Mobile Augmented Reality in Supporting Performance Assessment: An Implementation in a Cooking Course

Kuo-Hung CHAO [a], Chung-Hsien LAN[b*], Yao-Tang LEE[c], Kinshuk[d], Kuo-En CHANG[e] , & Yao-Ting SUNG[f]

[a]*Assistant Professor, Hsing Wu University, Taiwan*

[b]*Assistant Professor, Taoyuan Innovation Institute of Technology ,Taiwan*

[c]*Assistant Professor, Nan Kai University of Technology, Taiwan*

[d] *Professor, Athabasca University, Canada*

[e,f] *Professor, National Taiwan Normal University, Taiwan*

[f]

*chlan@tiit.edu.tw

Abstract: In this study, a conceptual system framework was developed by integrating augmented reality (AR) technology to reduce the limitations in observation and assessment during performance assessment. Thus, the efficiency and reliability of mobile performance assessment can be enhanced. The processes of student performance can be presented from diverse approaches by using the characteristics of mobile devices and AR and considering the categories and situations of student performance. In this study, a novel mobile performance assessment system that incorporates AR to the processes of observation and assessment in performance assessment was developed. The mechanism emphasises the approaches adopted to present student works and provides opportunities for enhancing student communication and interaction. In addition, the system enables students to explain their works and incorporate the feedback they receive into future work. More importantly, mobile AR can be applied to offer personalised features and appropriate information in particular areas. Hence, students can interact with real or virtual information based on their needs. During this process, students can observe their own works from varying perspectives, acquire vital knowledge, develop the skills of critical thinking, and transform the process into a substantial self-established learning process.

Keywords: Augmented reality, performance assessment, mobile learning

1. Introduction

Conventional paper-based tests are adopted extensively in school education because the scoring is objective and the tests are easy to implement and have relatively high reliability and validity. However, this type of test has been criticised for being fragmented, having situations removed from the learning process, overemphasising cognition, and confining learning. Based on the popular educational philosophy of allowing students to develop diverse capabilities and achieve active knowledge building, performance assessment should be considered as a vital link in teaching. In addition to assigning a final score to students, the purpose of assessment is to develop a highly in-depth understanding regarding the process that students undergo during learning and to provide feedback to assist in student growth. Performance assessment, which is an assessment method that became widely used again after the 1980s, emphasises assessing actual performance and behaviour. This approach can be employed to compensate for the inadequacy of the paper-based test approach and assess the diverse disposition- and

skill-based achievements of students. O'Neil (1992) indicated that assessment plays a vital role in teaching and that the process of assessment consists of goal setting, data collection, organisation, and result analysis. The results can be used to enhance teaching and report the actual progress of students. By conducting assessments, teachers can evaluate students' learning outcomes and identify relevant factors, which can be used to devise improved methods for enhancing the effectiveness and quality of teaching. Considering the value of assessment, establishing a sound assessment mechanism is imperative for achieving multiple purposes and functions. Most teachers have had the experience of being questioned by parents if their children have failed tests because the teachers are biased against their children. In the assessment of design works, particular students have the misconception that they should receive a passing grade as long as they submit their assignments. This phenomenon has generated the discussion whether a well-defined set of assessment criteria should be applied when teachers are assessing the design works of students, who are highly self-aware. These students directly question teachers regarding the assessment criteria once they realise that their grades are unsatisfactory. In other words, students are no longer passive recipients of assessment results, but rather are attentive regarding their learning outcomes.

Nevertheless, a major problem encountered by the education community is determining the appropriateness of educational evaluation. Beywl and Speer (2004) found that previous studies had emphasised that the various dimensions of assessment in the narrow and broad senses must be consistent for an evaluation to be substantial, indicating that the assessment methods employed should be diverse. Problem-solving abilities are often regarded as abilities that require high-level reasoning or thinking and intelligence. Performance assessment has been recognised as one of the most effective methods for assessing this type of high-level thinking because this approach emphasises the application and demonstration of abilities in problem-solving situations and the complexity of problem-solving processes (Wiggins, 1993).

Although less objective and convenient compared with conventional paper-based tests, performance assessment can result in encouraging education outcomes when implemented under certain conditions. Previous studies (Dunbar, Koretz, & Hoover, 1991; Jiang, Smith, & Nichols, 1997; Baker, 1996; Messick, 1992) have indicated that the primary limitations and disadvantages of the performance assessment approach include the lack of comparison, limited reliability, unsatisfactory economic performance, and low validity. However, the majority of these factors can be attributed to the subjective consciousness of the assessors and errors in the measured situations. By contrast, augmented reality (AR) technology can be employed to display, in real situations, real-time information that is necessary for assessing or learning. From the perspective of cognitive psychology, this approach can be applied to reduce the errors resulting from the process of performance assessment and to minimise the time and economic costs that teachers must bear when observing student behaviour. Therefore, we examined the meaning, relevant studies, and limitations of performance assessment before investigating the effects that incorporating AR technology exert on improving performance assessment systems. Subsequently, we applied an AR-based performance assessment system to a cooking course to explore the effects of the application. The results yielded by conducting the performance assessment and paper-based tests were compared before a conclusion and recommendations were provided. The results of this study can serve as a reference for implementing performance assessment in teaching.

2. Performance Assessment

Performance assessment requires that students apply the knowledge and skills they have learned to perform hands-on practice rather than simply revalidating and recollecting the experience of learning (Roeber, 1990). This assessment method satisfies the needs of the current trend of constructivist learning and teaching (Chang, 2002b). Performance assessment motivates students to integrate the knowledge, skills, and dispositions required in the subject, and the results of the assessment can reflect students' problem-solving abilities in real life and the interest and needs of the students. Performance assessments, which can be conducted to evaluate high-level cognitive abilities and the dispositions and skills of students, are more comprehensive in the contents tested compared with conventional paper-based tests. Performance assessment can be integrated into teaching activities rather than being separated from teaching. Thus, teaching can be performed without interruptions. Furthermore, by adopting specific performance and assessment criteria, teachers can provide students with specific feedback, which motivates students to take initiatives in learning and assume the responsibility to

critique their own works and strategies. Therefore, performance assessment is an effective approach for facilitating teaching and learning.

Performance assessment was initially called performance-based assessment when the concept was first applied to education. Specifically, performance refers to the process of completing a task. The concept emphasises the authenticity and representativeness of the assessment regarding particular abilities as well as the significance of learning and evaluating in meaningful and real situations. By contrast, the concept of performance emphasises the necessity in challenging the intelligence, knowledge, and skills of the learners and the necessity in mastering the contents, processes, and outcomes of learning (Wiggins, 1993). Airasian (1996), Fitzpatrick and Morrison (1971), and Wiggins (1992) have indicated that in performance assessment, students are required to participate in an activity or create a piece of work to demonstrate the knowledge and skills they possess. Thus, students are required to demonstrate what they know and are capable of in actual situations. Aschbacher (1991) argued that the performance assessment in teaching-learning situations refers to teachers using their professional judgment to assess students' learning performance, which includes students' responses to tasks, the works delivered, and the process of learning. The characteristics of the performance assessment in this context are listed as follows: (a) students are expected to perform tasks or create objects that require high-level thinking or problem-solving skills; (b) the tasks based on which the students are assessed are meaningful, challenging, and integrated with teaching activities; (c) the tasks based on which the students are assessed are connected with real life; (d) processes and products are often the focuses of assessment; and (e) the assessment criteria and standards, which are vital dimensions and standards of the assessment, must be defined in advance (Herman et al., 1990).

The purpose of performance assessment is to motivate students to engage in useful, beneficial, and meaningful activities. Regarding form, this type of assessment is a component of learning activities, which require high-level thinking skills, an understanding of relevant concepts, and the ability to link various forms of knowledge. In addition, this type of assessment involves a specific explanation regarding the bases upon which student works are assessed. Therefore, the assessment is essentially a process of standard building rather than standard testing. Thus, the ultimate objective of performance assessment is to motivate students to comprehend the teaching contents and reach achievement standards by participating in teaching activities (Dorn, 1999). When performance assessment is conducted, students are required to apply particular learning results to daily situations. During the process of problem solving, students can reference necessary knowledge and independently build subject-specific knowledge and evaluate the possibility of the results. By so doing, students are motivated to independently determine what they want to learn and thus acquire knowledge actively. During performance assessment, the processes of thinking and the results yielded from the processes are both assessed. When implementing performance assessment, teachers can evaluate students' understanding of the problems, involvement, problem-solving skills, and self-expression abilities. Thus, the learning outcomes and learning processes of students can be fully reflected.

The purpose of performance assessment is using assessment to promote student development. The tasks designed for performance assessment, such as design-related experiments and system operation tasks, are often difficult for students to complete independently. Instead, students must work in teams and cooperate with their team members. Hence, during assessment, allowing other students and teachers to provide feedback on the assessment standards, record the processes, and evaluate the progress made by the student under assessment offers more opportunities for teacher–student and peer interactions. During these interactions, students can communicate and explain their learning experience and contemplate the learning process. In addition, the process of teamwork enables students to develop the abilities to communicate and cooperate with their peers and to develop favourable work attitudes. The difference between teachers (experts) and novices is that experts understand how to effectively use the knowledge they have acquired. Performance assessment enables teachers to comprehend the thinking process students undergo by observing the process of students' operations. Thus, teachers can understand whether the students' operations comply with the prescribed procedure.

To determine the effectiveness of a test or an evaluation method, we must explore the effectiveness of the method and the results based on the intended purposes. Ou (2002) asserted that one of the purposes of performance assessment is to offer a real and specific situation for students to demonstrate their knowledge and skills and thus make correct inferences regarding students' learning achievements. In this context, the accuracy with which the assessment results reflect the teaching goals can be enhanced. One of the characteristics of performance assessment is that students are allowed to flexibly use days of

a week or several weeks to engage in high-level thinking before completing their tasks. In terms of presentation, no correct answers or uniform patterns apply in performance assessment. Instead, the ambiguity of situations enables students to adopt various approaches to demonstrate individual creativity.

Gronlund (1993) and Linn (1991) have indicated that the purpose of performance assessment is to establish a model that enables students to focus on real learning activities. A course-oriented assessment approach is established when performance assessment is closely integrated with a course. This form of assessment system can inspire students and teachers to strive to enhance learning outcomes. The methods used in performance assessment are often one of the following: (a) checklists, which are used to evaluate whether the assessees exhibit a particular behaviour; (b) rating scales, which are employed to evaluate whether the assessees exhibit a particular behaviour and to rate the behaviour based on the extent; and (c) anecdotal records, where texts are used to describe and interpret assessee behaviour.

3. Limitation of Performance Assessment in Learning

Performance assessment, where the rating is often performed by professionals based on their observation and judgment, is subjective, demanding, and low in reliability compared with paper-based tests. In addition, the fairness of performance assessment is often questioned because the results cannot provide immediate feedback on student performance. Therefore, a critical problem that necessitates solution in implementing performance assessment is devising fair and objective rating criteria that are easy to apply and can be used to provide feedback to students (Lu et al., 2005). According to Lu et al. (2005), the criteria must also be able to provide specific answers to parent questioning, enable students to understand the dimensions of their learning capabilities, provide students with information that can be used to examine and evaluate their performance, offer feedback on student performance, and enable teachers to determine whether the assessment results truly reflect the response processes of students. Performance assessment often simultaneously involves multiple rating standards, some of which are objective and quantitative (e.g., completion time, quantity of completed work, materials consumed, and error) and some are subjective and qualitative (e.g., the originality and comprehensiveness of the completed work, the proficiency of action, and safety; Ou, 2002).

The application of performance assessment ranges from classroom teaching to large-scale surveys such as those conducted for appraisal purposes. A major concern in these forms of assessments is general quality control (Dunbar, Koretz, & Hoover, 1991). A previous study regarding performance assessment found that errors in the generalisation of performance assessment is primarily affected by the following four factors: (a) the items or activities used in the assessment, (b) the assessors, (c) the situations in which the assessments are conducted, and (d) the unintentional influence of assessees or other people (Jiang, Smith, & Nichols, 1997).

To determine whether a student has mastered a skill, the evaluator must collect performance data on multiple occasions. The number of observations necessary for making decisions can be determined based on the importance of the decision, the amount of time that an observation consumes, and whether the teacher has collected sufficient samples for evaluating student performance and behaviour. Hence, carefully and comprehensively observing all details in a single observation is equally vital for the assessors and assessees. Thus, the assessors can obtain all details by conducting only a minimal number of observations, thereby reducing the cost of assessment. Simultaneously, the assessees can benefit from fair assessments performed based on records that contain all details regarding their performance.

Another question involves assessor selection: Should teachers or professionals act as the assessors or should students perform self-assessment or peer assessment? Regardless, the assessors should have received training on rating.

Performance assessment has long been extensively applied to various fields primarily because this assessment method has advantages that cannot be achieved by conducting paper-based tests. Nevertheless, this method has limitations that must be overcome. Previous studies have shown that the limitations of performance assessment include the following concerns: rating is subjective; the criteria adopted by various assessors are inconsistent; the assessors do not truly understand the connotation of the assessment; the assessors do not follow the standardised assessment processes; and the rating

standards are undefined and time- and energy-consuming. To enhance the fairness and objectivity of performance assessment, scholars have developed procedures for implementing performance assessment. Among the steps defined by multiple scholars, the following four steps are indispensable: (a) defining the purpose of assessment, (b) confirming the assessment standards, (c) designing tasks or activities, and (d) selecting a rating or assessment method (Chen & Martin, 2000).

In summation, when performance assessment is applied to assess the operation and production of actual works, the fairness, objectivity, convenience, and timeliness of assessment must be considered to overcome the limitations of this method. These factors were used as the references for developing the questionnaire employed in this study. The following paragraphs list the primary limitations of performance assessment and how AR technology was employed to solve the problems:

Lack of comparability: In conventional standardised tests, the results can be compared against established norms; therefore, result interpretation is specific and clear. By contrast, the results of performance assessment are often affected by the subjective judgments of teachers; additionally, the criteria employed are occasionally confusing, thereby increasing the difficulty involved in comparing and interpreting the assessment results. The process of student performance can be recorded and students and teachers can employ AR technology during the rating process. Thus, the assessors and assessees can appear in real-time situations and serve as the direct references for assessment processes, thereby enhancing the accuracy of assessment.

Limited reliability: The majority of manual assessment methods are subject to the subjective influence of the assessors. Unlike standardised tests, for which computer scoring can be adopted, performance assessment relies on assessor observation and judgment. Consequently, the reliability of the assessors should not be overlooked. The errors in assessor reliability result from the assessors, and a satisfactory rating system can reduce assessor errors (Baker, 1996). We can employ AR to present the processes of work production or the implicit details hidden in the works. Thus, assessment accuracies can be increased substantially and the risks of rating errors resulting from assessor negligence or excessively short observation time can be reduced.

Unsatisfactory economic performance: The amounts of time and money spent on performance assessment are considerably greater than those spent on paper-based tests. Using AR to present the production processes of works can reduce the travel costs that teachers would otherwise spend for conducting on-site observations. In addition, the assessors can watch videos repeatedly to reduce rating errors and the amount of effort that teachers must spend on assessments can be reduced. Furthermore, occasionally teachers must simultaneously observe multiple students, thereby rendering them unable to observe all details within a particular period. Adopting AR can prevent this problem.

Low validity: In performance assessment, ambiguous problem situations are designed to test the high-level thinking abilities of assessees. Nevertheless, the validity of ambiguous problems is difficult to control; consequently, the assessment can be irrelevant to the teaching contents. An AR-integrated system can show in real-time the rating standards and the feedback from teachers or peers; thus, the associated cognition of feedback materials and student works can be enhanced. Therefore, students can more effectively immerse themselves into the teaching situations, thereby improving the validity of assessments.

4. Methods of the Performance Assessment Conducted in Hands-on Performance Courses

The assessment conducted in implementation activities are considered assessments conducted during activities. Performance assessment is often based on observation; thus, it can also be referred to as the work evaluation method (Tsai, 1996).

Establishing a set of criterion to be used in performance assessment enables designers to perform self-assessment during the processes of creation and development and the completion stage and enables assessors to eliminate uncertainty and overcome the complexity involved in the assessment. According to Wolansky (1985), a U.S. scholar specialising in vocational-education studies, teachers must focus on the following concerns when evaluating student works or productions: (a) the performance of the works

should accomplish the required purposes; (b) records regarding the production processes of the final products should be made available for assessment purposes; and (c) teachers should provide well-defined explanations or standards regarding the quality of the final products. Wolansky emphasised that a standard table or a criteria table is a convenient and informative tool that enables students to understand the standards for excellent work and the contents of assessment. Yunghans (1981) indicated that the standards for evaluating art works should be (a) the purity and openness of expressions, (b) the problem-solving methods exhibited, (c) the duration of focused attention on production, and (d) the attention to detail in images.

Khattri et al. (1998) argued that the results of performance assessment differ when the methods or systems employed vary. Specifically, the following five characteristics of performance assessments should be considered: (a) the purpose of a performance assessment, (b) the format of assessment, (c) the subject areas being assessed, (d) the levels of students, and (e) the implementation of performance assessments.

Gronlund (1993) categorised performance assessments into the following types based on the extent to which the situations used in the tests are true to reality: (a) paper-and-pencil performance, (b) identification tests, (c) structured performance tests, (d) simulated performance, and (e) work samples. The experiments in this study were conducted using a structured performance test and simulated performance.

In summary, the specific steps of rating criterion design are (a) teachers must first determine the target of assessment, be it the process, the result, or both; (b) subsequently, teachers must identify the contents or scopes of observation, list the focuses of the observation and assessment, and explain the significance of the criteria adopted; (c) teachers can discuss the rating criteria with students and ensure that the students truly understand the connotations of the criteria; the rating criteria can also be established by students or jointly by students and teachers; and (d) before conducting a performance assessment, teachers must carefully examine the items regarding detailed behaviour and apply necessary revisions.

5. Performance Assessment with Mobile Augmented Reality

AR enables users to visualise real environments in a real world with the digital information overlaid on real environments (objects or locations), thereby improving user experiences (Berryman, 2012). The combination of additional information and real situations can enhance the senses of reality and presence for people. The theoretical basis for the mobile AR system that integrates human–computer-context interactions is situated cognition. The fundamental argument of the theory is that knowledge acquisition and learning occurs after people interact with situations, which include social environments such as people and social culture, and physical environments such as the contexts formed by scenes and artefacts (Brown, Collins, & Dugid, 1989; Greeno, Collins, & Resnick, 1996).
Applications of AR and high-tech products to situated teaching activities are lacking because the attention that users direct toward additional information and real scenes is difficult to balance. In addition, human–computer-context interactions are difficult to achieve. Participants may focus excessively on human–computer interactions and overlook human–context (objects in scenes and information contexts) interactions, which are more crucial than human–computer interactions in real situations. Therefore, the link between additional information and real environments should be emphasised in the virtual contents presented in AR (Klopfer & Squire, 2008; Chang et al., 2014; Zhang et al., 2014).

In informal learning, the application of mobile devices has recently attracted an increasing amount of attention (Semper & Spasojevic, 2002; Kwak, 2004; Cabrera et al., 2005; Chang et al., 2006; Sung et al., 2010a). However, studies regarding the application of AR navigation are scant (Barber et al., 2001; Sparacino, 2002; Damala et al., 2007; Damala et al., 2008; Portalés et al., 2009).
The mixed reality spectrum (Fig. 2-2) developed by Milgram and Kishino (1994) offers a valuable basis for exploring the integration of reality and virtual reality. AR is situated on the spectrum between virtual and real environments. Based on the definition of mixed reality, Milgram and Kishino developed the linear spectrum, showing the transition from real environments to virtual environments. With real

environments on the left end of the spectrum and virtual environment on the right, AR is located toward the left end of the spectrum, indicating that the main subject in AR is real objects and that virtual objects are additional and supplementary. When AR is applied to spaces, implicit spatial information is transformed to explicit spatial information by employing technologies that incorporate virtual objects with the real world. Hence, additional values and meaning are added to spaces.

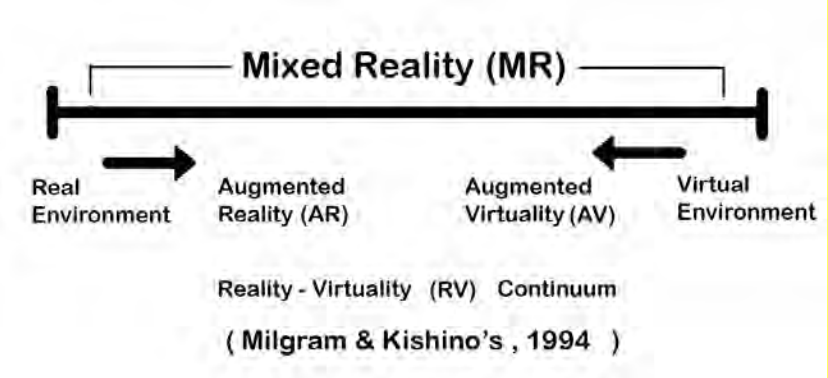

Figure 0. Reality-Virtual Continuum

AR is a technology that incorporates virtual and real objects in real environments (Azuma et al., 2001). AR enables users to visualise real environments in a real world with digital information overlaid on actual environments (objects or locations), thereby improving user experience (Berryman, 2012). Initially, most studies used head-mounted displays to present the results of virtual–real environment integration. Azuma (1997) defined three criteria for AR: (a) the combination of virtual and real environments, (b) real-time interaction, and (c) 3D referencing. Scholars generally agree that AR can be used to enhance the experience that users have when interacting with real environments. In addition, virtual information enables users to obtain information that otherwise cannot be directly acquired from the real world. Because of this feature, AR is considered an effective tool that users can employ to achieve objectives in the real world (Azuma, 1997).

Technically, AR presentation can be divided into the following two types: marker-based and marker-less identification. Specifically, marker-based identification operates based on the principle of quick response codes, which are 2D bar codes or dot matrices in square icons. The markers are locked and read using the cameras on mobile devices and identification software. Subsequently, the interaction is activated using 3D objects or videos. By contrast, marker-less identification is based on the global positioning system. Users can use mobile devices to locate objects that interest them and floating markers or chat boxes are shown to display information through the cameras installed in devices.

Since the 1990s, AR has been applied in various fields, including geography (Vlahakis et al., 2002; Portalés, Lerma, & Pérez, 2009; Priestnall, 2009), linguistics (Liu, 2009), social sciences (Hedley et al., 2002; Mathews, 2010; McCall et al., 2011), mathematical sciences (Wang, 2007; Yim & Seong, 2010), natural sciences (Klopfer & Squire, 2008; Liu, Tan, & Chu, 2009; Zhang et al., 2014), biomedicine (Vilkoniene, 2009; Strickland et al., 2011), arts and humanities (Shen, Ong, & Nee, 2010; Chang et al., 2014), leisure and recreation (Portalés et al., 2010; Wang & Chen, 2009), and advertising and marketing (Moltenbrey, 2011).

Barber et al. (2001) indicated that using smartphones to display additional information and placing the screens next to student works is essentially integrating virtual and real environments into the same view. Thus, the number of times users switch between the exhibited works and description plaques can be reduced and the number of searches users must make can also be minimised. In a study conducted by Dunleavy et al. (2008), the students engaged in role-playing tasks by using AR. Specifically, the students walked around campus while the cameras on the mobile devices they were holding displayed digital objects and virtual characters that were overlaid onto real spaces. Video, audio, and text files were used to provide clues and challenges for narration, navigation, and cooperation. Thus, the learning objectives for subjects such as math, language arts, and scientific literacy at junior and senior high schools can be accomplished.

Damala et al. (2008) agreed that the integration of virtual and real environments revolutionised the interaction between people and objects in an unanticipated manner. The tiny screens on the devices can represent complete environmental spaces and facilitate establishing a close relationship between the

appreciated works and additional information. However, AR technology requires improvement when applied to learning activities (Billinghurst et al., 2003; Dunleavy et al., 2008; Wang & Chen, 2009; McCall et al., 2011). For example, considering the coexistence of virtual and real environments in AR, the additional information presented in AR may be designed to attract participant attention so that the participants can see the information. Consequently, the participants may excessively focus on the contents shown in the AR system, particularly on the additional information, and ignore the actual environment and surroundings (Billinghurst et al., 2003; Dunleavy et al., 2008; Wang & Chen, 2009).
In summation, to design a performance assessment learning system that achieves human–computer-context interactions, we employed AR technology to develop a performance assessment system that enables peer assessment. In this system, the criteria based on which learners produce their works or assessors evaluate the works are predefined. Thus, students can evaluate their own works or peers' works based on sufficient information, thereby developing strong learning motivations and achieving great efficacy. In addition, teachers or assessors can spend comparatively less time and simultaneously evaluate assessees' works accurately and fairly.

6. System Realization and Illustrative Example

6.1 System Architecture

In the field of education, numerous situations cannot be experienced or represented in the classroom setting. AR is the most appropriate technology for incorporating or adjusting students' learning experience based on specific needs. AR is defined as a real-world environment whose elements are built upon computer-generated sensory input such as sound, video, graphics or GPS data. In this study, AR allows students to see virtual objects about peers' works or contents in a real world environment with the aid of camera during the assessment process. The overall framework of the use of mobile AR technique in performance assessment is described in figure 1.

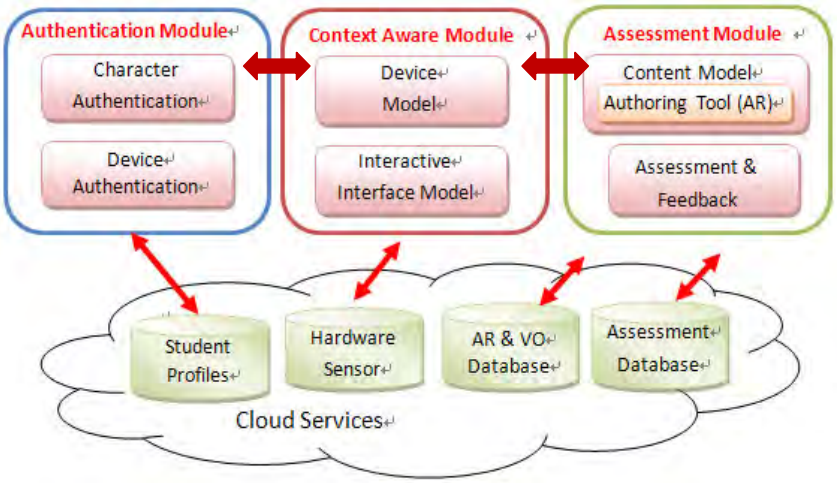

Figure 1. The architecture of the mobile AR technique

The entire processes of learning and assessment can be divided into the following three modules: the authentication module, the context-aware module, and the interactive assessment module. In addition, the process is supported by four databases on cloud servers, which are student profiles, a hardware sensor, the AR and virtual object database, and the assessment database. The authentication module enables authorised people to obtain appropriate information for completing corresponding tasks. The context-aware module enables assessors to employ appropriate device functions for accessing suitable information for performing rating. In the context-aware module, mobile devices list appropriate learning contents after detecting student locations and collecting onsite information. Thus, learners can select appropriate learning materials from the content model. The content model retrieves appropriate materials from the virtual object database before providing them to assessors. Subsequently, learners

can use the authoring tool provided by the system to establish an AR marker, work descriptions, and an AR context object. Once all steps are completed and the information is uploaded to the system, appropriate virtual information is used through AR technology to overlay images onto corresponding objects in the real world. Thus, assessors can rate the works conveniently and accurately. Through the AR work presentation technology employed in the system, the assessment module enables assessors to conveniently and directly observe the works of peers. Hence, assessors can provide feedback for the peers they evaluate. In addition, the system can be employed to develop a work-specific exhibition situation for peer references, thereby enabling peer assessors to provide feedback. Additionally, teachers can use the AR performance assessment system to understand the peer assessment performed by students before providing feedback for the assessors and assessees. More importantly, teachers can integrate previous cases to develop new teaching situations that are highly appropriate and inspiring.

6.2 Walk-through Illustrative Example

The methods for conducting performance assessment are diverse, including observation, document records, and real-time performance. The methods adopted in this study were real-time performance and peer assessment. Peers who possessed similar knowledge levels observed and learned from each other before offering recommendations. Specifically, a class of 50 sophomore students at the culinary department of a technical institute were recruited as the participants of this study. A performance assessment experiment was conducted during a training course for cooking licenses in Western cuisine. The students were divided into groups of five, obtaining a total of 10 groups. The group members divided the labour between themselves. The students were randomly assigned to the groups without considering sex or cooking skills. During class, the teacher designated an item from the licensing examination as a task. The teacher demonstrated the cooking procedure once, after which the completed set was recorded and used as an item marker in the AR performance assessment system. Before the students began the performance, they used mobile devices (tablet personal computers) to photograph the sample. Subsequently, the system displayed real-time information (learning mode) that corresponds

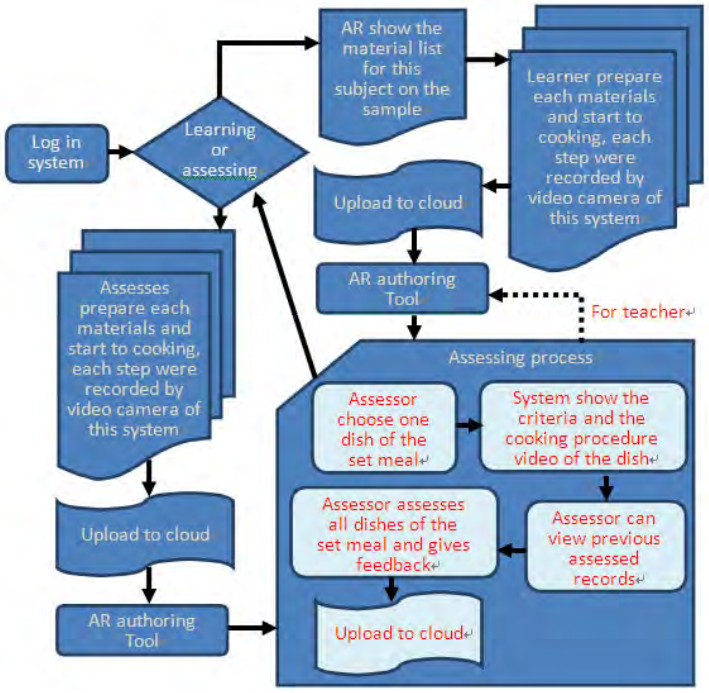

Figure 2. The process of the mobile augmented reality p assessment system

to the dish onto the dish image, such as the ingredients that should be prepared and the steps of cooking. Thus, students can follow the instructions during the performance and record the process of cooking by using mobile devices. Subsequently, the videos were uploaded to cloud servers and arranged by the

names of the dishes. After all the dishes were completed, the system integrated all the data and was prepared for peer assessment. The system enables assessors to review the records of assessed dishes for reference. After the assessment mode began, the system listed the content and assessment criteria of the set for the assessees rather than listing information regarding the dishes after the samples were scanned. The assessees prepared the ingredients for cooking the dishes and had every step recorded before the videos were uploaded to cloud servers. Subsequently, the system integrated the information for the teachers to provide ratings and feedback. The procedure of the experiment is shown in Fig. 2.

During assessment, the AR performance assessment system identifies each dish and lists the contents when the assessors use the cameras on their mobile devices to photograph the sets completed by the assessees. When an assessor selects the name of a particular dish, the video showing the cooking process is immediately shown on the screen. In addition, the assessment criteria are displayed simultaneously, enabling the assessors to perform the assessment intuitively and clearly. Because the entire cooking procedures were videotaped, the assessors were able to observe all the details that interested them. Thus, the assessors did not miss crucial details as they otherwise would when simultaneously observing several groups of students. Additionally, the assessment criteria adopted are consistent because they are shown in real-time. Hence, the errors in performance assessment can be minimised.

Furthermore, the assessors can provide real-time feedback and recommendations during assessment. The feedback can be uploaded to cloud servers immediately following assessment. Thus, the students can immediately review and share the feedback and recommendations regarding their works and further discuss among themselves by using the system. Real-time sharing and the real-time display of assessment criteria enable students to immediately understand the advantages and disadvantages of their works and to use the feedback to improve their works. Thus, the learning objectives were achieved.

7. Conclusion

This paper presents the novel framework of an enhanced performance assessment system complemented by the use of smart and mobile devices. Integrating the AR technology overcame some of the limitations of conventional performance assessment systems, such as the implementation method, excessively high costs, and substantial errors. In this framework, the AR technology enables students to observe how their peers completed their works by displaying videos of the cooking process over the completed dishes. During the assessment, students can determine whether their peers followed the instructions correctly by comparing the performance against the assessment criteria. By doing so, students can discover their own inadequacies or learn from other people's methods. In addition, the system provides each student stable and convenient information and digital content based on environmental parameters or the identification of particular objects. Thus, students can learn while engaging in activities based on which their performance is assessed. Students can obtain appropriate learning information by using mobile devices to photograph and identify target objects at appropriate moments and particular locations. The novel framework developed in this study, in which the AR technology was integrated, enables students to use various methods to observe the cooking processes and completed works of their peers. Simultaneously, the students can receive real-time feedback and recommendations regarding their own works. Hence, the barrier resulting from conflicting opinions between students can be eliminated, and students' understanding of each other's opinions can be enhanced. Thus, the accuracy of the results of performance assessment can be improved. From the perspective of cognitive psychology, showing assessment criteria and feedback in real-time situations during assessment enables students to develop strong impressions of the feedback. Therefore, the students are highly capable of incorporating the feedback into their future work to achieve improvement and growth. The novel AR-integrated framework used in this experiment is almost complete. However, additional work is necessary. For example, studies can be conducted using an experiment sample that is larger than that used in this study, a large number of performance tests in a classroom setting, and an enhanced system.

References

Lin, C. P., 1999, Teaching assessment in art courses: Using the Department of Art and Craft Education at Taipei Municipal Teachers College as an example. *Journal of Taipei Municipal Teachers College*, 30, 451-474.

Ou, C. H. 2002. *Educational Testing and Evaluation*. Taipei: Psychological Publishing Co., Ltd.

Lu, J. H., Chen, M. Y., & Wu, Y. Y., 2005, An Action Research of a Performance Assessment Design on Assessing Students' Ability of Conducting Experiments. *Contemporary Educational Research Quarterly*, 13, 4, 95-132.

Chang, L. L., 2002a, The promised land of assessment reform, virtual or real: The operation and criteria for performance assessment. *Journal of Education Research*, 93, 76-86.

Chang, L. L., 2002b, Establishing the validity of performance assessment based on the significance of scores. *Journal of Education Research*, 98, 76-86.

Tsai, D. C., 1996, Using group decision-making techniques to assess designs. Proceedings of the 2nd Conference on the Research and Practice of Design and Management. pp. 1-8.

Yen, J., 1997, Applying structured assessment methods to the assessment of visual communication designs. Proceedings of Visual Communication Design: Education, Culture, and Technology. Taipei: AsiaPac Books. pp. 231-236.

Li, D. W., 1986, *Measurement and Evaluation in Technical and Vocational Education*. Taipei: San Min Books Co., Ltd.

Beywl W. and Speer S., 2004, Data- and literature-based reflections on western European evaluation standards and practices, New Directions for Evaluation, No. 104, pp. 43-54.

Wiggins, G., 1993, Assessment: Authenticity, context, and validity, Phi Delta Kappa, pp. 200-214.

Airasian, P. W., 1996, Assessment in the classroom, McGraw-Hill, New York.

Fitzpatrick, R. and Morrison, E. J., 1971, Performance and product evaluation, In Thorndike, R. L. (Ed.) , Educational measurement, American Council on Education, Washington, DC, pp. 237-270.

Aschbacher, P. R., 1991, Performance assessment: state activity, interest, and concerns, Applied Measurement in Education, Vol. 4, No. 4, pp. 275-288.

Herman, J. L., Aschbacher, P. R., and Winters, L., 1990, Issues in developing alternative assessments, Paper presented at the annual meeting of the California Educational Research Association, Chicago.

Dorn, C., 1999, Mind in art: Cognitive foundations in art education, Mahway, Lawrence Erubaum Associates Publishers, New York.

Chen, Y. F. and Martin, M. A., 2000, Using performance assessment and portfolio assessment together in the elementary classroom, Reading Improvement, No. 37, pp.32-38.

Wolansky, W., 1985, Evaluating student performance in vocational education, The Iowa State University Press, Ames, IA.

Yunghans, M., 1981, A pull-out program for gifted elementary students, School Arts, Vol. 80, No. 8, pp. 50-51.

Khattri, N., Reeve, A. L., and Kane, M. B., 1998, Principles and practices of performance assessment, Lawrence Erlbaum, Mahwah, NJ.

Baker, E. L., Abedi, J. (1996). Dimensionality and Generalizability of Domain-Independent Performance Assessments. Journal of Education Research, 89(4), 197-205

Roeber E. D. (1990). Performance assessment: A national perspective. Policy Briefs Numbers 10 & 11. Special Double Issue. (ERIC Document Reproduction Service No. ED 370980)

Wiggins, G. (1993). Assessment: Authenticity, context, and validity. Phi Delta Kappan, November, 200-214.

Gronlund, N.E. (1993). How to make achievement tests and assessments (5nd ed.).Boston: Allyn & Bacon.

Linn, R. L., Baker, E. L., & Dunbar, S. B.（1991）. Complex, performance-based assessment: Expectations and validation criteria. Educational Research, 20(8), 15-21.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational Measurement, 13 –104.New York: Macmillan.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), Encyclopedia of educational research (6th ed., Vol. 4, pp. 1487-1495). New York: Macmillan.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education, 4, 289-304

Jiang, Y. H., Smith, P., & Nichols, P. (1997). Error sources influencing performance assessment reliability or generaliziability: A meta-analysis. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Popham, W. J., 1995, Classroom assessment: What teachers need to know, Allyn and Bacon, Boston.