

Development of an interactive mathematics learning system based on a two-tier test diagnostic and guiding strategy

Tzu-Chi Yang, Hseng-Tz Fu

National Central University

Gwo-Jen Hwang

National Taiwan University of Science and Technology

Stephen J. H. Yang

National Central University

Mathematical skills have been recognised as a core competence for engineering and science students. However, learning mathematics has been recognised as a difficult and challenging task for most students, in particular, calculus for first-year students in university. Consequently, the development of effective learning strategies and environments for mathematics courses has become an important issue. To this end, a mathematics learning system based on an instant diagnostic and guiding strategy is proposed to enhance students' calculus learning outcomes. Moreover, an experiment has been conducted in a university calculus course to evaluate the effectiveness of the proposed method. The experimental results show that the proposed approach not only improved the students' learning performance, but also improved their confidence in learning calculus. Further findings are also discussed.

Introduction

Mathematical skills have been recognised as a core competence for engineering and science students in university. Unfortunately, many mathematics, science, and engineering departments have observed a qualitative decline in the mathematical readiness of their first year students (Burton, 1989). Moreover, several studies have reported that many university students have great difficulty learning mathematics (Jourdan, Cretchley, & Passmore, 2007; Lawson, 2003). Consequently, researchers have argued the importance of exploring effective instructional approaches to teaching mathematics, such as new tools, pedagogical approaches, and models or methods (Lopez-Morteo & López, 2007; Pettersson & Scheja, 2008; Szendrei, 1996). Developing effective strategies or tools for teaching mathematics courses has become an important and challenging issue, especially for calculus at the higher education level (Le, Joordens, Chrysostomou, & Grinnell, 2010).

In the past decade, several attempts have been made to develop effective strategies to help students improve their mathematical skills (Roth, Ivanchenko, & Record, 2008), in particular, the use of online testing, which is an increasingly popular assessment approach in higher education for teaching mathematics (e.g., Ellis, Ginns, & Piggott, 2009). Scholars regard online testing as an effective and efficient way of improving students' learning performance by identifying their weaknesses and providing instant feedback to promote their motivation and their confidence in the subject (Gikandi, Morrow, & Davis, 2011).

Meanwhile, the problem of applying conventional online testing, which uses multiple-choice questions to evaluate students' learning performance, has been noted. It is quite possible that the students who correctly answer a multiple-choice question only have partial understanding of the relevant mathematical concept, or may even totally misunderstand the concept (Sealey, 2014); that is, it is difficult to use conventional online tests to precisely examine students' understanding. Researchers have further argued the importance of diagnosing students' misconceptions and delivering remedies rather than merely evaluating their learning achievement (Nyachwaya et al., 2011; Panjaburee, Triampo, Hwang, Chuedoung, & Triampo, 2013). Accordingly, two-tier testing has been recommended and widely adopted as an approach to diagnosing students' misconceptions, owing to both its efficiency and effectiveness (Chu, Hwang, Tsai, & Tseng, 2010; Wang, 2014). Incorporating the two-tier test mechanism in a learning guidance system allows us to not only examine students' reasoning processes so as to accurately diagnose their learning problems or misunderstandings, but also to provide corresponding feedback and remedies to help them overcome the problems. Therefore, this study aimed to provide an online mathematics learning system based on a two-

tier test strategy for improving students' learning, and to evaluate the effectiveness of the proposed approach by answering the following questions:

- Do the students who learn with the proposed approach show better learning achievement in mathematics than those who learn with the conventional online learning approach?
- Is there a difference in the confidence, motivation and anxiety of the students who learn with the proposed approach and those who learn with the conventional online learning approach when learning mathematics?

Literature review

Online learning and testing systems for mathematics

Previous research has indicated that there is a need to improve university students' mathematical skills; for example, Lawson (2003) noted a significant drop in university pre-calculus diagnostic scores; Jourdan, Cretchley and Passmore (2007) reported that students experienced substantial challenges in algebra, functions, and trigonometry. Kay and Kletschin (2012) further noted the decline in students' ability in basic algebra, trigonometry, exponents, and logarithms in higher education. To cope with these problems, various teaching strategies and learning activities have been applied in mathematics courses using computer and network technologies. For example, Lopez-Morteo and López (2007) created an online practicing environment for learning mathematics; Roth et al. (2008) reported the benefits of employing an online testing and feedback system for helping graduate students learn calculus. Le et al. (2010) also demonstrated the positive effects of an online learning environment on students' mathematics learning performance.

Among various approaches implemented in online learning environments, assessment-based learning strategies have been widely adopted (Gikandi et al., 2011). Cox, Imrie, Miller and Miller (2014) indicated that assessment could be of various forms, such as achievement tests, artwork evaluation, or learning portfolio assessment. Many researchers, such as Rod and Allan (2006) and Karaman (2011), have pointed out that online testing is an efficient approach to identifying students' knowledge levels and learning problems. Kay and Kletschin (2012) further reported the advantages of using online testing in calculus courses for facilitating practice and providing instant feedback.

On the other hand, the study of Sealey (2014) revealed that most students were able to find answers using mathematics formulae, but were unaware of the physical meanings of the calculations. In other words, evaluating students' understanding of a course by only judging their answers and scores possibly fails to reflect their actual understanding, revealing the limitation of conventional online test systems. In addition, researchers have indicated several factors that might affect students' online learning outcomes, such as computer literacy (DeBoer et al., 2014), online learning strategies (Zlatović, Balaban, & Kermek, 2015), and attitudes towards online testing (Stratling, 2015). Chu (2014) further pointed out the limitations of testing in terms of improving students' ability to solve real-world problems. Therefore, it is necessary to develop effective online testing systems by considering these factors.

The two-tier test approach

Diagnostic testing is an approach to examining students' learning problems or misconceptions via well-designed tests (Black & Wiliam, 1998). Black and Wiliam (1998) as well as other scholars (e.g., Treagust, 1995) have addressed the effectiveness of diagnostic tests in helping teachers evaluate individual students' learning status, and providing them with appropriate assistance.

Two-tier testing is a widely adopted diagnostic testing approach proposed by Treagust (1988). It involves tests consisting of a set of two-tier multiple-choice questions. The first tier assesses students' descriptive or factual knowledge of the phenomenon to be assessed, while the second tier investigates the reasons for their choice made in the first tier. Treagust, Jacobowitz, Gallagher and Parker (2001) indicated that, via exploring students' in-depth explanations of factual knowledge, teachers or researchers are able to assess in-depth understanding of the learning content. In the past decade, several applications of two-tier tests have shown the benefits of this approach. For example, Chou, Chan and Wu (2007) employed two-tier testing in an online learning system to assess students' understanding and alternative conceptions of cyber copyright

laws, and reported the feasibility and efficiency of the approach. Chu et al. (2010) further employed two-tier testing in a natural science course and reported the positive impacts of the approach on students' learning achievements.

Researchers have pointed out that solving a mathematics problem usually requires a certain breadth of knowledge such as knowing the content, generating relationships, and analysing claims (Ball, Thames, & Phelps, 2008; Webb, 1977), as well as showing the importance of examining the problem-solving logic. Andrade-Aréchiga, López and López-Morteo (2012) also emphasised the need to examine what students think about their solutions to mathematics problems. Zandieh, Roh and Knapp (2014) further indicated that examining students' reasoning plays a key role in calculus courses as it can precisely diagnose their learning problems.

To address this problem, the approach in this study did not fully engage with Treagust's work on two-tier testing, which was mainly proposed for assessment purposes. The approach adopted in the current study is a test-and-feedback learning guiding mechanism based on Treagust's two-tier tests. An online test and guidance system was developed based on this proposed approach. Previous studies (e.g., Hwang, Yang, & Wang, 2013; Ibabe & Jauregizar, 2010) have reported the potential of such a testing system for improving students' learning achievement, and attitudes (e.g., confidence, motivation, and anxiety), as well as perceptions (e.g., perceived usefulness and perceived ease of use). Therefore, to evaluate the efficacy of the proposed approach, an experiment was conducted to investigate students' learning achievements, affective factors of learning calculus, and their perceptions of the system for learning calculus.

Development of a mathematics learning system with a diagnostic and guiding mechanism

Based on the two-tier test approach, a web-based mathematics learning system was developed as shown in Figure 1. The system consists of the online test subsystem (OTS), the diagnosis subsystem (DS), and the learning content interactive subsystem (LCIS). The system also includes three databases: a test-item database, a diagnosing rule database, and a learning materials database. The OTS provides students with web-based exercises, practice, and tests corresponding to the class content. Accordingly, in the experimental group, the diagnosis subsystem judges students' answers and reasons to clarify their misconceptions or misunderstandings. The LCIS then generates feedback including answers to the test, suggestions and corresponding learning content. Moreover, the system enables students to browse their test results and related learning materials whenever they want. Each subsystem is described as follows.

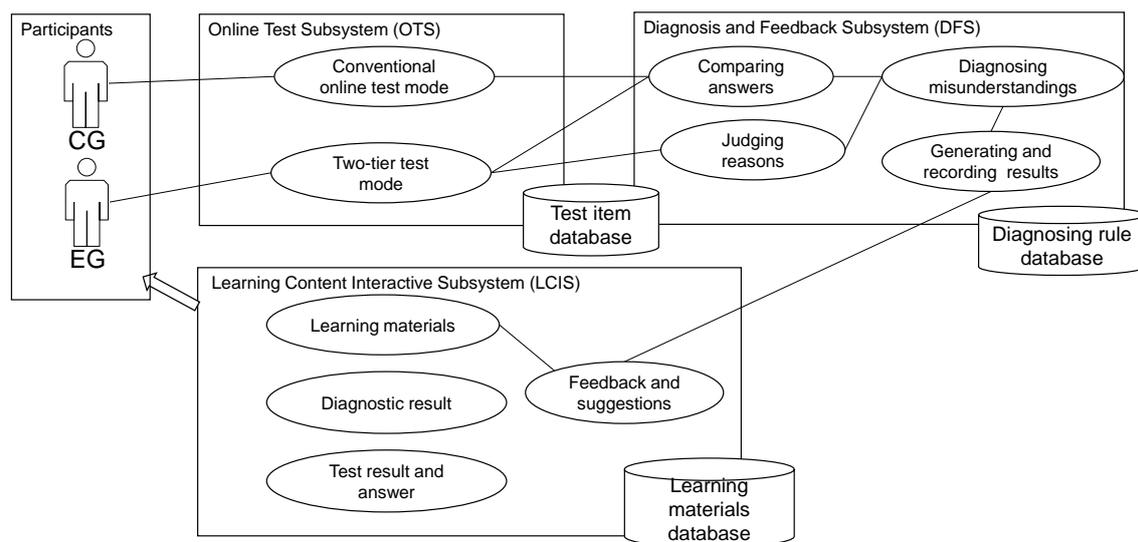


Figure 1. Structure of the two-tier test-based mathematics learning system

Online test facility

In order to prompt the students to use the system, after learning a unit or concept, they were asked to complete at least one test before continuing on to the next unit. The highest quiz scores were recorded as their homework grades. The online test subsystem generated two kinds of test: (1) conventional multiple-choice questions selected from the test item database; and (2) a two-tier test in which every test item had three or four choices in the first tier, followed by three or four reasons for each choice in the second tier. Table 1 and Table 2 show an example of a two-tier test item for a chain rule in calculus. Table 1 shows the first-tier multiple-choice item, which provides three choices (i.e., $3a^3 - 3 \cos^2 x \sin x$, $3 \cos 2x$, and $-3 \cos 2x \sin x$). Table 2 shows the second-tier items developed based on the first-tier choices. As three additional choices are used to assess the reasons for each item, a total of nine choices were developed for the second-tier items.

Table 1
Illustrative example of a conventional multiple-choice item

First-tier item	Choices
If $y = a^3 + \cos^3 x$, a is a constant, then $\frac{dy}{dx} = ?$	$3a^3 - 3 \cos^2 x \sin x$
	$3 \cos 2x$
	$-3 \cos 2x \sin x$

Table 2
Illustrative example of the second tier options (i.e., identifying the reason for the answer)

Second-tier item	Choices
Why did you choose $3a^3 - 3 \cos 2x \sin x$?	(1) The derivative of a constant is the same rule as for functions.
	(2) The derivative of a constant is zero.
	(3) The derivative of a constant is still a constant.
Why did you choose $3 \cos 2x$?	(1) The derivative of x^3 is $3x^2$
	(2) The derivative of a constant is zero.
	(3) The derivative of Y^3 is $3Y^2Y''$.
Why did you choose $-3 \cos 2x \sin x$?	(1) Because of the chain rule.
	(2) Because $(\cos 3x)' = 3 \cos 2x$
	(3) Because $(\cos 3x)' = 3 \cos 2x \sin x$

To prevent students from memorising the answers, the sequence of the choices in the test items was changed when generating test sheets. That is, few identical test sheets were generated during the learning process. When making choices in the second tier, the students were allowed to view the corresponding first-tier item to identify the reason for making their choice. The scope of the tests was related to the weekly course content. However, the students were able to extend the test scope by including more previously learned concepts or units. Figure 2 shows the system interfaces of the conventional (multiple-choice) online test and the two-tier test. Figure 3 shows the test results and feedback from the learning system.

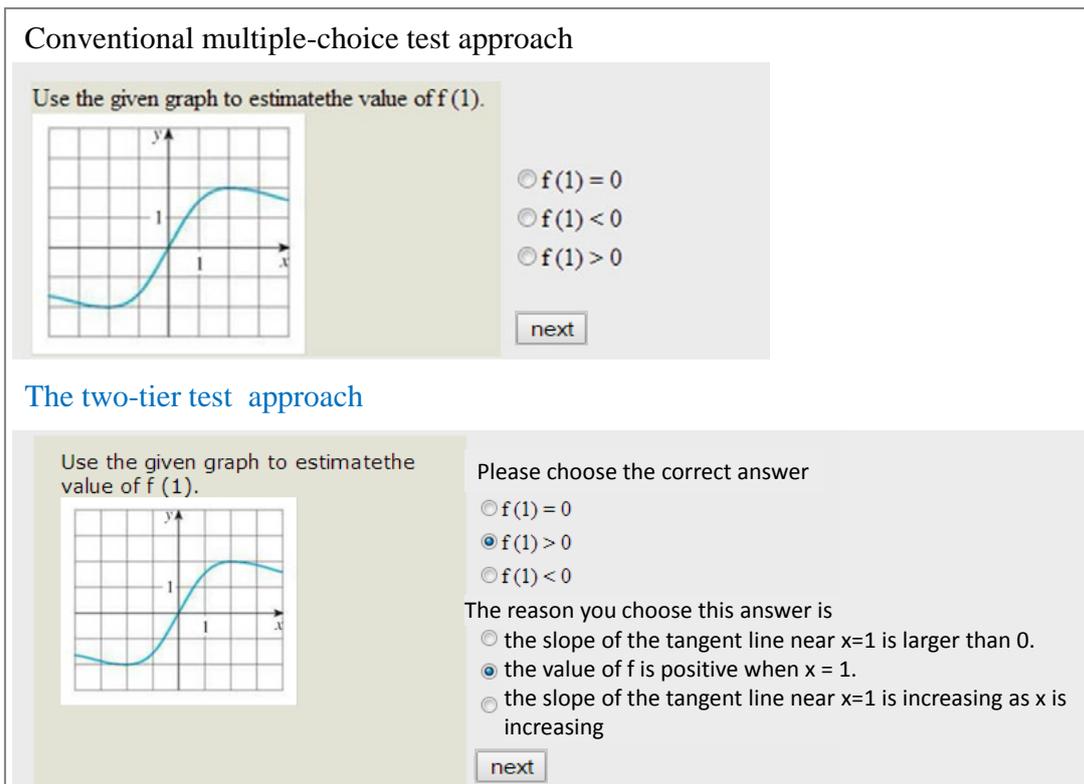


Figure 2. A comparison of a conventional online test and a two-tier test item

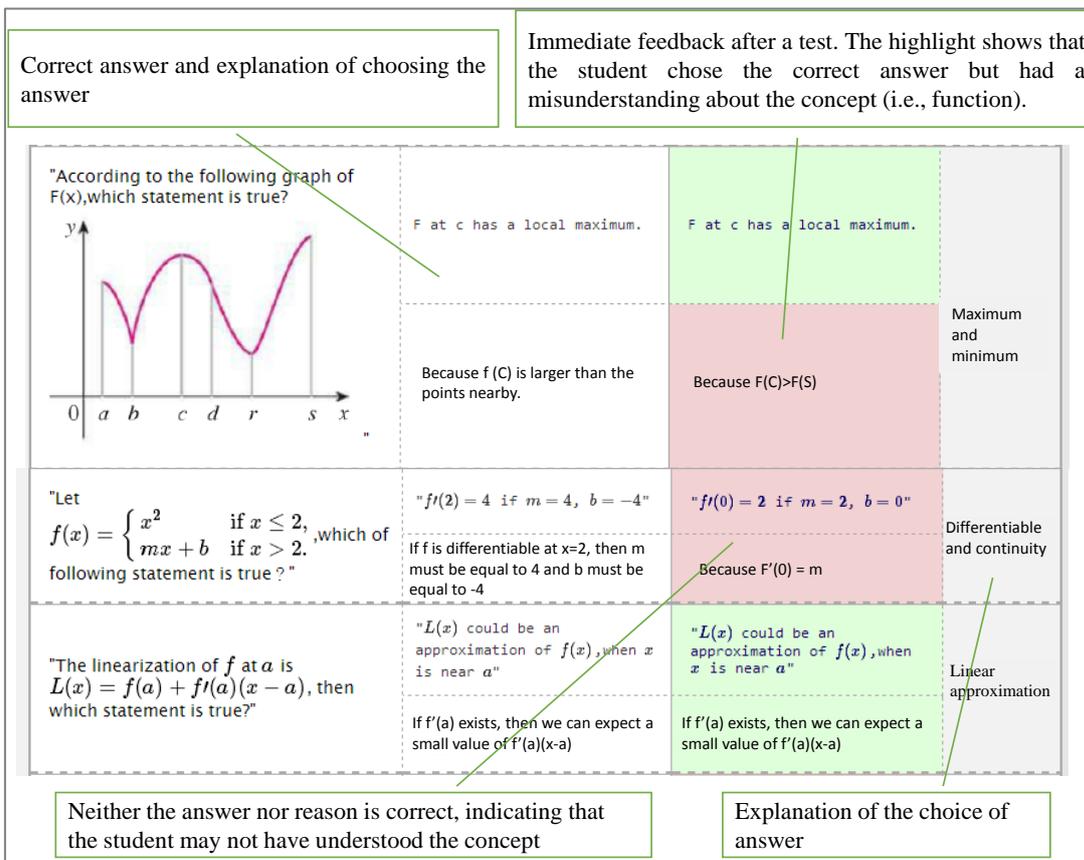


Figure 3. Illustrative example of a test result and feedback

Learning diagnosis and feedback facility

In the diagnosis and feedback subsystem (DFS), the first-tier item is used to assess the students' descriptive or factual knowledge of calculus, while the second tier probes the students' reasons for their choices made in the first tier. Hence, the choice made in the second-tier item not only confirms whether the students have fully understood a concept, but also helps them to explore their misunderstandings and misconceptions if incorrect reasons are selected. On this basis, a set of rules in a decision-tree is used in DFS for diagnosing students' misunderstandings or learning problems.

According to the decision-tree rules, the DFS can precisely ascertain students' learning status (e.g., misunderstanding, alternative concepts, and learning problems). An illustration of the diagnosis processing for judging the students' learning problems based on their answers is given in Figure 4. In this example, two concepts were tested, namely, the chain rule and constant differentiating. If students choose the first item, $3a^3 - 3 \cos^2 x \sin x$, it means they understand the chain rule but do not notice, or do not understand the concept of constant differentiating. If they choose the first item in the second tier, that is, if they think the symbol constant is differentiated as with functions, they may have a misconception of the meaning of symbol constants. If they choose the second item in the second tier, it means they know the differentiating of a constant is 0, but may have wrong concepts about the symbol constant. If they choose the third item in the second tier, it means they have wrong concepts about the differentiating of the constant. In other words, if students fail to answer a question correctly, the diagnosis subsystem not only identifies the misconception, but also exactly what the reason for the misconception is.

In addition, to provide feedback and suggest materials, a learning content recommending rule (LCRC), adapted from the Lossy Counting Algorithm (Manku & Motwani, 2002), was implemented, as shown in Figure 5. The algorithm is a data mining approach which aims to identify frequent patterns from a set of data that records the mistakes the students have made in chronological order. The concepts related to each test item that the students failed to answer correctly were provided as data for the LCRC. Due to the limited number of test records, the threshold of LCRC was set to 2 in this study. According to the rule, if a student failed a concept more than two times during a test, or failed more than three times in two consecutive tests, the concept would be recognised as a frequent mistake or misconception. As such, the learning content related to the misconception could be precisely selected and subsequently presented in the feedback with strong recommendations that the student read it. Otherwise, if there were no frequent mistakes or misconceptions flagged, those concepts which the student failed but which were not identified as frequent misconceptions, were presented in the feedback as review suggestions.

To ensure the reliability of the test items for the two-tier test, two postdoctoral researchers with more than 3 years' experience teaching calculus were employed to construct the test items for the system. Subsequently, all test items were reviewed by two university professors who had been teaching the calculus course for more than 5 years. Further, all ambiguous sentences were revised in order to enhance the efficiency of the test.

Learning content organising and presentation facility

The learning content interactive subsystem (LCIS) is mainly used to present the learning content based on the students' requirements. A list of concepts was provided to allow the students to select and compose some units as an e-book in PDF or webpage format. Thus, the students were able to review the relevant contents based on their requirements. Moreover, based on the students' learning status as evaluated by the DFS, the LCIS could highlight concepts they may not be familiar with and remind them to read the recommended learning materials. After doing so, students may overcome their misconceptions by studying relevant learning materials suggested by the system. In sum, the LCIS provides corresponding feedback and learning contents for students based on their learning status. Figure 6 shows an illustrative example of providing learning materials to individual students based on their test results and needs.

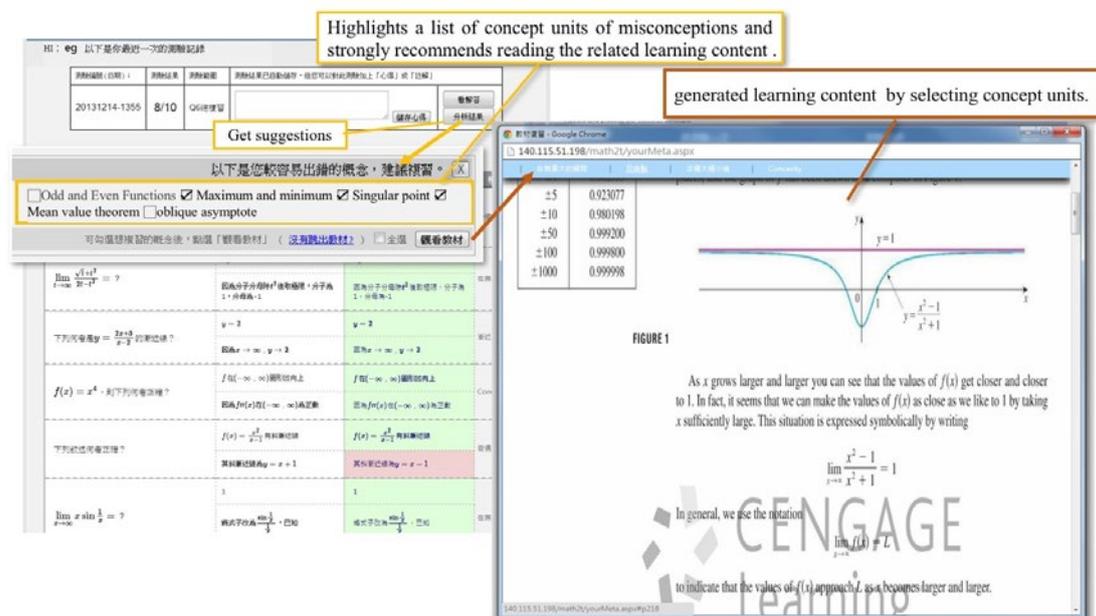


Figure 6. Illustrative example of providing personalised comments and learning materials

Experiment design

The experiment was conducted on the learning activities of the Introduction to Calculus course of a university in northern Taiwan. In the following sections, the details of the experiment are presented.

Participants

A total of 55 university students participated in the online programming activity. The average age of the students was 19. They were randomly divided into an experimental group and a control group. Eventually, 43 students completed all of the activities. There were 22 in the experimental group, and 21 in the control group. All the students were taught by the same teacher and used the same learning materials.

Experiment procedure

A first year university calculus course was taken as the research context for this study. The learning objectives included understanding functions and limits, derivatives, applications of differentiation, integrals,

application of integration, inverse functions, etc.

A total of 119 two-tier test items were administered in this study, covering functions and limits, derivatives, applications of differentiation, integrals, application of integration and inverse functions. To encourage students to participate in the online learning activity, the learning contents in class (i.e., book and examples) were segmented into many learning content units based on the corresponding concepts. Moreover, a checkbox list of the learning content units was provided to the students, who could select the units they wanted to read. These selected units were then composed as e-books in PDF format or as webpages. In addition, the test items were periodically modified to prevent generating identical test sheets during the learning process.

After taking the quiz, a diagnostic subsystem was employed to provide feedback to the students based on their quiz results. We believed that providing up-to-date content and useful information could help maintain the students' attention during the online course. Figure 7 presents the experimental procedure of this study. Both groups of students first received instruction on the basic knowledge of calculus, and then took a pre-test and completed a questionnaire to analyse their knowledge and perceptions (i.e., their confidence, motivation, and anxiety) of calculus before interacting with the proposed system.

In the first stage of the activity, both groups of students received face-to-face classroom instruction. The teacher provided instruction on the basic knowledge of calculus. The students were then asked to complete exercises based on what they had learned in the classroom.

In the second stage, both groups of students were asked to take a test on the system after each class. In general, they completed at least five exercises and one test each week. In this stage, the experimental group learned with the two-tier test-based learning system. On the other hand, the students in the control group learned with the conventional online learning approach; that is, they browsed the same learning materials and did the same exercises, but took conventional tests (multiple-choice) and received feedback in the web-based learning system. The learning activity was conducted for a period of 6 weeks for both groups after which both groups took a post-test. They were then asked to fill in the same questionnaire to measure if their attitudes towards calculus had changed.

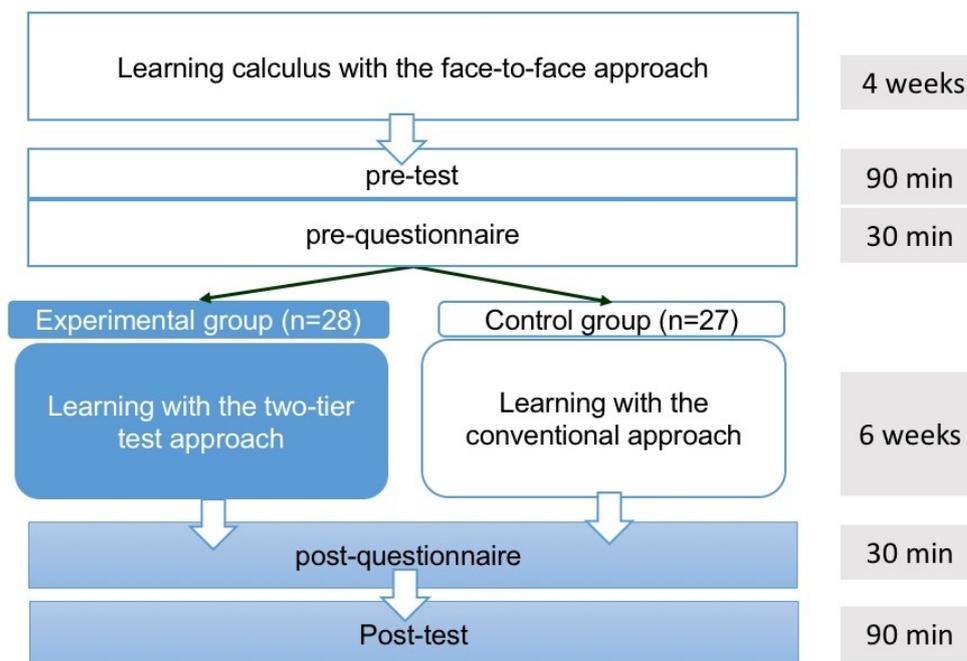


Figure 7. Experimental procedure

Instruments

Pre-test and post-test

Four types of test item were used to evaluate the students' learning performance; namely, finding or evaluating a value of an equation, correctly using a rule, proving an equation, and inducing a line/curve/equation. In order to further compare the learning performance of the students in the two groups, the revised Bloom's taxonomy of the cognitive process dimension (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956) was employed to classify the four question types. The former two were concerned with the students' understanding of concepts. To answer these two types of question, the students needed to recall or recognise the concepts or principles in the learning content, or interpret the learning contents based on their prior learning. According to the revised taxonomy, finding an answer to these types of question using specific principles is referred to as the *remember* (e.g., recalling principles), *understand* (e.g., comprehending information) and *apply* (e.g., selecting and using principles) abilities, which are recognised as the basic mathematical knowledge in this study. Figure 8 shows an illustrative example of answering these types of calculus question.

6. Find y'' if $y = x + \sqrt{x}$.

$$y = x + \sqrt{x}$$

$$= x + x^{\frac{1}{2}}$$

$$y' = 1 + \frac{1}{2}x^{-\frac{1}{2}}$$

$$y'' = \frac{-1}{4}x^{-\frac{3}{2}}$$

#

Figure 8. An example of a test item for evaluating the basic knowledge of calculus

To answer the latter two types of question, students need to reorganise their prior knowledge and select appropriate principles by considering the criteria for solving the problems, which are relevant to the *analyse* (e.g., organise) and *evaluate* (e.g., check and critique) competences, and are recognised as advanced mathematical skills in this study. Figure 9 shows an illustrative example of a test item for evaluating advanced mathematical skills in calculus.

The pre-test consisted of 15 items with a total score of 105. It aimed to evaluate the students' basic knowledge of calculus (e.g., functions and limits, derivatives) before the learning activity. The post-test, a further achievement test for calculus, also consisted of 15 items with a perfect score of 105, of which 10 items were to evaluate basic knowledge such as finding the value of a limit, while the other 5 examined further mathematical skills such as applying definitions for proving an equation. Both the pre- and the post-test were developed by the Calculus Centre of the university, which specifically teaches calculus and evaluates the calculus learning performance of all first year students.

An analysis of covariance (ANCOVA) was used to compare the two groups' test scores with regard to their learning performance while controlling for the pre-test scores. That is, the impact of the pre-test scores can be excluded when comparing the two groups' post-test scores. Before applying ANCOVA, the Levene's test for equality of variances was used to confirm the homogeneity of the samples.

7. Given the facts $\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1$ and $\lim_{h \rightarrow 0} \frac{1 - \cos h}{h} = 0$, use the definition of derivative to show that $\frac{d}{dx} \sin x = \cos x$.

$$\begin{aligned} \frac{d}{dx} \sin x &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin x}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sin x \cdot \cos h + \cos x \cdot \sin h - \sin x}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sin x \cdot (\cos h - 1) + \cos x \cdot \sin h}{h} \\ &= \lim_{h \rightarrow 0} \left(\sin x \cdot \frac{1 - \cos h}{h} + \cos x \cdot \frac{\sin h}{h} \right) \\ &= \lim_{h \rightarrow 0} (-\sin x) \cdot \lim_{h \rightarrow 0} \frac{1 - \cos h}{h} + \lim_{h \rightarrow 0} \cos x \cdot \lim_{h \rightarrow 0} \frac{\sin h}{h} \\ &= -\sin x \cdot 0 + \cos x \cdot 1 \\ &= \cos x \end{aligned}$$

Figure 9. An example of a test item for evaluating advanced mathematical skills in calculus

Questionnaires

Previous studies have highlighted that online testing could improve students' motivation and confidence (Denvir & Brown, 1987; Nguyen & Kulm, 2005) as well as reduce their anxiety in learning mathematics (Nguyen, Hsieh, & Allen, 2006). In this study, the questionnaire developed by Fennema and Sherman (1977) was adopted. It consisted of three dimensions, that is, confidence (e.g., I am sure that I can learn math), motivation (e.g., I would like to think about the exercise that I cannot resolve in the class) and anxiety (e.g., I feel good when I'm in the math class) in learning mathematics. Each dimension consisted of 12 items. Another questionnaire adopted in this study was the questionnaire of technology acceptance developed by Hwang et al. (2013). It included two dimensions, *perceived usefulness* and *perceived ease-of-use*, each of which consisted of 14 items.

A 5-point Likert scale was used to rate the items in the two questionnaires, where 5 meant strong agreement or positive feedback, and 1 represented high disagreement or negative feedback. Moreover, the Cronbach's alpha value of the questionnaire was 0.89, showing the good reliability of this questionnaire in internal consistency.

In addition, the students were asked to answer an open-ended question, "Do you think the online test system benefits you in terms of learning calculus? If so, please write three sentences to describe how. If not, please also give your reasons." Via collecting these qualitative data, the researchers could obtain additional information related to whether the students agreed that the system was helpful to them in learning calculus, as well as their perceptions of learning with the proposed system.

Results and discussion

Learning achievement

The Levene's test for equality of variances showed that the groups were homogeneous with $F(1, 42) = 0.051, p = .822$, confirming the homogeneity of the regression coefficient. The ANCOVA result shows that the students in the experimental group had significantly better learning performance than those in the

control group, as shown in Table 3, implying that the two-tier test approach benefits students learning calculus. One possible reason is that those students who learned with the two-tier test approach were asked to provide reasons for their choices, meaning that they were not only prompted to recall what they had learnt, but that they also had to evaluate their solutions. Thus, they developed advanced knowledge via the guidance of the approach, and demonstrated better learning achievement than the students in the control group. Along the same lines as earlier research (Gálvez, Guzmán, & Conejo, 2009), such meaningful and efficient feedback benefits students in terms of understanding the learning content and overcoming misconceptions. Another possible reason is that the students were guided to recall what they had learnt and to reorganise their knowledge. In the meantime, the feedback mechanism of the system provided hints; therefore, the students who used the two-tier test system improved their learning performance. This finding has also been argued by previous research in this field (e.g., Gaytan & McEwen, 2007). That is, such a self-evaluating test approach and the guidance strategy possibly had an interactive effect on the students' learning outcomes. The results also echo previous studies in this field, which found that such learning guidance or prompts benefit students in science or mathematics courses (Chu et al., 2010; Panjaburee et al., 2013).

Table 3
ANCOVA result of the post-test scores of the two groups

Group	<i>N</i>	<i>M</i>	<i>SD</i>	Adjusted Mean	<i>SE</i>	<i>F</i>	<i>d</i>
Experimental group	22	84.64	11.97	84.33	2.45	4.79*	0.56
Control group	21	76.53	16.45	78.84	2.39		

* $p < .05$

In addition, the students who learned with the two-tier test approach showed significantly better learning performance at a higher cognitive level (i.e., inducing equations or target functions) than those who learned with the conventional approach, because providing the correct reason for the answer not only requires understanding a rule or definition, but also correctly explaining and using it. On the basis of the aforementioned, the learning performance of the two groups was further compared. The score of each post-test item was divided into two parts with regard to the comprehension level (i.e., items 1, 4, 6, 8, 9, 10, 11, 12, 13, and 14) and the application level (i.e., items 2, 3, 5, 7, and 15), as shown in Figure 10. Moreover, an ANCOVA with the pre-test as the covariate was employed to compare the learning performance of the two groups based on the comprehension level and application level.

According to the ANCOVA results shown in Table 4, the experimental group demonstrated higher learning performance at the application level than the control group. In other words, the two-tier test approach had a significant effect on developing the students' advanced mathematical skills in calculus (e.g., proving skills). This may be because both groups were provided with the same learning materials for the students to memorise and comprehend, which refers to the comprehension level, while the two-tier test further engaged them in not only frequently recalling the memorised concepts, but also in thinking about how to solve the problems by organising the concepts and rules. Therefore, those students who learned with the two-tier test approach exhibited better learning performance in solving application level problems.

The findings of this study not only confirm that learning with online testing could provide opportunities for students to self-evaluate their knowledge, and that it benefits their learning in mathematics (e.g., Farrell & Leung, 2004), but also further demonstrates that the proposed strategy may improve students' mathematics skills, in particular the high-level cognitive process skills required for solving mathematics questions. However, the effect size (Cohen's *d*) is 0.32, which indicates that the approach had minor effects on developing the students' mathematical skills in calculus. There is thus a need to further investigate the influence of the two-tier test approach on improving high level cognitive ability (i.e., the application level).

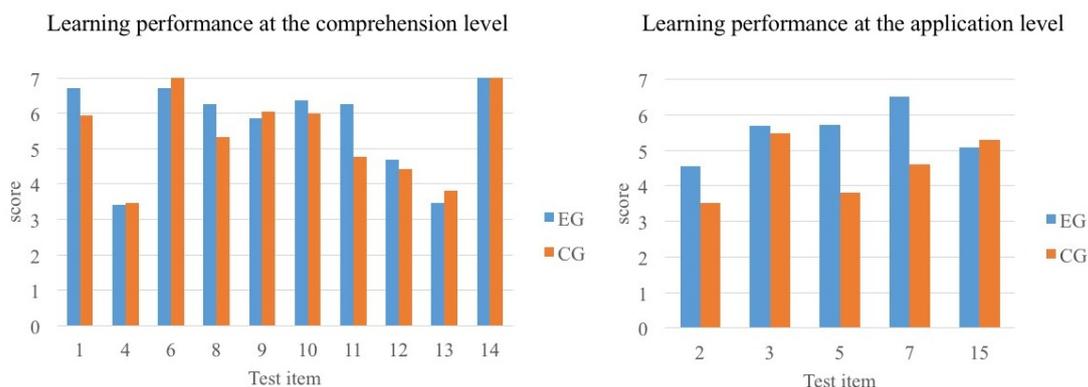


Figure 10. A comparison of learning performance at the comprehension and application levels

Table 4
The ANCOVA results of the post-test for the control group and experimental group

	Group	N	M	SD	Adjusted Mean	SE	F	d
Comprehension level	Experimental group	22	5.67	0.75	5.66	0.16	1.20	0.15
	Control group	21	5.38	1.10	5.40	0.17		
Application level	Experimental group	22	5.51	1.56	5.48	0.26	5.64*	0.32
	Control group	21	4.54	1.32	4.56	0.27		

* $p < .05$

Perceptions of learning calculus

Table 5 shows the ANCOVA result of the students' perceptions of learning calculus. It was found that the experimental group students showed significantly higher confidence than the control group after the learning activity. This finding conforms to previous studies' reports that effective learning guidance strategies or mechanisms are helpful to students in terms of improving their learning attitudes as well as their learning achievements (e.g., Chu et al., 2010; Hwang, Wu, Tseng, & Huang, 2011). Regarding the anxiety dimension, no significant difference was found between the two groups. This result is in agreement with previous research (e.g., Cassady, Budenz-Anders, Pavlechko, & Mock, 2001) which demonstrated that online testing did not have a significant impact on students' level of anxiety. Such a finding suggests that despite the two-tier test seeming to be more complex than the multiple-choice test, it does not increase students' anxiety. That is, the negative effect on students' anxiety could be eliminated when applying a two-tier test instead of a conventional multiple-choice test. However, there was no significant difference between the two groups' motivation. As Nguyen et al. (2006) argued that online testing might improve students' learning motivation, this issue could be further investigated in future research in this field.

Table 5
The post-questionnaire ANCOVA results of the experimental and control groups

	Group	N	M	SD	Adjusted Mean	SE	F
Confidence	experimental group	22	3.73	0.67	3.73	0.07	8.71**
	control group	21	3.41	0.57	3.41	0.08	
Motivation	experimental group	22	3.51	0.97	3.51	0.11	0.95
	control group	21	3.35	0.92	3.35	0.11	
Anxiety	experimental group	22	3.58	0.65	3.58	0.08	0.38
	control group	21	3.51	0.68	3.52	0.09	

** $p < .01$

Technology acceptance for the proposed system

The *t*-test result of the technology acceptance ratings revealed that the students in the experimental group perceived the system as being easier to use than did those in the control group, as shown in Table 6. It should be noted that the two-tier test approach may seem complex, but the students perceived it as being easy to use. One possible reason could be that the guiding method matched their learning strategies, and information was easy to find from the provided feedback and learning materials.

Table 6
The *t*-test results of the technical acceptance questionnaire for the control and experimental groups

	Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i>
Perceived usefulness	Experimental group	22	3.71	0.60	1.20
	Control group	21	3.38	0.64	
Perceived ease-of-use	Experimental group	22	5.51	1.56	5.64*
	Control group	21	4.54	1.32	

**p* < .05

Benefits of the learning system

The analysis results of the qualitative data showed that 17 (77.27%) students in the experimental group and 15 (71.43%) in the control group agreed that the proposed online test system benefited their learning of calculus. Regarding the students’ perceptions, five were classified based on their responses, i.e., promoting practice, providing reviewing guidance, improving understanding of concepts, finding learning problems and promoting deep thinking. The proportion of each viewpoint is shown in Figure 11.

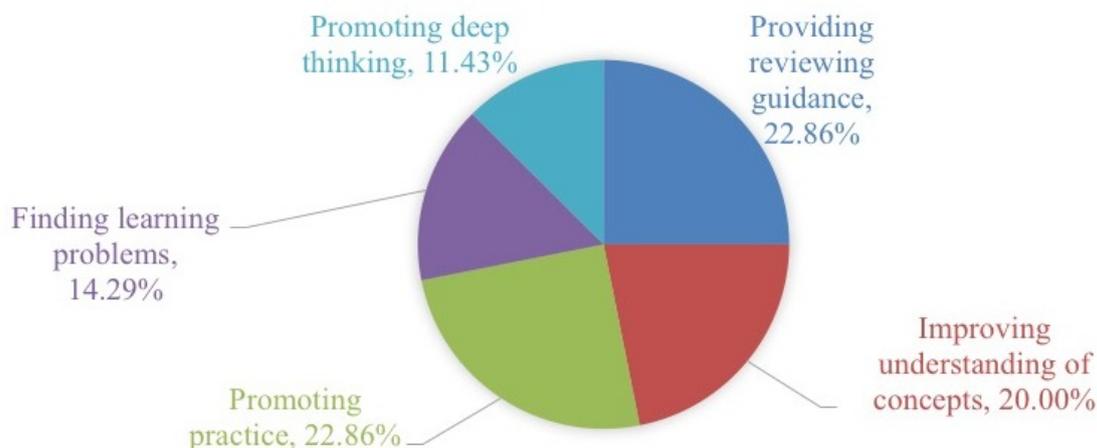


Figure 11. Distribution of the students’ perspectives on the benefits of the system for the two groups

According to the students’ responses, both the control group and the experimental group students shared the same perceptions that the online learning system engaged them in practice and helped them identify their learning problems. More specifically, more than one third of the students in both groups perceived that the learning approach could promote practice of what they had learnt. For example, one student said, “I can have a lot of practice”. Another student mentioned, “I will read the book on schedule and do more practice”, while one student stated, “By using this system, I can maintain my learning performance and get familiar with different types of questions”. Meanwhile, students in both groups shared a similar viewpoint regarding identifying their learning problems (27.27% for the experimental group; 28.57% for the control group). Typical responses included, “I can know which part I have to strengthen”; “I can know which part I have to improve”; and “This system helps me understand the problems I encounter”.

The students in the two groups also expressed some different perceptions, including providing reviewing guidance, improving understanding of concepts and encouraging deep thinking. Regarding providing reviewing guidance, 18.18% and 42.86% of students in the experimental group and the control group respectively indicated that the learning system provided a form of navigation for reviewing the learning contents. This finding shows that the students in the control group may have mainly considered the learning system as a reviewing tool. For example, one student in the control group indicated that, “I can review the lectures by doing quizzes”.

Regarding improving understanding of concepts, 40.91% of the experimental group students commented that their understanding of calculus concepts was improved owing to the use of the learning system, while only 14.29% of the control group students shared the same perception. This may be reflected by the fact that the experimental group students had better learning outcomes in solving higher-level questions. They were engaged in evaluating their knowledge when giving reasons for their answers. By doing so, they may have had more opportunities to review the contents and clarify some confusing issues. Typical responses included “The system helped me clarify the learning concepts” and “This system helped me understand the learning materials”. That is to say, compared with the control group, students in the experimental group perceived that the two-tier test not only provided a chance for them to practice, but also helped them improve their understanding. A possible reason for this improvement is that the students needed to give reasons for their answers. By doing so, they were forced to practice and think about what they had learned. During the process, they may have faced a cognitive conflict. Subsequently, their misunderstandings could be corrected and their understanding of calculus hence improved.

Regarding encouraging deep thinking, 31.82% of the students in the experimental group shared the same view that the learning system encouraged them to think about the reasons for their answers, while only 4.76% of the control group students indicated this point. This implies that the two-tier test approach engaged the students in deep thinking. In other words, they were encouraged to reflect on solving the calculus exercises. For example, one student emphasised that, “It helped me think about the questions critically and understand the questions”. Another student mentioned that, “solving problems in the right direction is the most important”. However, only 4.76% of the students in the control group shared the same perception.

The influences of using the two-tier test-based learning system

The aforementioned findings suggest that the students in the experimental group considered the benefits of the two-tier test-based learning system as improving understanding of concepts and prompting deep thinking, while those who learned with the multiple-choice test mainly considered it as promoting practice and providing reviewing guidance. This finding suggests that the two-tier test encouraged the students to think about what they had learnt rather than merely practice. It may also explain why those students who learned with the two-tier test performed better on the application level questions.

According to the qualitative results, it is concluded that the two-tier test approach benefited the students by correcting their misunderstandings of calculus and guiding them to effectively improve their advanced mathematical skills in calculus during the learning process. On the contrary, the students who learned with the conventional online test approach mainly focused on practicing and getting high scores in the exercise. This may explain why students who learned with the two-tier test approach improved their confidence and demonstrated higher learning achievement than those who learned with the conventional online test.

This finding suggests that the proposed learning system based on the two-tier test approach can improve students' advanced mathematical skills via enhancing their logical reasoning, and by helping them make connections between their prior knowledge (e.g., the definitions and rules) and meta-knowledge (e.g., selecting principles for proving equations). More specifically, the two-tier test-based learning system had effects on the students' learning, especially their learning perceptions and achievement.

Conclusions

In this paper, an interactive two-tier test diagnostic and guiding strategy is proposed, and a learning system with misconception diagnostics and a feedback mechanism was implemented based on the proposed strategy. To identify the influences of the proposed strategy, an experiment was conducted on a first year

university calculus course to evaluate the effectiveness of the proposed approach by comparing the learning performance of the students who learned with the two-tier test-based system and those who learned with the conventional technology-enhanced (multiple-choice test) learning approach.

From the experimental results, it was found that the two-tier test approach significantly improved the students' learning achievement, in particular, in solving the application level questions, showing the effectiveness of the proposed approach for helping students learn calculus at a higher cognitive level. In addition, by analysing the qualitative data, it was found that most of the students (77.27%) agreed that the use of the system benefited their learning of calculus in terms of improving their understanding of mathematics and encouraging them to think deeply about what they had learnt. Based on a prior study (Schoenfeld, 1988) which highlighted that high school and college students employ guess and test loops as a common problem solving strategy, the two-tier test approach seems to provide an alternative learning strategy for teachers to cope with this issue. In sum, the aforementioned findings suggest that the proposed learning system based on a two-tier test diagnostic and guiding strategy can enhance students' learning in mathematics courses.

Although the proposed system benefited the students in this application, there are some limitations to be noted. First, generalisation of the findings may be limited to populations of a similar nature, but may not be so applicable to other learner groups in different educational settings or with different cultural backgrounds. Second, the number of students is rather small, meaning that further studies need to be undertaken with a larger sample to provide stronger evidence in the future. Third, to provide feedback and judgment, the teachers need to spend time reviewing and refining the test items for evaluation purposes, as well as the digital learning materials for providing learning supports. Therefore, it is advisable to extend the approach to other learning areas, such as scientific and technological courses. It is also necessary to conduct long-term and large-scale experiments to evaluate the performance of the approach. Moreover, the provision of supporting systems for teachers could be helpful to them in developing two-tier test items and the related supplementary materials.

In addition, beyond the advantages, online tests may have some constraints. For example, technical issues and IT support for students could be problems when accessing online tests (Buchan & Swann, 2007). Meanwhile, it is more difficult to ensure that an online test is valid, reliable and fair than it is with pen-and-paper tests because students may cheat more due to feeling more distant, as indicated by Karman (2011). This study, however, did not uncover any effects of these issues. Thus, further research is needed which takes these issues into consideration when employing online test approaches.

References

- Andrade-Ar chiga, M., L pez, G., & L pez-Morteo, G. (2012). Assessing effectiveness of learning units under the teaching unit model in an undergraduate mathematics course. *Computers & Education*, 59(2), 594-606. <http://dx.doi.org/10.1016/j.compedu.2012.03.010>
- Ball, D. L., & Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407. <http://dx.doi.org/10.1177/0022487108324554>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, 5(1), 7-74. <http://dx.doi.org/10.1080/0969595980050102>
- Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York, NY: David McKay.
- Buchan, J. F., & Swann, M. (2007). A bridge too far or a bridge to the future? A case study in online assessment at Charles Sturt University. *Australasian Journal of Educational Technology*, 23(3), 408-434. <http://dx.doi.org/10.14742/ajet.1260>
- Burton, M. B. (1989). The effect of prior calculus experience on "introductory" college calculus. *The American Mathematical Monthly*, 96(4), 350-354. <http://dx.doi.org/10.2307/2324093>
- Cassady, J. C., Budenz-Anders, J., Pavlechko, G., & Mock, W. (2001, April). *The effects of internet-based formative and summative assessment on test anxiety, perceptions of threat, and achievement* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- Chou, C., Chan, P. S., & Wu, H. C. (2007). Using a two-tier test to assess students' understanding and alternative conceptions of cyber copyright laws. *British Journal of Educational Technology*, 38(6), 1072-1084. <http://dx.doi.org/10.1111/j.1467-8535.2006.00695.x>

- Chu, H. C. (2014). Potential negative effects of mobile learning on students' learning achievement and cognitive load - A format assessment perspective. *Educational Technology & Society*, 17(1), 332-344.
- Chu, H. C., Hwang, G. J., Tsai, C. C., & Tseng, J. C. R. (2010). A two-tier test approach to developing location-aware mobile learning systems for natural science courses. *Computers & Education*, 55(4), 1618-1627. <http://dx.doi.org/10.1016/j.compedu.2010.07.004>
- DeBoer, G. E., Quellmalz, E. S., Davenport, J. L., Timms, M. J., Herrmann-Abell, C. F., Buckley, B. C. ... & Flanagan, J. C. (2014). Comparing three online testing modalities: Using static, active, and interactive online testing modalities to assess middle school students' understanding of fundamental ideas and use of inquiry skills related to ecosystems. *Journal of Research in Science Teaching*, 51(4), 523-554. <http://dx.doi.org/10.1002/tea.21145>
- Denvir, B., & Brown, M. (1987). The feasibility of class administered diagnostic assessment in primary mathematics. *Educational Research*, 29(2), 95-107. <http://dx.doi.org/10.1080/0013188870290202>
- Ellis, R. A., Ginns, P., & Piggott, L. (2009). E-learning in higher education: Some key aspects and their relationship to approaches to study. *Higher Education Research and Development*, 28(3), 303-318. <http://dx.doi.org/10.1080/07294360902839909>
- Farrell, G., & Leung, Y. K. (2004). Innovative online assessment using confidence measurement. *Education and Information Technologies*, 9(1), 5-19. <http://dx.doi.org/10.1023/B:EAIT.0000024258.29560.3c>
- Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization, and affective factors. *American Educational Research Journal*, 14(1), 51-71. <http://dx.doi.org/10.3102/00028312014001051>
- Gálvez, J., Guzmán, E., & Conejo, R. (2009). A blended E-learning experience in a course of object oriented programming fundamentals. *Knowledge-Based Systems*, 22(4), 279-28. <http://dx.doi.org/10.1016/j.knosys.2009.01.004>
- Gaytan, J., & McEwen, B. C. (2007). Effective online instructional and assessment strategies. *The American Journal of Distance Education*, 21(3), 117-132. <http://dx.doi.org/10.1080/08923640701341653>
- Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4), 2333-2351. <http://dx.doi.org/10.1016/j.compedu.2011.06.004>
- Hwang, G. J., Wu, C. H., Tseng, J. C., & Huang, I. (2011). Development of a ubiquitous learning platform based on a real-time help-seeking mechanism. *British Journal of Educational Technology*, 42(6), 992-1002. <http://dx.doi.org/10.1111/j.1467-8535.2010.01123.x>
- Hwang, G. J., Yang, L. H., & Wang, S. Y. (2013). A concept map-embedded educational computer game for improving students' learning performance in natural science courses, *Computers & Education*, 69, 121-130. <http://dx.doi.org/10.1016/j.compedu.2013.07.008>
- Ibabe, I., & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, 59(2), 243-258. <http://dx.doi.org/10.1007/s10734-009-9245-6>
- Imrie, B. W., Cox, K., Imrie, B. W., Miller, A., & Miller, A. (2014). *Student assessment in higher education: A handbook for assessing performance*. London, UK: Routledge.
- Jourdan, N., Cretchley, P., & Passmore, T. (2007). Secondary-tertiary transition: What mathematics skills can and should we expect this decade? In J. Watson, & K. Beswick (Eds.), *Proceedings of 30th Annual Conference of Mathematics: Essential research, essential practice*. (Vol. 2, pp. 463-442). Hobart, Australia: MERGA.
- Karaman, S. (2011). Examining the effects of flexible online exams on students' engagement in e-learning. *Educational Research and Reviews*, 6(3), 259-264. <http://dx.doi.org/10.5897/ERR2014.1923>
- Kay, R., & Kletskin, I. (2012). Evaluating the use of problem-based video podcasts to teach mathematics in higher education. *Computers & Education*, 59(2), 619-627. [10.1016/j.compedu.2012.03.007](http://dx.doi.org/10.1016/j.compedu.2012.03.007) <http://dx.doi.org/10.1016/j.compedu.2012.03.007>
- Lawson, D. A. (2003). Changes in student entry competencies 1991-2001. *Teaching Mathematics and Its Applications*, 22(4), 171-175. <http://dx.doi.org/10.1093/teamat/22.4.171>
- Le, A., Joordens, S., Chrysostomou, S., & Grinnell, R. (2010). Online lecture accessibility and its influence on performance in skills-based courses. *Computers & Education*, 55(1), 313-319. <http://dx.doi.org/10.1016/j.compedu.2010.01.017>
- Lopez-Morteo, G., & López, G. (2007). Computer support for learning mathematics: A learning environment based on recreational learning objects. *Computers & Education*, 48(4), 618-641. <http://dx.doi.org/10.1016/j.compedu.2005.04.014>

- Manku, G. S., & Motwani, R. (2002). Approximate frequency counts over data streams. *Proceedings of the 28th international conference on Very Large Data Bases, Hong Kong, China*, 346-357. <http://dx.doi.org/10.1016/B978-155860869-6/50038-X>
- Nguyen, D. M., Hsieh, Y. C., & Allen, G. D. (2006). The impact of web-based assessment and practice on students' mathematics learning attitudes. *Journal of Computers in Mathematics and Science Teaching*, 25(3), 251-279. Chesapeake, VA: Association for the Advancement of Computing in Education.
- Nguyen, D. M., & Kulm, G. (2005). Using web-based practice to enhance mathematics learning and achievement. *Journal of Interactive Online Learning*, 3(3), 1-16.
- Nyachwaya, J. M., Mohamed, A. R., Roehrig, G. H., Wood, N. B., Kern, A. L., & Schneider, J. L. (2011). The development of an open-ended drawing tool: an alternative diagnostic tool for assessing students' understanding of the particulate nature of matter. *Chemistry Education Research and Practice*, 12(2), 121-132. <http://dx.doi.org/10.1039/C1RP90017J>
- Panjaburee, P., Triampo, W., Hwang, G. J., Chuedoung, M., & Triampo, D. (2013). Development of a diagnostic and remedial learning system based on an enhanced concept effect model. *Innovations in Education and Teaching International*, 50(1), 72-84. <http://dx.doi.org/10.1080/14703297.2012.746514>
- Pettersson, K., & Scheja, M. (2008). Algorithmic contexts and learning potentiality: A case study of students understanding of calculus. *International Journal of Mathematical Education in Science and Technology*, 39(6), 767-784. <http://dx.doi.org/10.1080/00207390801986908>
- Rod, B., & Allan, E. (2006). The prevalence and characteristics of online assessment in Australian universities. *Australasian Journal of Educational Technology*, 22(1), 104. <http://dx.doi.org/10.14742/ajet.1309>
- Roth, V., Ivanchenko, V., & Record, N. (2008). Evaluating student response to WeBWorK, a web-based homework delivery and grading system. *Computers & Education*, 50(4), 1462-1482. <http://dx.doi.org/10.1016/j.compedu.2007.01.005>
- Schoenfeld, A. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist*, 23(2), 145-166. http://dx.doi.org/10.1207/s15326985Sep2302_5
- Sealey, V. (2014). A framework for characterizing student understanding of Riemann sums and definite integrals. *The Journal of Mathematical Behavior*, 33(4), 230-245. <http://dx.doi.org/10.1016/j.jmathb.2013.12.002>
- Stratling, R. (2015). The complementary use of audience response systems and online tests to implement repeat testing: A case study. *British Journal of Educational Technology*. <http://dx.doi.org/10.1111/bjet.12362>
- Szendrei, J. (1996). Concrete materials in the classroom. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education*. Part 1 (pp. 411-434). Dordrecht, The Netherlands: Kluwer Academic Publishers. http://dx.doi.org/10.1007/978-94-009-1465-0_13
- Treagust, D. F. (1995) Diagnostic assessment of students' science concepts. In S. Glynn, & R. Duit (Eds.) *Learning science in the schools: Research reforming practice*. (pp. 327-346). New Jersey, NJ: Lawrence Erlbaum Associates.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159-169. <http://dx.doi.org/10.1080/0950069880100204>
- Treagust, D. F., Jacobowitz, R., Gallagher, J. J., & Parker, J. (2001). Using assessment as a guide in teaching for understanding: A study of middle school science class learning about sound. *Science Education*, 85(2), 137-157. [http://dx.doi.org/10.1002/1098-237X\(200103\)85:2<137::AID-SCE30>3.0.CO;2-B](http://dx.doi.org/10.1002/1098-237X(200103)85:2<137::AID-SCE30>3.0.CO;2-B)
- Tsai, C. C., & Chou, C. (2002). Diagnosing students' alternative conceptions in science. *Journal of Computer Assisted Learning*, 18(2), 157-165. <http://dx.doi.org/10.1046/j.0266-4909.2002.00223.x>
- Wang, T. H. (2014). Developing an assessment-centered e-Learning system for improving student learning effectiveness. *Computers & Education*, 73, 189-203. <http://dx.doi.org/10.1016/j.compedu.2013.12.002>
- Webb, N. L. (1997). Determining alignment of expectations and assessments in mathematics and science education. *NISE Brief*, 1(2), n2.
- Zandieh, M., Roh, K. H., & Knapp, J. (2014). Conceptual blending: Student reasoning when proving "conditional implies conditional" statements. *Journal of Mathematical Behavior*, 33(4), 209-229. <http://dx.doi.org/10.1016/j.jmathb.2013.11.007>

Zlatović, M., Balaban, I., & Kermek, D. (2015). Using online assessments to stimulate learning strategies and achievement of learning goals. *Computers & Education*, 91, 32-45.
<http://dx.doi.org/10.1016/j.compedu.2015.09.012>

Corresponding author: Gwo-Jen Hwang, gjhwang.academic@gmail.com

Australasian Journal of Educational Technology © 2017.

Please cite as: Yang, T. C., Fu, H. T., Hwang, G. J., & Yang, S. J. Y. (2017). Development of an interactive mathematics learning system based on a two-tier test diagnostic and guiding strategy. *Australasian Journal of Educational Technology*, 33(1), 62-80. <https://doi.org/10.14742/ajet.2154>