity. The measures we developed for analyzing 3D models examined the shape complexity from different perspectives, as described in the next subsection.

2) MIXED MODEL FEATURES

To better assess quality and complexity, we extract five features for model representation as follows: number of connected components, largest component ratio, shape complexity, partition complexity, and block-ratio complexity, as shown in Fig. 12.

a: NUMBER OF CONNECTED COMPONENTS

First of all, we can tell whether a model is structurally sound by computing the number of connected components in the model. A model with a large number of components will have a higher risk of failure during printing and result in waste of materials. Fig. 13 demonstrates two models with different numbers of connected components, representing the tendency of making errors respectively.

b: LARGEST COMPONENT RATIO

Even though students often create multiple components in a model, there exist differences in their performance. We choose the component with the largest number of facets and compute the largest component ratio (*LCR*) by Eq. 4. A larger ratio indicates a higher degree of completion, as depicted in Fig. 14.

$$LCR = \frac{\text{Number of largest component faces}}{\text{Number of total faces}} \tag{4}$$

c: MODEL COMPLEXITY

According to [36], complexity is directly related to entropy, based on information theory. However, if we intend to use 2D methods in 3D space, the issue we might encounter is the difference in their representation. For example, a 2D image is divided into pixels as a unit, and each position of a pixel is a discrete integer, while a 3D model consists of vertices and faces, and that position measure is continuous.

After the largest component has been retrieved, we obtain the bounding box of the component and then divide each axis into 64 equal parts as the units of the axis. Moreover, we introduce Fiolka *et al.*'s method [35] to calculate normal vector entropy instead of intensity entropy. We generate N vectors as histogram bins by dividing equidistant azimuth angles on a uniform sphere. We then obtain the smallest inner product of normal vectors of the model and generated vectors to build a histogram and compute the normal entropy, as illustrated in Fig. 15(a). Fig. 15(b) visualizes all normal vectors on each vertex of this model. In our implementation, the equidistant azimuth angle is set to 10 degrees.

We carry out the partition algorithm based on normalized entropy calculation, as given in Eq. 5. First, we segment the component into homogeneous regions with similar entropy

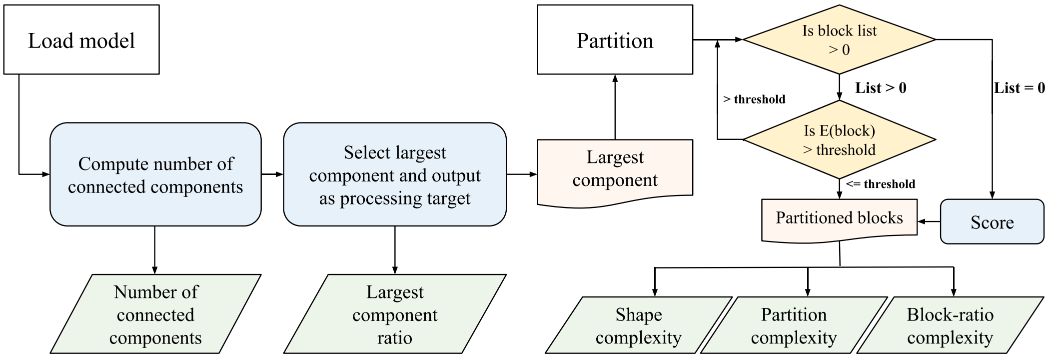
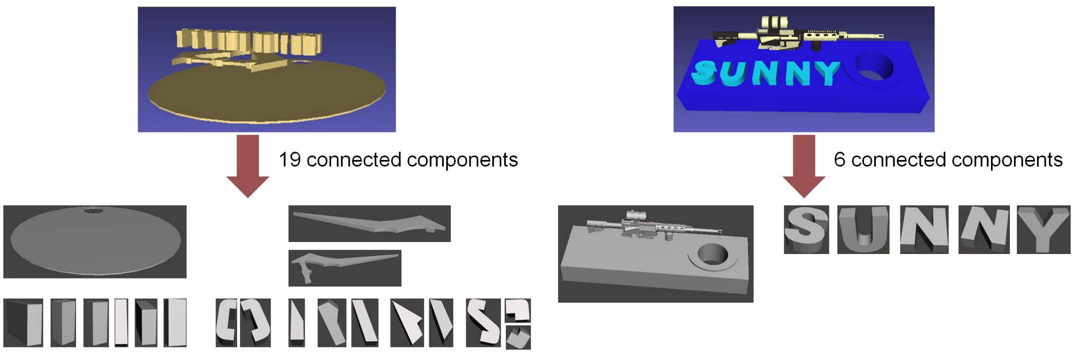
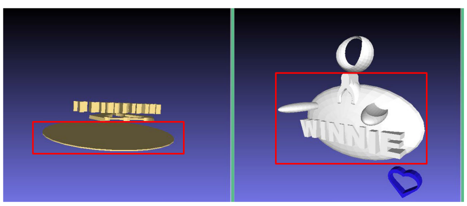


FIGURE 12. **Workflow of model analysis.**



(a) (b)

FIGURE 13. **Results of number of connected components of two Tinkercad models. (a) 19 (b) 6.**

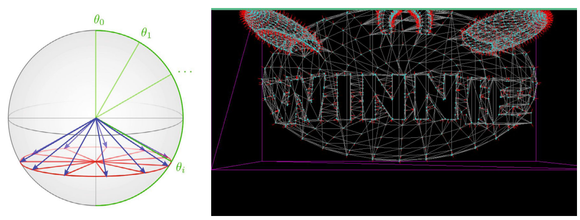


(a) (b)

FIGURE 14. **Results of largest component ratio (in red rectangle) of two Tinkercad models. (a) 0.298 (b) 0.837.**

values. The partition can be applied according to the directions of the three axes. We iterate through each unit of axes and compute the corresponding normalized entropy. Each normalized entropy is used to determine whether the partition should be continued. If the minimum of the sum of the entropy of two partitioned blocks is smaller than a threshold, the segmentation process stops. In this paper, the threshold is set to 0.1. Otherwise, the partition proceeds according to Eq. 6.

$$\begin{aligned}\text{Entropy} &= -\sum_i p_i log_2 p_i \\ \text{Maximum entropy}(bins) & \\ &= -log_2 \frac{1}{bins} \\ NE &= \frac{\text{Entropy}}{\text{Maximum entropy}}\end{aligned} \tag{5}$$



(a) (b)

FIGURE 15. **Normal vector entropy calculation using Fiolka *et al.*'s method [35]. (a) Sphere is equally divided into N vectors, labeled as blue arrows, for building a histogram (b) The red parts are the normal vectors on vertices.**

$$\begin{aligned}E_x &= Argmin(NE(Part(X(0,i), Y(0,Max), Z(0,Max)) \\ &\quad + NE(Part(X(i,Max), Y(0,Max), Z(0,Max))) \\ E_y &= Argmin(NE(Part(X(0,Max), Y(0,j), Z(0,Max)) \\ &\quad + NE(Part(X(0,Max), Y(j,Max), Z(0,Max))) \\ E_z &= Argmin(NE(Part(X(0,Max), Y(0,Max), Z(0,k)) \\ &\quad + NE(Part(X(0,Max), Y(0,Max), Z(k,Max)) \\ Block&_{part\ decision} \\ &= Argmin(E_x, E_y, E_z) \\ &\quad \forall\, i,j,k = 1 \sim (Max-1)\end{aligned} \tag{6}$$

where Max is number of division in each axis, as 64 in the paper. E_x, E_y, E_z are the optimal results of partition according to axis-x, y and z respectively, and we can therefore obtain $Block_{part\ decision}$ as the two partitioned blocks.

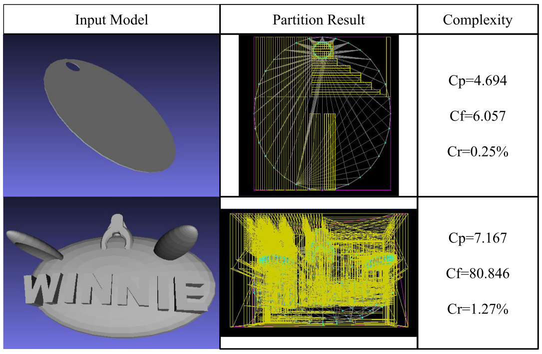


FIGURE 16. Complexity measures of the two models.

(a) (b)

FIGURE 17. Two versions of the expert evaluation interface for the two stages respectively. (a) Stage 1: from 0-1 point by each expert (b) Stage 2: from 0-2 points by each expert.

Once the partition is done, the complexity score can be computed accordingly. The first complexity measure, denoted as partition-based complexity, is directly related to the distribution of the volume of the partitioned blocks and can be computed using Eq. 7. If the volumes of the partitioned blocks are diverse, C_p of the model will be higher.

$$C_p = \text{Entropy}(Volume(Block_i)) \tag{7}$$

The next feature, shape complexity, is simply summed and denoted as C_f, as expressed in Eq. 8.

$$C_f = \sum_i NE(Block_i) \tag{8}$$

The total number of partitions also reflects the complexity of a model. Thus, we define the feature block-ratio, which is denoted as C_r and can be computed using Eq. 9.

$$C_r = \frac{R}{N}, \text{where } R\text{=Total number of partitions, } N = 64^3 \tag{9}$$

Fig. 16 shows the three complexity measures of the two 3D models. We can observe that different complexity indicators represent different characteristics of a model.

IV. THE EXPERIMENTAL DESIGN

As mentioned above, we designed two experimental stages, each with their respective mission. In this section, the details of the two stages, including dataset collection, expert evaluation, and extra field observation, are presented and the data thus gathered are employed for subsequent qualitative evaluation and quantitative analysis.

A. THE FIRST STAGE-QMODEL CREATOR WORKSHOPS

We hosted two workshops to introduce Qmodel Creator to K-12 students and collected user logs for the usability analysis of Qmodel Creator, as well as investigating the learning patterns of modeling. The finished models and operating records were used as our dataset.

1) DATASET COLLECTION

a: LANYU PRIMARY AND JUNIOR HIGH SCHOOL

The experiment involved 62 students aged from 9 to 15 years and was separated into three sessions. Of the 62 students, 49 used Qmodel Creator on an iPad, and 13 did so using an Android tablet. To motivate students and increase interest, we prepared 3D printed models as gifts for active participants. None of the students had prior experience with 3D modeling software.

b: SANCHONG HIGH SCHOOL

The experiment involved 8 students aged from 16 to 17 years. All the students used Qmodel Creator on an iPad. These students had solid training in the arts. They had been instructed in creating 3D models using Tinkercad and Sculptris. However, this experiment was their first exposure to Qmodel Creator.

2) EXPERT EVALUATION

To facilitate analysis and comparison of the model features, we need a model evaluation system informed by expert opinion. To judge the quality of 3D models in the two workshops, we designed a web interface and asked three experts (our researchers in the study) to label the models as bad or not. If more than half of the votes received by a model were ''bad'' votes, it was classified as a bad model. If a model was judged as bad, the students modeling skill was deemed to be at the novice level. Fig. 17(a) demonstrates the evaluation interface of the first stage. The degree of mutual agreement of expert evaluations of 70 models in the dataset was calculated using Eq. 10, and the reliability was 0.914.

$$\begin{aligned} DMA_{AB} &= \frac{2 \times M_{AB}}{N_A + N_B} \\ Reliability &= \frac{N \times DMA_{average}}{1 + (N-1) \times DMA_{average}} \end{aligned} \tag{10}$$

where DMA is the degree of mutual agreement, M_{AB} is number of agreements by both A and B, N_i is the number of i that should agree, and N is the total number of participating experts.

B. THE SECOND STAGE-INTEGRATED 3D PRINTING COURSES

We collaborated with Affiliated Experimental Elementary School (AEES) of National Chengchi University to implement our lesson plan for teaching modeling software and 3D printing for one semester. The details are presented in Table 1.

TABLE 1. 90-Minute lesson plan incorporated with courses in AEES.

Week	Software	Task	Num. of models
#1	Tinkercad	Key tag	14
#2	123 Design	Box	14
#3	Tinkercad	Stationery box	14
#4	Qmodel Creator	Animal and its limbs	30
#5	Qmodel Creator	Bird and its wings	29

1) DATASET COLLECTION

These experiments involved 15 students (eight girls and seven boys) aged 11 and 12 years. At the end of the semester, they were to print their designs out for presentation on stage. We obtained screen recordings (for Tinkercad and 123 Design) and log files (for Qmodel Creator) during all the courses.

2) EXPERT EVALUATION

To assess the quality of the 3D models objectively, we coordinated five researchers of our 3D printing project as experts to evaluate the models according to the following criteria:

- The model is beautiful and can be printed successfully,
- The model is ordinary but can be printed successfully, and
- The model looks bad or is impossible to print.

Fig. 17(b) also demonstrates the evaluation interface of the second stage. The degree of mutual agreement of expert evaluations of 101 models in the dataset was also calculated using Eq. 10, and reliability was 0.818.

3) INTERVIEW WITH STUDENTS

For identifying the key factors affecting students' learning experiences and performance, we also designed a face-to-face interview with students. During the interview, we mainly focused on the following four issues for qualitative analysis,

- Q.1: What modeling software do you prefer to use?
- Q.2: Are you good at drawing?
- Q.3: Do you like art?
- Students were asked to draw a picture of a river, a tree and a house with a pen on a paper in five minutes, as depicted in Fig. 18. We also had five experts judge these hand drawings and assign them into three classes.

To sum up, the proposed indicators could be shared; however, the dataset, the scale of expert evaluation, and the experimental objectives were quite different between the two stages. Therefore, the analytic results for these stages will be presented and discussed in separate sessions.

V. RESULTS OF THE FIRST STAGE-QMODEL CREATOR WORKSHOPS

In this session, first, a statistical overview will be provided to examine our skill acquisition assumption. In addition, the usability of Qmodel Creator and learning patterns of K-12 students are specifically evaluated with the following questions:

1) How long does the user take to finish a modeling task?
2) Which function is used more frequently: intuitive modeling or traditional 3D editing (such as adding and deleting voxels)?
3) When users create models, is the process smooth? Is trial-and-error required?
4) For models created with Qmodel Creator, what is the degree of completion?

In the analysis process, F-test and T-test are both adopted to examine the usage pattern and model related indicators.

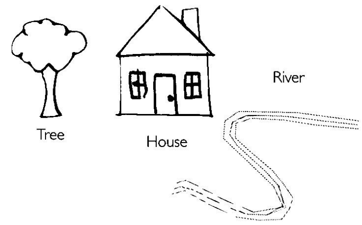


FIGURE 18. A hand-drawing example.

A. OVERALL DISTRIBUTION

Table 2 presents an overview of the dataset of the first stage. As aforementioned, *TEP*, *IP*, and *EOP* are the indicators of usage patterns; *DoD* is the model related indicator; and *Score* (ranging from 0 to 3) is the evaluation results of the models from our three experts.

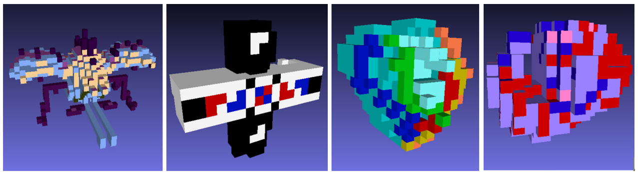
TABLE 2. Statistics of the first stage.

	Lanyu		Sanchong		Total	
	μ	σ	μ	σ	μ	σ
TEP (s)	289.622	314.565	188.387	196.741	278.052	305.129
IP (s)	224.026	215.568	165.496	157.131	217.337	210.54
EOP (s)	522.398	392.466	359.848	259.796	503.82	383.164
Score(0-3)	2	1.136	2.75	0.661	2.086	1.118
DoD	3.054	1.423	3.687	1.566	3.126	1.454
N	62		8		70	

To examine the skill acquisition assumption in Section I, we divided the dataset into different skill groups according to the experts' evaluations. Based on the scores in Table 2, we set the threshold as 2. That is, if a model received less than 2 points, it was judged as a bad model. When their models were classified as bad, the students' modeling skills were deemed to be at novice level. Fig. 19 illustrates some examples based on the evaluation results.

Of the 62 students from Lanyu Primary and Junior High School, 20 were novices, while the other 42 had advanced skill. Of the 8 students from Sanchong High School, only 1 was categorized as a novice. The other 7 were categorized as advanced. Statistics of the further categorized results are summarized in Tables 11 and 12 in Appendix VIII.

As an alternative perspective, we also calculated the correlation coefficients with evaluated score of a 3D model, including indicators of learning pattern and model feature,



(a) (b) (c) (d)

FIGURE 19. Examples of evaluation. According the votes by experts, we classified (a)(b) as the advanced group, and (c)(d) as the novice group. (a) 3 points (b) 2 points (c) 1 point (d) 0 points.

TABLE 3. Correlation with scores.

	Total	Lanyu			Sanchong
		Total	H.G.[1]	L.G.[2]	
DoD	0.24	0.224	0.122	0.257	0.184
TEP	0.051	0.065	0.015	0.014	0.298
IP	0.149	0.164	-0.133	-0.098	0.348
EOP	0.128	0.149	-0.067	-0.021	0.424
Mean of *TEP* step	-0.006	-0.048	-0.059	0.087	0.403
Mean of *IP* step	0.108	0.09	-0.127	0.294	0.121
Mean of *EOP* step	0.191	0.212	-0.002	0.362	-0.005
N	70	62	42	20	8

corr. $\leq$ -0.5 : ; -0.5 $<$ corr. $\leq$ -0.4 :
-0.4 $<$ corr. $\leq$ -0.3 : ; -0.3 $<$ corr. $\leq$ -0.2 :
0.2 $\leq$ corr. $<$ 0.3 : ; 0.3 $\leq$ corr. $<$ 0.4 :
0.4 $\leq$ corr. $<$ 0.5 : ; 0.5 $\leq$ corr. :

***Abbreviation**
1. H.G.: High-score Group
2. L.G.: Low-score Group

as shown in Table 3. Firstly, we could observe that *DoD* is weakly correlated with the score. Moreover, in terms of learning patterns, the *EOP* speed of the low-score group of Lanyu exhibit a higher degree of positive correlation, representing that the lower the score, the faster the student's operation. At last, in Sanchong group, there is a strong positive correlation in the *TEP*, *IP*, *EOP*, and the operation speed of *TEP*, indicating that the slower the operation during the trial-and-error period, the higher the score. Due to the small data size of Sanchong group, the correlation result is slightly different from that of Lanyu group. We will conduct further analysis in next subsections.

B. HOW LONG DOES THE USER TAKE TO FINISH A MODELING TASK?

Performance was evaluated with *TEP*, *IP*, and *EOP*. In these two datasets, the longest *EOP* was 1,748 seconds (nearly 30 minutes), and the shortest *EOP* was 14 seconds. Detailed statistics are provided in Tables 11 and 12, from which we discovered that in all three sessions, the advanced group from Lanyu students spent slightly more time to create their models as compared to the other groups. Table 4 lists the F-test and T-test results of operation periods from three user groups: the novice group from Lanyu, the advanced group from Lanyu, and the advanced group from Sanchong. The results show that only the advanced group from Lanyu and the novice group from Lanyu had an almost significant *IP* gap.

TABLE 4. F and T test results of operation indicators between three groups of Lanyu and Sanchong datasets.

	Lanyu Advanced vs. Lanyu Novice		Lanyu Advanced vs. Sanchong Advanced		Lanyu Novice vs. Sanchong Advanced	
	F	T *(2-tails)*	F	T *(2-tails)*	F	T *(2-tails)*
TEP	0.669	0.614	0.3	0.474	0.44	0.693
IP	0.014*	0.031	0.396	0.448	0.478	0.597
EOP	0.692	0.146	0.327	0.285	0.464	0.924

From the above results, we can conclude that the operation of the advanced group from Lanyu was significantly slower than that of the novice group from Lanyu, suggesting that advanced group from Lanyu was more careful in creating their models. Furthermore, comparing datasets from Lanyu and Sanchong, the three periods (*TEP, IP* and *EOP*) exhibited no detectable differences. Consequently, we believe that Qmodel Creator would provide an easy start on 3D model creation for K-12 students of all ages.

C. WHICH IS USED MORE FREQUENTLY: INTUITIVE MODELING OR TRADITIONAL 3D EDITING?

One of the objectives of usability analysis is to determine whether the K-12 students preferred the intuitive modeling function. By analyzing the commands recorded in the log of Qmodel Creator, the operation indicators and timespan for intuitive modeling could be calculated. As the time required for each operation was different, operation counting was not employed. Instead, we adopted the operation time span and observed the ratio of operation time span to total time span. The two groups of Lanyu Primary and Junior High School and Sanchong High School, as well as comparisons of the ratio of those who used the intuitive modeling functions, are also recorded in Tables 11 and 12 (in the column *I-R*).

The results indicated that not all the students used the intuitive modeling function frequently, so traditional 3D editing functions are still necessary. However, when comparing *TEP* and *IP* in the Lanyu and advanced Sanchong datasets, we discovered that a considerable proportion of the users reduced their use of intuitive operation during *IP*, as illustrated in Table 5.

TABLE 5. Change in the ratio of intuitive operation between ***TEP*** and ***IP*** in Lanyu and Sanchong datasets.

Trend	Lanyu		Sanchong	
	Advanced	**Novice**	**Advanced**	**Novice**
Decrease	31	13	5	0
Increase	5	5	2	1
No *TEP*	6	2	0	0

Although they used intuitive modeling less in the Implementation Period, the average ratios of intuitive operation of *IP* in both groups of Lanyu students were higher than 25%, while the average ratio of intuitive operation of *IP* in the advanced group of the Sanchong dataset was almost 40%, as also can be observed in the Appendix VIII. Therefore, high school students who had used other modeling software and understood basic 3D spatial concepts used intuitive modeling functions more frequently, while primary and junior high

TABLE 6. Results of significant change in operation speed in Lanyu and Sanchong datasets.

Trend	Lanyu				Sanchong		
	Advanced		Novice		Advanced		Novice
	Count	Ratio	Count	Ratio	Count	Ratio	Count
Speed Up	13	30.952% (13/42)	10	50% (10/20)	2	28.571% (2/7)	0
Slow Down	4	9.524% (4/42)	2	10% (2/20)	2	28.571% (2/7)	0

$\alpha = 0.05$

TABLE 7. The cases of no Trial-and-Error in Lanyu dataset.

Advanced Group			Novice Group		
DoD	*IP* (s)	**Remove count**	*DoD*	*IP* (s)	**Remove count**
1.459	860.703	485	2.356	40.452	0
1.507	249.554	108	2.517	39.864	0
1.976	214.154	63			
2.328	72.6	0			
3.152	14.378	0			
6.045	439.217	7			

TABLE 8. Tests for detail of degree.

Dataset	Advanced		Novice		Tests	
	μ	σ	μ	σ	**F-test**	**T-test**
Lanyu	3.228	1.541	2.688	1.045	0.4*	0.059
Sanchong	3.796	1.646	2.922		Not available	Not available

school students who were creating 3D models for the first time preferred traditional 3D editing functions.

Regarding this phenomenon, we propose a possible explanation as follows. Intuitive operations require users to understand some 3D concepts beforehand in order to gain the confidence to create models in accordance with their expectations. Because the Lanyu students were willing to try out unfamiliar functions, the average ratios of intuitive operations of *TEP* were higher than 40% in both groups. However, the main purpose of this event was to submit a finished model, so the primary and junior high school students preferred the complicated but controllable operations of adding and removing voxels incrementally.

D. WHEN USERS CREATE MODELS, IS THE PROCESS SMOOTH? IS TRIAL-AND-ERROR REQUIRED?

We calculated the means and standard deviations of the *TEP* and the *IP*. Because each series of *TEP* and *IP* was retrieved from the same student, they should be considered as two distributions having the same variation. Thus, we adopted the 2-tailed T test to examine if the operation speeds were consistent during *TEP* and *IP*.

The 2-tailed T test results confirmed that the means of the *TEP* and *IP* steps were unequal, as shown in Table 6. The test results also indicated the direction of shifting means in the same tables. If the mean of the *IP* step was sufficiently less than the mean of the *TEP* step, then we call this case ''speed up'', and vice versa. In Table 6, we found that users in the novice group in the Lanyu dataset tended to speed up in the Implementation Period. This could have been due to external reward expectations or deadline pressure to complete the model-making quickly. It is suggested that these factors should be considered in the design of lecture plans for 3D modeling.

In Table 5, it is worth noting that a few users from Lanyu skipped Trial-and-Error. All cases are listed in Table 7. We also observed that some students had their own patterns of Trial-and-Error, such as using ''remove voxels'' to delete content entirely. The operation counts are also included in Table 7. After excluding these cases, the remaining cases contained models of relatively low degrees of detail and low Implementation Periods (less than 90 seconds).

Here, we offer several possible reasons for why the students skipped Trial-and-Error, irrespective of whether they were classified as novices or advanced users:

1) They were not familiar with the *Clear* and *Undo* functions, so they used other buttons (e.g., *Remove*) to purge their trial model(s). Such cases, however, still belonged to the Trial-and-Error Period.
2) They just wanted to finish the model quickly (e.g., to get a reward).
3) They had previously used similar 3D modeling applications or advanced 2D graphics software.

According to the above results, we believe that users need to be given sufficient time for trial-and-error. The finished models will contain more details, which means that users will have better achievements in the learning process.

E. WHAT IS THE DEGREE OF COMPLETION?

The proportion of the advanced group in Lanyu dataset was 67.7%, while that of the advanced group in the Sanchong dataset was 87.5%. The degree of achievement using Qmodel Creator in the two datasets was specified. We then compared these two groups (novice and advanced) using the Degree of Detail of the models, as shown in Table 8. After analyzing the Lanyu dataset with the F test and 1-tailed T test, we found that the models created by the advanced group possessed higher degrees of detail.

F. SUMMARY OF THE STAGE

We investigated the usage patterns of a 3D modeling software program to understand user behavior and requirements. Operation logs of Qmodel Creator were recorded and analyzed. The characteristics of different user groups were observed using quantitative measures derived from the log files. The results showed no significant difference in operation period between students of Lanyu Primary and Junior High School and Sanchong High School, making this software an easy-to-use tool for all K-12 students.

According to the user behavior analysis, the intuitive modeling function of Qmodel Creator greatly eases the learning curve for K-12 children. If complemented by suitable lesson plans, this software could be widely adopted by elementary school students. The quantitative indicators developed in this stage could be applied to the evaluation of other 3D modeling software and serve as a reference for designing lecture cases.

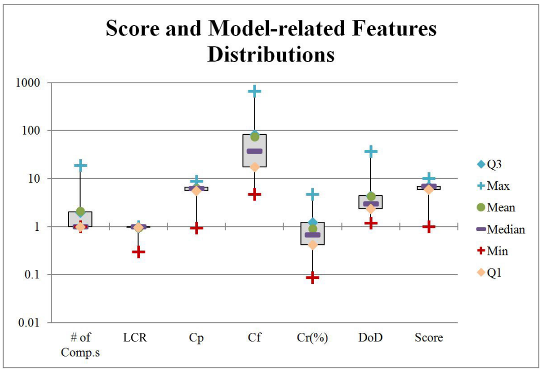


FIGURE 20. Boxplot of distribution of model features and scores.

VI. RESULTS OF THE SECOND STAGE-INTEGRATED 3D PRINTING COURSES

To begin with, an overall analysis of models was conducted. We investigated which factors were related to the scores of the models. Because we found the variance of the effective operating period between users as too large (as shown in Table 13), we considered only model-related features and qualitative factors of interviews.

In addition, we conducted statistical tests on the high-score and low-score groups (according to the results of the scores in Table 13, we set the threshold as 7) for the features proposed in this paper, to examine the skill acquisition assumption as same as the first stage.

Finally, based on usage pattern metrics, we roughly categorized users into the following four categories so that the eligible cases in the dataset could be compared using model-related features as well as usage-pattern-related features:

1) Novice with little interest: Mean of trial-and-error period (*TEP*) steps is large, and mean of implementation period (*IP*) steps is also large.
2) Novice with caution: Mean of *TEP* steps is small, but mean of *IP* steps is large.
3) Intermediate: Mean of *TEP* steps is large, but mean of *IP* steps is small.
4) Advanced: Mean of *TEP* steps and mean of *IP* steps are both small.

In Appendix VIII, statistics of the further categorized results are summarized in Table 13. Moreover, detailed performance (such as learning patterns, model indicators, score, etc.) of the four eligible cases are also presented in Table 14. Additionally, the three-point-scale Likert results and basic information of all students (with labelled eligible cases) are shown in Table 15.

A. OVERALL DISTRIBUTION

In the beginning, we analyzed the models in the dataset and compared the results of model features as well as expert scores for each model, as shown in Fig. 20.

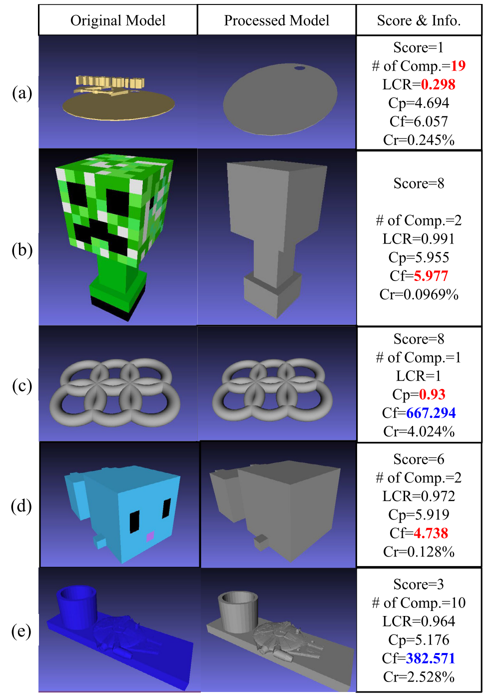


FIGURE 21. Models with extreme values.

We can observe that the indicators of C_f, C_r, and *DoD* had a larger interquartile range, which implies that these indicators had higher sensitivity to the differences between models. Moreover, we found a wide range of extreme values for each indicator. We then checked these models with extreme values. Some of them received low scores. An example is shown in Fig. 21(a). This model was not only structurally separated but also had the lowest largest component ratio (*LCR*), and the three complexity indicators were low as well.

However, an extreme value of one indicator or two does not necessarily lead to a low score. Fig. 21(b) and (c) are examples of works that received high scores in the extreme value area, whereas Fig. 21(d) and (e) are examples of the low-scoring group. We confirmed that there was still a huge range of distribution of model complexity, even in the group with similar scores. In addition, the quantitative values did indeed reflect the appearance of the model.

We can conclude that a model can be judged as having a low degree of completeness when the number of connected components is large, and the largest component ratio, C_p, C_f, and C_r are small, a rule that can be applied for fast filtering in practice.

B. CORRELATION WITH SCORE

To investigate what factors were related to the scores of the models, we conducted three correlation analyses. We computed correlation coefficients including each model feature and measure of interviews. In the second trial, we divided the model data into three groups according to the software used to create them, while in the third trial we divided data according to groups of the score. If the score was greater than or equal

TABLE 9. Correlation with scores.

	Trial #1	Trial #2			Trial #3	
	Total	Tinkercad	123D	Q.C.	H.G.	L.G.
# C.s	-0.41	**-0.671**	-0.472	-0.169	-0.153	**-0.579**
LCR	0.321	0.491	0.237	0.399	0.036	**0.577**
C_p	0.169	0.46	-0.354	0.06	0.382	0.213
C_f	0.003	-0.24	0.076	0.095	-0.218	-0.163
C_r	0.196	0.277	0.006	0.022	0.23	-0.084
DoD	-0.112	-0.385	-0.101	0.079	0.111	-0.282
Q.1	0.226	0.274	N.A.	0.109	0.194	0.224
Q.2	-0.076	-0.058	N.A.	-0.094	-0.006	0.17
Q.3	0.052	0.237	N.A.	-0.069	0.085	-0.286
S.H.D.	-0.084	0.094	N.A.	-0.197	0.144	-0.061

corr. $\leq$ -0.5 : -0.5 $<$ corr. $\leq$ -0.4 :
-0.4 $<$ corr. $\leq$ -0.3 : -0.3 $<$ corr. $\leq$ -0.2 :
0.2 $\leq$ corr. $<$ 0.3 : 0.3 $\leq$ corr. $<$ 0.4 :
0.4 $\leq$ corr. $<$ 0.5 : 0.5 $\leq$ corr. :

to 7, the work was assigned to the high-score group, and vice versa. The results are summarized in Table 9.

As labeled in Table 9, the number of connected components was negatively correlated with the scores, and the largest component ratio of a model was positively correlated with scores. But the correlation in the high-score group was not the same as that in the low-score group. A plausible explanation is that a low-score model usually cannot be printed. If a model has been ruled out of key factors that lead to failed printing, such as controlling the number of connected components to a minimum, the model will obtain a higher score.

In addition, the aforementioned widely-distributed C_f, C_r, and *DoD* showed an inconsistent degree of correlation in Trial 2. This inconsistency may have been related to the functions and properties of the modeling software. In Qmodel Creator, for example, the models made with this software are cubic style, so it is difficult to create as much detail as could be created in other software within the same period. This difference is reflected in the lower correlation with the score in the group of Qmodel Creator.

Question 1 of the interview, ''What modeling software do you prefer to use?'', yielded weak correlation results in all trials. After excluding the possibility of bias factors in interviews, we think that students' preferences and motivation are, at least most of the time, the keys to learning.

C. DIFFERENCES BETWEEN GROUPS OF SCORES

Because we divided works according to their evaluated scores, we wondered if there were any differences in the proposed features between the two groups. The results of the tests are listed in Table 10.

TABLE 10. Tests for high-score and low-score groups.

	High scores		Low scores		Tests	
	μ	σ	μ	σ	F-test	T-test
C_p	6.165	1.332	6.187	0.86	2.69E-03*	0.92
C_f	89.681	128.118	57.545	77.025	5.24E-04*	0.128
C_r	0.011	9.78E-03	7.56E-03	5.06E-03	9.24E-06*	**0.0497***
DoD	4.398	2.922	4.287	5.833	2.66E-06*	0.904

As labeled in the table, C_r had a significant difference in the T-test. C_f of the two groups had a great gap, but it was not significant. Because in the previous summary, C_r and C_f were both sensitive to the difference between models, we can conclude that the two features are appropriate for representing model complexity.

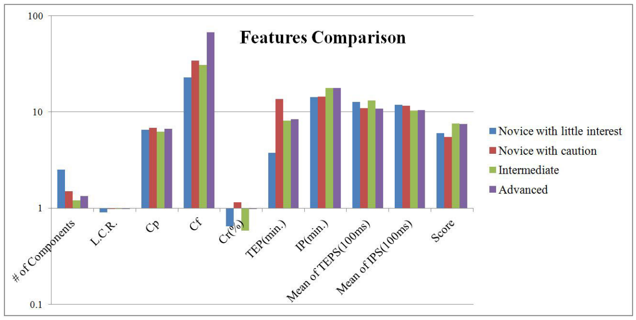


FIGURE 22. Comparison among four user categories.

D. USER CATEGORIES COMPARISON

According to the definition of the 4 user categories, we identified one eligible case for each category and retrieved models and usage pattern features of users, and we averaged and compared these quantitative values, as presented in Fig. 22. Further detailed results are available in Table 14 in Appendix VIII.

Mean of *TEPS* and mean of *IPS* were the criteria for user classification, so the two items were in line with expectations. However, averaging the results shortened the gap between them. In addition, the advanced students' models were the highest on the C_f item. It is interesting to note that the works by novices with caution had the highest C_r, while those of the intermediate students had the lowest C_r. The novices with little interest made the works with the largest number of connected components, which had the smallest ratios of largest component.

Although we used usage patterns to select students, the features of the model partially reflect its creator's characteristics. We believe that teachers can better monitor the learning status of students by inspecting both model features and usage patterns.

E. SUMMARY OF THE STAGE

We proposed measures to quantify the complexity of a 3D model that are better than those in the previous stage, as well as the generalized procedure for computing usage pattern features, to judge students' work and evaluate their performance regardless of the type of software used. In the workflow of model analysis, we defined the steps for computing model complexity, namely shape, partition, and block-ratio complexity. Meanwhile, we implemented an adjustable threshold on a hierarchical video segmentation system to make these features extractable and applicable to general cases in the work-flow of usage pattern analysis.

We confirmed that the number of connected components is negatively correlated with the evaluated score of a model, while the largest component ratio is positively correlated with the score. The two indicators can be used to automatically

TABLE 11. Statistics of Lanyu dataset.

	Advanced Group									Novice								
	Female			Male			Total			Female			Male			Total		
	μ	σ	*I-R*	μ	σ	*I-R*	μ	σ	*I-R*	μ	σ	*I-R*	μ	σ	*I-R*	μ	σ	*I-R*
TEP (s)	226.255	264.27	44.02%	374.27	356.160	39.73%	303.786	324.294	41.25%	52.962	22.805	49.68%	282.867	297.739	42.37%	259.876	290.849	43.1%
IP (s)	246.889	217.961	22.21%	269.195	252.103	31.62%	258.573	236.723	27.14%	136.636	14.692	34.07%	153.126	143.606	26.68%	151.477	136.405	27.42%
EOP (s)	480.393	299.246	31.86%	657.111	453.428	35.57%	572.96	397.65	33.8%	194.977	5.8	42.44%	440.799	370.201	34.67%	416.217	358.868	35.45%
DoD	2.574	0.863		3.822	1.764		3.228	1.541		3.024	0.169		2.65	1.094		2.688	1.045	
N	20			22			42			2			18			20		

TABLE 12. Statistics of Sanchong dataset.

	Advanced Group									Novice	
	Female			Male			Total			Female	
	μ	σ	*I-R*	μ	σ	*I-R*	μ	σ	*I-R*	*record*	*I-R*
TEP (s)	273.569	205.643	55.85%	52.87	24.635	53.34%	210.512	200.799	55.13%	33.509	63.88%
IP (s)	237.814	158.586	28.95%	57.116	24.053	62.77%	186.186	157.457	38.62%	20.669	93.87%
EOP (s)	517.696	203.25	38.12%	110.836	0.565	65.28%	401.45	251.575	45.88%	68.635	80.52%
DoD	3.591	1.893		4.307	0.389		3.796	1.646		2.922	
N	5			2			7			1	

TABLE 13. Statistics of AEES dataset.

	Tinkercad						QModel Creator						123D					
	Female		Male		Total		Female		Male		Total		Female		Male		Total	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
TEP (s)	488.66	460.32	356.58	313.23	427.7	397.32	429.4	553.3	285.76	267.16	371.39	461.67	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
IP (s)	1206.53	528.73	996.13	603.15	1109.42	562.92	705.97	515.94	691.26	500.47	700.03	504.84	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
EOP (s)	1695.19	637.24	1352.48	577.45	1537.02	623.08	1176.96	802.15	995.9	518.35	1103.84	701.36	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Step of TEP (s)	1.026	0.339	0.706	0.349	0.878	0.374	1.074	0.35	1.35	0.38	1.185	0.384	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Step of IP (s)	1.008	0.146	0.895	0.089	0.956	0.133	1.036	0.263	1.27	0.575	1.131	0.429	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Step of EOP (s)	0.994	0.128	0.884	0.089	0.943	0.123	1.059	0.25	1.228	0.21	1.127	0.247	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
# C.s	1.47	1.36	4.31	5.15	2.79	3.85	1.97	1.45	1.74	0.98	1.86	1.25	1.38	1.06	2.33	2.34	1.79	1.72
LCR	98.7%	4.23%	87.17%	21.57%	93.35%	15.82%	95.83%	9.24%	98.91%	2.52%	97.24%	7.13%	94.31%	16.08%	81.76%	29.14%	88.93%	22.53%
C_p	7.2	1.6	6.46	1.05	6.86	1.1	6.06	0.75	6.14	0.57	6.1	0.67	4.89	1.7	5.48	1.89	5.14	1.74
C_f	62.13	45.65	59.97	102.3	61.13	75.71	37.09	37.996	39.16	16.18	38.04	32.86	314.1	204.3	169.09	91.62	251.95	176.78
C_r	1.49%	1.31%	0.91%	0.72%	1.22%	1.1%	0.597%	0.398%	0.698%	0.373%	0.643%	0.387%	1.6%	1.13%	1.24%	0.58%	1.44%	0.92%
DoD	6.12	4.2	7.6	9.03	6.81	6.78	2.38	0.79	2.82	0.69	2.58	0.77	5.17	1.86	9.07	7.44	6.84	5.21
Score **(0-10)**	8	1.6	5.62	2.14	6.89	2.2	5.94	1.93	6.37	1.39	6.14	1.71	6.88	0.83	6	1.26	6.5	1.09
N	16		14		30		32		28		60		8		6		14	

detect who needs more instruction. We also conducted statistical tests on the high-score and low-score groups for the proposed features. The results showed that the two features are appropriate for representing model complexity. Finally, we divided works into four categories according to the characteristics of usage patterns, and representative data of students were retrieved and compared. The results showed that the model-related features can be partially related to the learning patterns or the characteristics of the students (the creators).

VII. LIMITATIONS

Due to the resource limitation and the challenge in collecting/processing operation logs, we can only perform case-study rather than modeling the learning patterns. Even so, the analysis shows that we can retrieve a comprehensive perspective of students' learning conditions using both model-related features and usage-pattern features.

VIII. CONCLUSION

To employ integrated STEAM courses in K-12 classrooms, we developed modeling software, Qmodel Creator, and proposed appropriate indicators for the purpose of observation and evaluation of learning behaviors as well as STEAM lesson plans. Designing suitable 3D modeling software for children is a challenging mission. Evaluation with an integrated lesson plan is another difficult task. Based on the features we propose in this paper, we built an analysis system with more automatic processes and fewer human interventions, and we also collected more data to contribute to research on learning behaviors associated with various 3D modeling software packages.

According to the above summaries of our two stages, the results were remarkable when the trial-and-error patterns was employed as an indicator for the operation sequence, suggesting that it is reasonable and significant to analyze the learning behavior of trial-and-error. We confirm that in all cases, model-related features or usage-pattern features should be considered together, as doing so could highlight the deficiencies of students. We believe that teachers can gain a comprehensive understanding of their students' learning conditions and statuses by referencing the proposed features.

Since learning behavior is a continuously interactive process between human beings and the environment, which may result in diligent or stagnant, what is the key factor? Combined with the analysis of user operation records as primary sources or field evidence, we believe, the trial-and-error approach (defined as untargeted attempts in the paper) is a kind of concrete action representing people are interested and focusing in the learning task. In the future work, we will investigate different tasks, and further verify the qualitative meaning of measuring the trial-and-error behavior at each skill level.

ACKNOWLEDGMENT

This article was presented in part at the Tenth International Conference on E-Learning, MCCSIS, Madeira, Portugal, in July 2016, and in part at the 18th IEEE International Conference on Advanced Learning Technologies (ICALT), Mumbai, India, in July 2018.

TABLE 14. **Mean performance of the four eligible cases.**

	Novice with little interest			Novice with caution			Intermediate			Advanced		
	T.C.[a]	Q.C.[b]	Total	T.C.	Q.C.	Total	T.C.	Q.C.	Total	T.C.	Q.C.	Total
TEP (s)	547.7	64.28	225.42	695.6	847.907	817.445	0	485.42	485.42	777.4	361.456	500.104
IP (s)	1052.4	754.9	854.06	1524.2	700.289	865.072	2268	764.253	1065	1608.5	782.492	1057.828
EOP (s)	1600.1	842.02	1094.71	2219.8	1621.757	1741.366	2268	1272.513	1018.178	2385.9	1211.505	1602.97
Step of TEP (s)	0.905	1.444	1.264	0.817	1.161	1.092	0	1.317	1.317	1.056	1.093	1.08
Step of IP (s)	0.969	1.299	1.189	0.93	1.207	1.151	0.836	1.083	1.034	0.978	1.068	1.038
Step of EOP (s)	0.94	1.283	1.169	0.891	1.257	1.184	0.836	1.145	1.083	1	1.078	1.052
# C.s	3.5	2	2.5	1	1.75	1.5	1	1.25	1.2	2	1	1.333
LCR	78.597%	96.258%	90.371%	100%	98.567%	99.045%	100%	99.765%	99.812%	91.841%	100%	97.28%
C_p	7.233	6.097	6.476	7.212	6.656	6.842	7.243	5.993	6.243	7.635	6.181	6.665
C_f	18.045	25.173	22.797	68.313	16.684	33.894	22.259	32.765	30.664	63.256	69.278	67.27
C_r	0.921%	0.511%	0.648%	2.666%	0.382%	1.144%	0.595%	0.584%	0.586%	0.984%	0.983%	0.983%
DoD	5.158	2.481	3.373	7.577	1.906	3.796	4.276	3.096	3.332	3.699	2.7	3.033
Q.1	Tinkercad			Tinkercad			123D			Tinkercad		
Q.2	1			2			2			2		
Q.3	1			3			3			3		
S.H.D.	2			3			2.5			1.5		
Score	6.5	5.75	6	7.5	4.5	5.5	7	7.75	7.6	9	6.75	7.5

[a]Tinkercad
[b]Qmodel Creator

TABLE 15. **Participants' profiling and questionnaire(Three-Point scales).**

ID	Gender	Age	Age Start Using Computer	Computer Pref.	Modeling Software Preference[1]	Teaching Style Preference	Gaming Pref.	Art Pref.[3]	Painting Talent (self evaluation)[2]	Hand Drawing[4]
00129	Male	11	7	3	Tinkercad	Free-Style	3	3	2	1
00130	Female	12	5	3	Tinkercad	Free-Style	2	3	3	2
00131	Male	11	7	2	123D Design	Free-Style	2	3	1	2
00132	Female	12	8	3	QModel	Free-Style	3	3	1	1
00133[a]	Male	11	6	3	123D Design	Free-Style	3	3	2	2.5
00134[b]	Female	11	7	3	Tinkercad	Topic	1	3	1	1.5
00135	Male	11	8	3	QModel	Topic	2	2	1	2
00136[c]	Female	11	8	2	Tinkercad	Topic	1	3	2	3
00139	Male	11	5	3	Tinkercad	Topic	3	3	2	3
00140	Female	12	8	2	Tinkercad	Topic	2	2	1	1
00141[d]	Male	11	6	3	Tinkercad	Free-Style	3	1	1	2
00142	Female	11	7	3	QModel	Free-Style	3	3	2	2.5
00143	Male	11	4	3	QModel	Free-Style	3	1	1	2
00144	Female	11	10	2	Tinkercad	Topic	1	2	2	2.5
00145	Female	11	8	3	Tinkercad	Topic	1	3	2	1

The analyzed items in the work:

1: What modeling software do you prefer to use?
2: Are you good at drawing?
3: Do you like art?
4: Students were asked to draw a picture of a river, a tree and a house with a pen on a paper in five minutes.

The students' usage patterns have been analyzed as:

[a]Intermediate
[b]Advanced
[c]Novice with caution
[d]Novice with little interest

STATISTICS OF *TEP*, *IP*, *EOP*, *DoD*, AND INTUITIVE OPERATION RATIO IN THE FIRST STAGE

See Tables 11 and 12.

STATISTICS OF FEATURES AND FACE-TO-FACE INTERVIEW IN THE SECOND STAGE

See Tables 13–15.

REFERENCES

[1] S. B. Merriam and L. M. Baumgartner, *Learning in Adulthood: A Comprehensive Guide*. Hoboken, NJ, USA: Wiley, 2020.

[2] V. Marcy, "Adult learning styles: How the VARK learning style inventory can be used to improve student learning," *J. Physician Assistant Edu.*, vol. 12, no. 2, pp. 117–120, 2001.

[3] A. L. de Oliveira Pires, "Higher education and adult motivations towards lifelong learning," *Eur. J. Vocational Training*, vol. 46, no. 1, pp. 129–150, 2009.

[4] M.-H. Lin, H.-G. Chen, and K.-S. Liu, "A study of the effects of digital learning on learning motivation and learning outcome," *EURASIA J. Math., Sci. Technol. Edu.*, vol. 13, no. 7, pp. 3553–3564, Jun. 2017.

[5] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemp. Educ. Psychol.*, vol. 25, no. 1, pp. 54–67, Jan. 2000.

[6] National Research Council, *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC, USA: National Academies Press, 2000.

[7] M. Honey, G. Pearson, and H. Schweingruber, *STEM Integration in K-12 Education: Status, Prospects, and an Agenda for Research*. Washington, DC, USA: National Academies Press, 2014.
[8] G. Cesaretti, E. Dini, X. De Kestelier, V. Colla, and L. Pambaguian, ''Building components for an outpost on the lunar soil by means of a novel 3D printing technology,'' *Acta Astronautica*, vol. 93, pp. 430–450, Jan. 2014.
[9] A. Mahmoud and M. Bennett, ''Introducing 3-Dimensional printing of a human anatomic pathology specimen: Potential benefits for undergraduate and postgraduate education and anatomic pathology practice,'' *Arch. Pathol. Lab. Med.*, vol. 139, no. 8, pp. 1048–1051, Aug. 2015.
[10] D. Thornburg, N. Thornburg, and S. Armstrong, *The Invent to Learn Guide to 3D Printing in the Classroom: Recipes for Success*. Torrance, CA, USA: Constructing Modern Knowledge Press, 2014.
[11] J. Irwin, J. Pearce, G. Anzalone, and D. Oppliger, ''The RepRap 3-D printer revolution in STEM education,'' in *Proc. ASEE Annu. Conf. Expo.*, 2014, pp. 1224–1242.
[12] Y. Anzai and H. A. Simon, ''The theory of learning by doing,'' *Psychol. Rev.*, vol. 86, no. 2, p. 124, 1979.
[13] R. Maloy, S. Kommers, A. Malinowski, and I. LaRoche, ''3D modeling and printing in history/social studies classrooms: Initial lessons and insights,'' *Contemp. Issues Technol. Teacher Edu.*, vol. 17, no. 2, pp. 229–249, 2017.
[14] Y.-H. Chien and P.-Y. Chu, ''The different learning outcomes of high school and college students on a 3D-printing STEAM engineering design curriculum,'' *Int. J. Sci. Math. Edu.*, vol. 16, no. 6, pp. 1047–1064, Aug. 2018.
[15] C.-C. Hu, T.-H. Wei, Y.-S. Chen, Y.-C. Wu, and M.-T. Chi, ''Intuitive 3D cubic style modeling system,'' in *Proc. SIGGRAPH Asia Posters (SA)*, 2015, p. 27.
[16] I.-S. Lin, T.-Y. Li, F.-C. Liang, and Y.-T. Lin, ''A collaborative learning system for sharing 3D models: 3D model co-learning space,'' in *Proc. IEEE 17th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2017, pp. 502–506.
[17] Autodesk. *Tinkercad*. Accessed Apr. 29, 2018. [Online]. Available: https://www.tinkercad.com/ and https://www.tinkercad.com/
[18] A. Holzinger, ''Usability engineering methods for software developers,'' *Commun. ACM*, vol. 48, no. 1, pp. 71–74, Jan. 2005.
[19] J. D. Vermunt and V. Donche, ''A learning patterns perspective on student learning in higher education: State of the art and moving forward,'' *Educ. Psychol. Rev.*, vol. 29, no. 2, pp. 269–299, Jun. 2017.
[20] F. Coffield *et al.*, ''Learning styles and pedagogy in post-16 learning: A systematic and critical review,'' Learn. Skills Res. Centre, London, U.K., Tech. Rep. TD/TNC 79.71, 2004.
[21] C. Evans and J. D. Vermunt, ''Styles, approaches, and patterns in student learning,'' *Brit. J. Educ. Psychol.*, vol. 83, no. 2, pp. 185–195, Jun. 2013.
[22] J. D. Vermunt and Y. J. Vermetten, ''Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations,'' *Educ. Psychol. Rev.*, vol. 16, no. 4, pp. 359–384, Dec. 2004.
[23] J. Nemiro, C. Larriva, and M. Jawaharlal, ''Developing creative behavior in elementary school students with robotics,'' *J. Creative Behav.*, vol. 51, no. 1, pp. 70–90, Mar. 2017.
[24] E. A. Kirkpatrick, *Fundamentals of Child Study*. New York, NY, USA: Macmillan, 1908.
[25] C. L. Hull, ''Simple trial and error learning: A study in psychological theory,'' *Psychol. Rev.*, vol. 37, no. 3, p. 241, 1930.
[26] A. Amsel, ''Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension,'' *Psychol. Rev.*, vol. 69, no. 4, p. 306, 1962.
[27] R. R. Schmeck, ''Error-produced frustration as a factor influencing the probability of occurrence of further errors,'' *J. Exp. Psychol.*, vol. 86, no. 2, p. 153, 1970.
[28] H. L. Dreyfus, S. E. Drey-fus, and L. A. Zadeh, ''Mind over machine: The power of human intuition and expertise in the era of the computer,'' *IEEE Expert*, vol. 2, no. 2, pp. 110–111, Jun. 1987.
[29] S. E. Dreyfus and H. L. Dreyfus, ''A five-stage model of the mental activities involved in directed skill acquisition,'' Berkeley Oper. Res. Center, California Univ., Berkeley, CA, USA, Tech. Rep. ORC-80-2, 1980.
[30] H. L. Dreyfus and S. E. Dreyfus, ''The ethical implications of the five-stage skill-acquisition model,'' *Bull. Sci., Technol. Soc.*, vol. 24, no. 3, pp. 251–264, Jun. 2004.
[31] H. L. Dreyfus and S. E. Dreyfus, ''From Socrates to expert systems: The limits of calculative rationality,'' in *Philosophy and Technology II*. Dordrecht, The Netherlands: Springer, 1986, pp. 111–130.
[32] B. Valentan, T. Brajlih, I. Drstvensek, and J. Balic, ''Basic solutions on shape complexity evaluation of stl data,'' *J. Achievements Mater. Manuf. Eng.*, vol. 26, no. 1, pp. 73–80, 2008.
[33] J. Rossignac, ''Shape complexity,'' *Vis. Comput.*, vol. 21, no. 12, pp. 985–996, Dec. 2005.
[34] J. D. Gardiner, J. Behnsen, and C. A. Brassey, ''Alpha shapes: Determining 3D shape complexity across morphologically diverse structures,'' *BMC Evol. Biol.*, vol. 18, no. 1, p. 184, Dec. 2018.
[35] T. Fiolka, J. Stückler, D. A. Klein, D. Schulz, and S. Behnke, ''Sure: Surface entropy for distinctive 3D features,'' in *Proc. Int. Conf. Spatial Cognition*. Berlin, Germany: Springer, 2012, pp. 74–93.
[36] W.-H. Liao and P.-M. Chen, ''Analysis of visual elements in logo design,'' in *Proc. Int. Symp. Smart Graph.* Springer, 2014, pp. 73–85.
[37] G. Bradski, ''The OpenCV library,'' *Dr. Dobb's J. Softw. Tools*, vol. 25, no. 11, 2000, Art. no. 2236121.

YI-CHIEH WU received the B.E. degree in psychology from National Taiwan University, in 1999, and the M.S. degree in computer science from National Chengchi University, Taiwan, in 2014, where she is currently pursuing the Ph.D. degree. She was with Industry as a Software Engineer for over 16 years. Her research interests include user interaction analysis, computer vision, and human-computer interfaces.

WEN-HUNG LIAO (Member, IEEE) received the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, The University of Texas at Austin, in 1991 and 1996, respectively. He has been with National Chengchi University, Taiwan, since 2000, where he is currently an Associate Professor with the Department of Computer Science. His research interests include computer vision, pattern recognition, human-computer interaction, and multimedia signal processing.