Integration of Mobile AR Technology in Performance Assessment

Chao Kuo-Hung¹, Chang Kuo-En²*, Lan Chung-Hsien³, Kinshuk⁴ and Sung Yao-Ting⁵

¹National Taiwan Normal University, Taiwan // ²National Taiwan Normal University, Taiwan // ³Taiyuan Innovation Institute of Technology, Taiwan // ⁴Athabasca University, Canada // ⁵National Taiwan Normal University, Taiwan //

stefanckh@gmail.com // kchang@ntnu.edu.tw // chlan@tiit.edu.tw // kinshuk@ieee.org // sungtc@ntnu.edu.tw

*Corresponding author

(Submitted January 22, 2015; Revised October 26, 2015; Accepted March 8, 2016)

ABSTRACT
This study was aimed at exploring how to use augmented reality (AR) technology to enhance the effect of performance assessment (PA). A mobile AR performance assessment system (MARPAS) was developed by integrating AR technology to reduce the limitations in observation and assessment during PA. This system includes three modules: Authentication, AR Context Aware, and AR Interaction. It identifies the character of system users, recognises subjects and activities that users should join in, and follow and assists users in showing their works and doing assessment anywhere anytime. Both qualitative and quantitative research were applied in this study, and it presents an action research conducted in PA institutions that, through a questionnaire, gave information in connection to the valuation of mobile AR technology and the achieved effects. The proposed mechanism of PA 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. From the observation of survey results, 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

Introduction

Based on the popular educational philosophy of allowing students to develop diverse capabilities and achieve active knowledge building, performance assessment (PA) should be considered 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 provide feedback to assist in student growth. Toptas (2011) indicated that an effective evaluation of the students who answered the questions in a particular period of time will be insufficient. If we want to correct this weakness, the performances of the students must be measured with the observation of the process as well. O’Neil and Osif (1993), and VanTassel-Baska (2014) 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. Turgut and Baykul (2012) point out that the process can be measured alongside the results of learning outputs by measuring the performances. In addition, it is asserted that the measurement of students’ performance gives them the opportunity to effectively learn the concepts, complex events, and their structures.

Nevertheless, a major problem encountered by the education community is determining the appropriateness of educational evaluation. PA 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; VanTassel-Baska, 2014).

Previous studies (Bay, Küçükoğlu, Kaya, Gündoğdu, Köse, Ozan, & Taşun 2010; Jiang, Smith, & Nichols, 1997) have indicated that the primary limitations and disadvantages of the PA 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 PA 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 PA before investigating the effects that incorporating AR technology exert on
improving PA systems. Subsequently, we applied an AR-based PA system to a cooking course to explore the effects of the application. The results of this study can serve as a reference for implementing PA in teaching.

**Performance Assessment (PA)**

Performance Assessment (PA) requires students to apply the knowledge and skills they have learned to perform hands-on practice rather than simply revalidating and recollecting the experience of learning (VanTassel-Baska, 2014). This assessment method satisfies the needs of the current trend of constructivist learning and teaching (Chang, 2002). PA 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 compared with conventional paper-based tests. When evaluating a student for periodic checks or a promotion, there has to be a list of measurable performance criteria that can be applied consistently to all members of a particular class. (Kirovska & Qoku, 2014). Therefore, 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, PA is an effective approach for facilitating teaching and learning.

Students who acquire a comprehensive set of motor skill proficiencies and continually engage their skills are more likely to become fit adults (Kalaja, Jaakkola, Luikkonen, & Watt, 2010). Strong factors include competence at performing and confidence in using motor skills, both of which are established through early experiences in physical activity and sport (Rink & Hall, 2008).

The purpose of PA is using assessment to promote student development. 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. PA 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.

**Limitation of performance assessment in learning**

PA, 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. Therefore, a critical problem that necessitates solution in implementing PA is devising fair and objective rating criteria that are easy to apply and can be used to provide feedback to students (Lu, Chen, & Wu, 2005; VanTassel-Baska, 2014).

A previous study regarding PA found that errors in the generalisation of PA are 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 assesses or other people (Jiang, Smith, & Nichols, 1997).

To determine whether a student has mastered a skill, assessors must collect performance data on multiple occasions. They can obtain all details by conducting a minimal number of observations, thereby reducing the cost of assessment. Simultaneously, assesses can benefit from fair assessments performed based on records that contain all details regarding their performance. When PA 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. Table 1 lists the primary limitations of PA and how AR technology was employed to solve the problems.

The assessment conducted in implementation activities are considered assessments conducted during activities. PA is often based on observation; thus, it can also be referred to as the work evaluation method. However, the traditional assessment methods have been replaced by the evaluation activities based on the authentic
apprehension (Bay et al., 2010). The concept of assessment brought by the constructivist learning approach must be varied according to the traditional methods, handled as an element of the learning-teaching processes (not independent from the process), and be a part of educational activities.

### Table 1. Limitations of PA and solutions

<table>
<thead>
<tr>
<th>Limitations of traditional PA</th>
<th>How AR solve the problems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subjective judgments of the evaluator:</strong>&lt;br&gt;The results of PA 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.</td>
<td>The process of student performance can be recorded, and students and teachers can employ AR technology during the rating process. Thus, the assessor and assesses can appear in real-time situations and serve as the direct references for assessment processes, thereby enhancing the accuracy of assessment.</td>
</tr>
<tr>
<td><strong>Limited reliability:</strong>&lt;br&gt;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, PA 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.</td>
<td>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.</td>
</tr>
<tr>
<td><strong>Unsatisfactory economic performance:</strong>&lt;br&gt;The costs spent on PA are considerably greater than those spent on paper-based tests.</td>
<td>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. 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.</td>
</tr>
<tr>
<td><strong>Low validity:</strong>&lt;br&gt;In PA, ambiguous problem situations test the high-level thinking abilities of assesses. Nevertheless, the validity of ambiguous problems is difficult to control; consequently, the assessment can be irrelevant to the teaching contents.</td>
<td>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.</td>
</tr>
</tbody>
</table>


### Mobile augmented reality

AR enables users to visualise 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 (Greeno, Collins, & Resnick, 1996). In addition, human–computer-context interactions are difficult. 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 information presented in AR (Chang, Chang, Hou, Sung, Chao, & Lee, 2014).

In informal learning, the application of mobile devices has recently attracted an increasing amount of attention (Kwak & Stoddard, 2004; Cabrera et al., 2005). However, studies regarding the application of AR navigation are scant (Damala, Marshal, & Houlier, 2007; Portalés & Perales, 2009). The mixed reality spectrum (Figure 1)
developed by Milgram, Takemura, Utsumi, & 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.

![Mixed Reality (MR) Continuum](image)

*Figure 1. The mixed reality spectrum and mobile AR technique (Milgram, Takemura, Utsumi & Kishino, 1994)*

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). 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). Figure 2 shows how to use AR to display appropriate information.

![AR Technology](image)

*Figure 2. The application of AR technology*

As showing in Figure 2-1, for the purpose of locating virtual information in the right place in real word, tags or markers are necessary for recognition. AR recognises the tag and gets its position as the position of the corresponding virtual information. Figure 2-2 shows how outdoor AR uses GPS location information and compass direction as the tag to locate virtual information.

Since the 1990s, AR has been applied in various fields, including geography (Portalés et al., 2010; Priestnall & Cowton, 2009), linguistics (Liu, 2009), social sciences (Mathews, 2010), mathematical sciences, natural sciences, biomedicine, arts and humanities, leisure and recreation, and advertising and marketing.

In summation, to design a PA learning system that achieves human–computer-context interactions, we employed AR technology to develop a PA system that enables peer assessment. We predefined the criteria on which learners produce their works or assessors evaluate the works in this system. 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 asseesees’ works accurately and fairly.

**System realisation and illustrative example**

**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 the mobile AR technique in PA is described in Figure 3.

The entire processes of learning and assessment can be divided into the three modules: the authentication module, the context-aware module, and the interactive assessment module. In addition, the process is supported by three databases on cloud servers, which are student profiles, the AR and virtual object database, and the assessment database. The authentication module enables authorised people to obtain appropriate information for completing corresponding tasks. In the AR context-aware module, mobile devices list appropriate learning contents after detecting users’ cooking subjects. Thus, learners can select appropriate learning materials from the databases. The authoring tool retrieves appropriate materials from the virtual object database before providing them to assessors. AR interaction module assists users in showing their works and doing assessment. Then they can discuss in this system anywhere anytime and give feedback.

For teachers and students, the authentication module shortens the gaps between learners and assessees. The AR context-aware module is situated cognition by using AR to integrate human–computer-context interactions, and it reduces the assessment effort while assessors assess every assesse at the same time. The AR interaction module assists teachers and students in sharing and discussing feedback more easily and frequently.

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 PA 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.

**Walk-through illustrative example**

The methods for conducting PA 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 PA experiment was conducted during a training course for cooking licences in Western cuisine. The students were divided into groups of five, obtaining a total of ten 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 PA system. The procedure of the experiment is shown in Figure 4.

Students first choose the character (assessor or assesse) they want to play. Before students begin the performance, they use mobile devices (tablet and personal computers) to photograph the sample. Subsequently, the system displays real-time information (learning mode) that corresponds 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 or cook by themselves and record the process using mobile devices. Then the videos are uploaded to cloud servers. After all the dishes are completed, the system integrates all the data for peer assessment. The system enables assessors to review the records of assessed dishes for reference. In the assessment mode, the system lists the content and assessment criteria of the set for assesses. Assesses prepare the ingredients for cooking the dishes and record every step. Finally, the system integrates all the information for the teachers to provide ratings and feedback.

During assessment, the AR PA system identifies each dish and lists the contents when assessors use the cameras on their mobile devices to photograph the sets completed by 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 have 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 PA 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.
Procedure and evaluation

Setting and participants

In this study, we wanted to know whether the effect of PA can be enhanced by AR system. Could AR technology solve the problems of PA? Thus, both qualitative and quantitative research designs were used in this study. To validate the proposed framework and the effectiveness of the system, we conducted a survey. In the study, a class of 50 sophomore students in the culinary department of a technical institute were recruited as the participants of this study in Taiwan. And 47 survey results were recorded from these 50 participants. According to Van Zundert et al. (2010), peer assessment is facilitated by working in small groups of three or four students. These students are better able to compare feedback from different peers to determine its relevance. The grouping size of a cooking team cannot be small, and there must be an odd number of participants for the sake of voting convenience when there is a decision to make. A five-person grouping is the closest grouping size that fits all the situations. Therefore, in our study, the students were arranged in ten groups of five students. The experiment arranged two activities of PA: pen-and-paper assessment and mobile AR assessment. All groups undertook peer assessment on paper first, while the same groups employed the system to do peer assessment one week later. Finally, a survey considering the amount of time spent focusing on peer discussion and students’ attitude regarding the use and acceptance of MARPAS was assigned to students to evaluate the effectiveness of the system. The grades of assessment were compared between assessment by paper-and-pen and assessment using MARPAS.

During the experiment, all participants had to complete the training of peer assessment in order to understand the process of observation, assessment, and interaction. Subsequently, each group joined two activities: assessment on papers and assessment on mobile devices by using MARPAS. The teacher demonstrated the cooking procedure once, after which the completed set was recorded and used as an item marker in the AR PA system. The AR application constructed the relation between the image of the draft and the exposition. However, assessment criteria defined by the teacher were different based on the various dishes. The assessment criteria represented the teacher’s requests and were also the basis for marking peers’ works.

After students filled out the questionnaire, interviews were arranged with each group, where the aim was to explore the participants’ attitudes toward the strengths and limitations of the system and any suggestion students may have for improvement. And the results of assessments were recorded to measure the progress of students’ works. A qualitative approach was used to analyse the participants’ feedback. Then, the interviews about MARPAS were arranged with teachers who assigned these performance assessments.

Data collection and analysis

After the experiment, the assessment process and results were recorded. In order to validate the proposed framework, the experimental results were analysed. In addition, a survey was conducted to collect additional data from participants. A structured questionnaire was developed for this purpose and was sent to five experts to evaluate reliability and validity. These experts included two professors in cognitive psychology, one professor in computer science, and two cooking teachers. Subsequently, the questionnaire was distributed to each student in the ten groups. This questionnaire contained three sections: the first section related to the experience in joining peer assessment, while the second section dealt with the feedback in using the mobile service. The third section figured out how AR technology assisted students in demonstrating their work. The reliability of this questionnaire was checked by testing the value of Cronbach’s alpha. Cronbach’s alpha of the questionnaires was estimated as internal consistency reliability. The results showed that the reliability of the questionnaires was good (whole questionnaire: alpha = 0.937, section 1: alpha = 0.925, section 2: alpha = 0.896, section 3: alpha = 0.781). All show that the questionnaire is reliable.

In the questionnaire, there were three questions pertaining to personal information, 16 questions for the first section, 8 questions for the second section, 5 questions for the third section, and 4 questions for the fourth section. The questionnaire used five-point Likert scale to register students’ answers. Collected data was analysed using correlations, associations, and descriptive statistics in order to assess the relationships between variables.
Findings and discussion

According to Kothari (2009), for sample sizes of more than 30, the t-distribution is so close to the normal distribution that one can use to approximate the t-distribution (Kothari, 2009). Therefore, the t-test was used to validate the effectiveness of the system. Seven items in the first section of the questionnaire were selected to analyse students’ attitudes toward the use of peer assessment. The hypotheses are listed below:

\( H_0: \) Students did not agree on effect of performance assessment \((\mu \leq 3)\)
\( H_1: \) Students agreed on effect of performance assessment \((\mu > 3)\)

**Table 2. Analysis results for the effect of PA**

<table>
<thead>
<tr>
<th>Test value</th>
<th>( t )</th>
<th>( df )</th>
<th>Sig. (2-tailed)</th>
<th>Mean difference</th>
<th>95% Confidence Interval of the Difference</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I could understand the teacher’s requests in this course more clearly from performance assessment activities.</td>
<td>8.076</td>
<td>46</td>
<td>.000</td>
<td>.894</td>
<td>.67 to 1.12</td>
<td>3.89</td>
<td>.759</td>
</tr>
<tr>
<td>I could understand peers’ recognition of my work from peer assessment activities.</td>
<td>9.913</td>
<td>46</td>
<td>.000</td>
<td>1.000</td>
<td>.80 to 1.20</td>
<td>4.00</td>
<td>.692</td>
</tr>
<tr>
<td>Performance assessment activities increased my learning motivation.</td>
<td>11.524</td>
<td>46</td>
<td>.000</td>
<td>1.021</td>
<td>.84 to 1.20</td>
<td>4.02</td>
<td>.608</td>
</tr>
<tr>
<td>Performance assessment activities engaged my attention in the course.</td>
<td>8.398</td>
<td>46</td>
<td>.000</td>
<td>.915</td>
<td>.70 to 1.13</td>
<td>3.91</td>
<td>.747</td>
</tr>
<tr>
<td>Performance assessment activities increased the interaction with the teacher.</td>
<td>9.141</td>
<td>46</td>
<td>.000</td>
<td>.915</td>
<td>.71 to 1.12</td>
<td>3.91</td>
<td>.686</td>
</tr>
<tr>
<td>Performance assessment activities increased the interaction with peers.</td>
<td>10.960</td>
<td>46</td>
<td>.000</td>
<td>1.000</td>
<td>.82 to 1.18</td>
<td>4.00</td>
<td>.626</td>
</tr>
<tr>
<td>The suggestions from peers were helpful to me.</td>
<td>10.446</td>
<td>46</td>
<td>.000</td>
<td>.979</td>
<td>.79 to 1.17</td>
<td>3.98</td>
<td>.642</td>
</tr>
</tbody>
</table>

The second section of the questionnaire related to students’ attitudes toward the use of mobile service for PA. The hypotheses are listed below:

\( H_0: \) Students did not agree on effect of mobile service for performance assessment \((\mu \leq 3)\)
\( H_1: \) Students agree on effect of mobile service for performance assessment \((\mu > 3)\)

The experimental results reported in Tables 2 show that students agreed on effect of mobile service, which made PA activities more convenient and fairer.

**Table 3. Analysis results for the effect of mobile service**

<table>
<thead>
<tr>
<th>Test value</th>
<th>( t )</th>
<th>( df )</th>
<th>Sig. (2-tailed)</th>
<th>Mean difference</th>
<th>95% Confidence Interval of the Difference</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using mobile service could reduce the cost of traffic during performance assessment.</td>
<td>6.948</td>
<td>46</td>
<td>.000</td>
<td>.809</td>
<td>.57 to 1.04</td>
<td>3.81</td>
<td>.798</td>
</tr>
<tr>
<td>Performance assessment activities could progress anytime anywhere by using mobile service.</td>
<td>9.936</td>
<td>46</td>
<td>.000</td>
<td>.979</td>
<td>.78 to 1.18</td>
<td>3.98</td>
<td>.675</td>
</tr>
<tr>
<td>It is fair to assess peers’ works regardless of the relationship with peers by using mobile service.</td>
<td>6.933</td>
<td>46</td>
<td>.000</td>
<td>.787</td>
<td>.56 to 1.02</td>
<td>3.79</td>
<td>.778</td>
</tr>
<tr>
<td>The limitation of the hardware location could be eliminated by using mobile service.</td>
<td>5.657</td>
<td>46</td>
<td>.000</td>
<td>.787</td>
<td>.51 to 1.07</td>
<td>3.79</td>
<td>.954</td>
</tr>
</tbody>
</table>

246
The third section of the questionnaire clarifies whether the use of AR technology to demonstrate students’ works during mobile PA is helpful. The section includes five items and associated hypotheses are listed below:

- **H0**: AR technology did not assist the activities of mobile performance assessment ($\mu \leq 3$)
- **H1**: AR technology assists the activities of mobile performance assessment ($\mu > 3$)

The experimental results in Table 3 reveal that participants thought using AR technology to enhance the work demonstration is helpful for students in the course.

<table>
<thead>
<tr>
<th>Item</th>
<th>Test value</th>
<th>$t$</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean difference</th>
<th>95% Confidence Interval of the Difference</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I could get information about work immediately by AR technology.</td>
<td>7.230</td>
<td>46</td>
<td>.000</td>
<td>.745</td>
<td>.54</td>
<td>.95</td>
<td>3.74</td>
<td>.706</td>
</tr>
<tr>
<td>I could review peers’ works more clearly by using AR technology.</td>
<td>6.680</td>
<td>46</td>
<td>.000</td>
<td>.766</td>
<td>.54</td>
<td>1.00</td>
<td>3.77</td>
<td>.786</td>
</tr>
<tr>
<td>I could know the future application of peers’ works by using AR technology.</td>
<td>6.203</td>
<td>46</td>
<td>.000</td>
<td>.723</td>
<td>.49</td>
<td>.96</td>
<td>3.72</td>
<td>.800</td>
</tr>
<tr>
<td>I could acquire detailed information about work to assess accurately by using AR technology.</td>
<td>3.545</td>
<td>46</td>
<td>.001</td>
<td>.468</td>
<td>.20</td>
<td>.73</td>
<td>3.47</td>
<td>.905</td>
</tr>
<tr>
<td>I could know how work was created by using AR technology.</td>
<td>6.424</td>
<td>46</td>
<td>.000</td>
<td>.702</td>
<td>.48</td>
<td>.92</td>
<td>3.70</td>
<td>.749</td>
</tr>
</tbody>
</table>

The results demonstrate that PA activities enable students to understand the teacher’s requests and engage their attention in learning. In addition, PA activities increase positive interaction through their discussions with peers. And regarding the attitudes toward the use of the system, most students believed that PA helps them to acquire more information about self-work and peers’ works and helps them to propose their viewpoints anytime anywhere by using the mobile service and AR technology.

For the progress of assessment grade, the average grade of peer assessment that had used MARPAS was 3.17, while the average grade of the 57 students who used pen-and-paper assessment was 2.43. To test the significance of this difference we have applied Mann-Whitney-Wilcoxon non-parametric test (since grades were not distributed normally), which showed that this difference is significant ($U = 662, p = .02$) in Table 5 and Figure 5.

<table>
<thead>
<tr>
<th>Method</th>
<th>Grade</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARPAS</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Pen-and-paper</td>
<td>14</td>
<td>10</td>
</tr>
</tbody>
</table>

**Figure 5. Aggregate results on two assessments**

Through the experiment between pen-and-paper assessment and MARPAS assessment, the data collected in the survey demonstrated that students not only agreed about the positive effect of AR and mobile service, but also confirmed the usefulness of AR technology in learning. Students that used MARPAS had much better achievement, average $M = 8.95$ versus $M = 5.79$ points, (see Figure 5). This difference is also significant (test statistics $U = 426$, significance $p = .00$). However, this difference can be attributed to the fact that PA activities
enabled students to understand the teacher’s requests and increase positive interaction through their discussions with peers. Most students believed that PA helped them to acquire more information about self-work and peers’ works and helped them to propose their viewpoints anytime anywhere by using the mobile service and AR technology. In addition, the qualitative analysis revealed that most participants thought that the system provided high autonomy and good visual effects. Importantly, the system helped students in acquiring rich and proper information while reviewing work, interacting with peers, and receiving assessment results. Moreover, use of the mobile service enabled students to propose their viewpoints anytime anywhere. The approach also eliminated the limitation of time, space, and devices.

By contrast, most students expressed that it was inconvenient to interact with peers in pen-and-paper peer assessment, and thus no student would like to use it. After the experiment, these students continued using MARPAS to interact with peers. Students that assessed their homework traditionally could collaborate, but they were not encouraged to do it and usually they worked independently, even on complex designs.

For the qualitative analysis, the feedback from the interviews indicated the following opinions about the use of MARPAS.

M01: “I can get the procedure and images of cooker on the work immediately! It’s so cool. I like this way of demonstration!”
M02: “There were no classmates beside when I assessed works by using this system. There was no pressure at all. And blind assessment could make me tell the truth.”
M07: “I am able to operate this system at home or at school. Even on the way to school in traffic, I can do it, too. This is very convenient. This convenience motivates me to give opinions to other classmates, and I hope to receive suggestions from classmates, too.”
M09: “Because of my inarticulateness, I cannot explain my works well. Now by using this system, I could prepare my illustration information and combine it with my works in advance. This is great to me. This will make me feel more confident about my works.”
M03: “I am not good at memory, and I forgot things quickly. The information showed on the works can be read repeatedly by me. So I won’t forget the detail of the works. This would make me do the correct assessment.”
M04: “The virtual information about the work can be procedure recorded video and even 3D digital model. It showed me things I have never seen before and things I would not see in physical work. This might affect my assessment.”

The above examples clearly show that AR technology did assist students during the assessment process.

Analysis of the interviews revealed that most participants credited the system with five advantages: high autonomy, good visual effects, prompt responses and rich assessment information from all assessors, convenient content management, and flexibility in using the system anytime anywhere.

Teachers provided the following feedback on the use of MARPAS:
T01: This system can assist teachers to arrange assessment activities more flexibly and make students more attentive to presentation, interaction, and feedback in the assessment process.
T02: For teachers, we can know how students review peers’ work from the AR situations and give more feedback to assessors and assessees. Moreover, teachers can create integrated AR situations for further applications.

**Conclusion**

This paper presents the novel system of an enhanced PA system complemented by the use of smart and mobile devices. Integrating the AR technology overcame some of the limitations of conventional PA systems, such as the implementation method, excessively high costs, and substantial errors. In this system, 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 PA can be improved.

Students who used MARPAS were encouraged to use its collaborative work feature. All students worked independently on their works. More complex cooking skill required students to know and combine many isometric exercises. Students who were using MARPAS were asked to do these exercises collaboratively, in groups of five students. Students usually communicated only with other members of their group, and had asked questions that other group members answered.

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. In instructional view, this system integrated the AR technology to overcome some of the limitations of conventional PA systems, such as subjective judgments of the evaluator, limited reliability, unsatisfactory economic performance, and low validity.

Limitation and future plan

“While it is important and useful to measure technical expectations, educators must beware of a blinkered vision that limits their ability to recognise and assess the expressive qualities that are enabled when technical skills are in place,” according to DeLuca and Bolden (2014). In fact, the National Research Council advocated in a recent report that moving to assessments of the next standard of science must be led by classroom-based assessments rather than by this complex endeavour with large-scale assessments (Pellegrino et al., 2014). The mobile AR technology is particularly suited to pilot this new approach, given their intensive efforts in implementing complex performance assessments. The information learned through competency-based assessments would then be used to support accountability determinations and, hopefully, better inform school improvement (e.g., Hargreaves & Braun, 2013). Thus, we will apply this system to some other courses, such as scene design, architecture, art design, and so on. Since different courses need different ways of PA, and this study did not undergo this research. Thus the dimensions of assessment should be studied as the next step, and we will use experimental research design to strengthen this study. And the experiment about how MARPAS assists teacher to manage the performance assessment and the course will be enforced.

Sometimes it’s hard to record the whole process of students’ performance by oneself. Students need a photographer for them if they are not working in a group. But, new technical devices such as Google Glass can solve this problem in the near future. This is convenient for all kind of students and courses.

References


