Computers & Education 73 (2014) 189–203

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Developing an assessment-centered e-Learning system for improving student learning effectiveness

Tzu-Hua Wang

Department of Education and Learning Technology, National Hsinchu University of Education, No. 521, Nanda Road, Hsinchu City 300, Taiwan, ROC

ARTICLE INFO

Article history:
Received 15 January 2013
Received in revised form
28 November 2013
Accepted 5 December 2013

Keywords:
Assessment-centered e-Learning
Web-based assessment
Dynamic assessment
Two-tier diagnostic assessment
GPAM-WATA e-Learning system

ABSTRACT

This research used Web-based two-tier diagnostic assessment and Web-based dynamic assessment to develop an assessment-centered e-Learning system, named the 'GPAM-WATA e-Learning system.' This system consists of two major designs: (1) personalized dynamic assessment, meaning that the system automatically generates dynamic assessment for each learner based on the results of the pre-test of the two-tier diagnostic assessment; (2) personalized e-Learning material adaptive annotation, meaning that the system annotates the e-Learning materials each learner needs to enhance learning based on the results of the pre-test of the two-tier diagnostic assessment and dynamic assessment. This research adopts a quasi-experimental design, applying GPAM-WATA e-Learning system to remedial Mathematics teaching of the 'Speed' unit in an elementary school Mathematics course. 107 sixth-graders from four classes in an elementary school participated in this research (55 male and 52 female). With each class as a unit, they were divided into four different e-Learning models: (1) the personalized dynamic assessment and personalized e-Learning material adaptive annotation group ($n = 26$); (2) the personalized dynamic assessment and non-personalized e-Learning material adaptive annotation group ($n = 28$); (3) the non-personalized dynamic assessment and personalized e-Learning material adaptive annotation group ($n = 26$); and (4) the non-personalized dynamic assessment and non-personalized e-Learning material adaptive annotation group ($n = 27$). Before remedial teaching, all students took the prior knowledge assessment and the pre-test of the summative assessment and two-tier diagnostic assessment. Students then received remedial teaching and completed all teaching activities. After remedial teaching, all students took the post-test of the summative assessment and two-tier diagnostic assessment. It is found that compared to the e-Learning models without personalized dynamic assessment, e-Learning models with personalized dynamic assessment are significantly more effective in facilitating student learning achievement and improvement of misconceptions, especially for students with low-level prior knowledge. This research also finds that personalized e-Learning material adaptive annotation significantly affects the percentage of reading time students spend on the e-Learning materials they need to enhance learning. However, it does not appear to predict student learning achievement and improvement of misconceptions.

1. Introduction

e-Learning has become an important trend in recent years. In addition to providing richer resources than the traditional classroom to facilitate learning, e-Learning also overcomes the limitations of time and space of traditional teaching. e-Learning allows learners to learn independently, meaning that it lacks the supervision and enforcement mechanisms of traditional teaching (Wang, 2011a). Given this, learners in an e-Learning environment must be highly self-regulated and independent, or their e-Learning effectiveness may be low (Kauffman, 2004; Wang, 2011a). Self-regulated learning plays an important role in both traditional and e-Learning environment. It is especially important in an e-Learning environment which lacks teacher's supervision and enforcement mechanisms (Jonassen, Davidson, Collins, Campbell, & Haag, 1995; King, Harner, & Brown, 2000; Puzziferro, 2008; Wang, 2011a).

E-mail address: tzuhuawang@gmail.com.

0360-1315/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compedu.2013.12.002

The major characteristic of self-regulated learning is that learners intentionally make an effort to manage and direct complicated learning activities (Gordon, Dembo, & Hocevar, 2007; Kauffman, 2004; Wang, 2011a). Puzziferro (2008) further pointed out that by providing proper learning strategies, an e-Learning environment could facilitate learners to perform self-regulated learning. Paris and Paris (2001) believed that self-assessment was an effective strategy to help learners perform self-regulated learning because learners were better able to evaluate their learning conditions if they could assess themselves during the learning process. They can also further monitor and correct their course of learning, and as a result, improve their learning effectiveness. Self-assessment is also the major feature of the assessment-centered learning environment (Bransford, Brown, & Cocking, 2000). Bradsford et al. contended that, in a successful assessment-centered learning environment, teachers incorporate assessment into teaching activities to facilitate students in performing self-assessment, and advance student learning effectiveness via interaction with the timely feedback from self-assessment. Since frontline teachers often teach many learners and are pressured into following a teaching schedule, it is difficult for learners to perform effective self-assessment and receive meaningful feedback. An assessment-centered learning environment is therefore difficult to be constructed in a traditional learning environment. However, with the help of information communication technology, when learners encounter difficulties during assessment, the system can help teachers provide learners with timely feedback. If the database of the Web-based assessment system is equipped with well-designed feedback data, it can provide learners with more effective feedback and facilitate e-Learning (Wang, 2007). In other words, learners can directly interact with a Web-based assessment system to perform effective self-assessment. Referring to the viewpoints of Bransford et al., Paris and Paris, Puzziferro and Wang, this research develops an assessment-centered e-Learning environment where learners can perform effective self-assessment by interacting with Web-based assessment system, expecting to improve learners' learning effectiveness.

The interactive model of learners and assessment can be constructed by adopting dynamic assessment. The theoretical basis of dynamic assessment is the 'Zone of Proximal Development (ZPD)' proposed by L. S. Vygotsky (Elliott, 2003; Haywood, Brown, & Wingenfeld, 1990). ZPD refers to the difference between the cognition level learners can achieve with and without the assistance of others such as teachers and outstanding peers (Elliott, 2003; Vygotsky, 1978; Wang, 2010, 2011b). According to Elliott, early dynamic assessment was mainly used to evaluate examinees' real ability and categorize them for specific training and education. In recent years, dynamic assessment has been more commonly applied to education. It is used to develop individualized educational interventions and in turn assist teaching (Elliott, 2003; Wang, 2011b). In most cases, dynamic assessment is performed in the form of test-teach-retest (Moore-Brown, Huerta, Uranga-Hernandez, & Peña, 2006; Wang, 2010). Campione and Brown (1985, 1987, pp. 92–95) proposed the 'graduated prompt approach' to develop and perform dynamic assessment. The key of 'graduated prompt approach' is to deliver instructional interventions with personalized prompts. When answering dynamic assessment items incorrectly, they can get personalized prompts to assist their learning. These personalized prompts are displayed in a pre-set sequence based on their levels of explicitness (Bransford, Delclos, Vye, Burns, & Hasselbring, 1987). They start with 'general hints' and gradually become 'specific hints'. General hints offer relatively little specific information about the solution, while a specific hint offers a detailed blueprint from which learners can generate the correct answer (Campione & Brown, 1985, 1987, pp. 92–95). Wang (2010, 2011b) followed the 'graduated prompt approach' in developing a Web-based dynamic assessment system, the 'Graduated Prompting Assessment Module of the WATA system (GPAM-WATA),' on the architecture of the 'Web-based Assessment and Test Analysis system (WATA system) (Wang, Wang, & Huang, 2008; Wang, Wang, Wang, Huang, & Chen, 2004).' GPAM-WATA can assist teachers in incorporating important learning concepts into the design of the instructional item and the instructional prompt, allowing learners to acquire substantial learning through online self-assessment by answering the item and receiving the prompt (for a detailed introduction of GPAM-WATA, see Section 2.1). Wang (2010) integrated GPAM-WATA into an e-Learning environment for a elementary school nature science course. Learners can log into the system for self-assessment after reading the e-Learning materials. Wang (2011b) also employed GPAM-WATA in remedial teaching for a junior high school Mathematics course. Learners log into the system for self-assessment after completion of traditional Mathematics instruction. Both studies showed that GPAM-WATA can help improve learning effectiveness.

GPAM-WATA in Wang (2010, 2011b) is a Web-based dynamic assessment system that allows teachers to create items and prompts, and administer online dynamic assessments, but does not allow teachers to compose e-Learning materials online, nor does it provide learners with the personalized assessment scenario. This means all learners logging in must answer the same set of items. This research further enhances the design of the teaching and learning strategies in GPAM-WATA to construct an assessment-centered e-Learning system, named the 'GPAM-WATA e-Learning system (GPAM-WATA_EL).' This e-Learning system allows teachers to construct items and prompts and compose e-Learning materials online and allows learners engaging in e-Learning to perform assessment-centered e-Learning. GPAM-WATA_EL is centered on assessment for all learning and teaching activities. For learning, the system provides personalized dynamic assessment, which relies on a two-tier diagnostic assessment. It automatically constructs a personalized dynamic assessment based on the results of the pre-test of the two-tier diagnostic assessment. In personalized dynamic assessment, examinees only need to answer the items related to the concepts they answered incorrectly in the pre-test of the two-tier diagnostic assessment. When they answer dynamic assessment items incorrectly, they can obtain progressive prompts in sequence and therefore achieve learning. For teaching, the system provides personalized e-Learning material adaptive annotation. Based on how each examinee answers items in the pre-test of the two-tier diagnostic assessment and dynamic assessment, personalized e-Learning material adaptive annotation reminds examinees by marking 'recommended reading' for the e-Learning materials requiring enhanced learning. These two designs are discussed in greater detail in Section 3.3.1.

In a traditional learning environment, as required by a teaching schedule, teachers often need to teach more than one student at a time. Therefore, it is not possible to provide effective teaching feedback based on students' personal needs. This research expects to apply GPAM-WATA_EL to assisting teachers' teaching, especially the remedial teaching. After conventional teaching process, teachers can leverage this e-Learning system to perform remedial teaching. Learners are allowed to compensate the deficiencies of their learning with personalized learning. This research applies GPAM-WATA_EL to the remedial teaching of the 'Speed' unit of an elementary school Mathematics course and investigates how the two designs of GPAM-WATA_EL, personalized dynamic assessment and personalized e-Learning material adaptive annotation, help improve student learning effectiveness. This research investigates student learning effectiveness from the perspectives of learning achievement and improvement of misconceptions. This research develops four e-Learning models (see Section 3.4) based on the two designs and investigates the effectiveness of the four models. This research also explores how the four e-Learning models assist learners with different levels of prior knowledge. This research answers three questions:

1. How effective are the four different e-Learning models developed based on personalized dynamic assessment and personalized e-Learning material adaptive annotation in the remedial teaching of elementary school Mathematics?
2. How do learners with different levels of prior knowledge differ in their learning effectiveness when learning in the four different e-Learning models developed based on personalized dynamic assessment and personalized e-Learning material adaptive annotation?
3. How does the personalized e-Learning material adaptive annotation provided by GPAM-WATA_EL affect student reading time of e-Learning materials?

2. Literature review

2.1. Web-based dynamic assessment: GPAM-WATA (Wang, 2010, 2011b)

The effectiveness of Web-based assessment in facilitating learning has been demonstrated in many studies (Buchanan, 2000; Deutsch, Herrmann, Frese, & Sandholzer, 2012; Hwang & Chang, 2011; Justham & Timmons, 2005). Based on the PsyCAL of Buchanan (2000), Hwang & Chang used the 'repeated answering', 'non-answer provision', and 'immediate feedback' strategies to develop FAML (Formative Assessment-based Mobile Learning). They found that FAML delivered good learning effectiveness when assisting students to learn local culture in a mobile learning environment. Deutsch et al. conducted Web-based mock examination, finding that examinees had positive attitudes toward Web-based assessment. Deutsch et al. argued that in higher education, Web-based assessment could be a strategy of computerized or Web-based learning which was attractive to learners. Justham and Timmons (2005) applied WebCT (http://www.Webct.com) to cultivating Web-based assessment environment. This environment was found to be quite effective in helping students improve their knowledge and understanding about statistics. Wang (2010, 2011b) improved the design of functions related to the interaction between Web-based assessment and learners and the feedback mechanisms. Wang's system offers examinees progressive feedback based on how they answer the items. Following the architecture of the 'Web-based Assessment and Test Analysis system (WATA system) (Wang et al., 2004, 2008)', Wang developed the 'Graduated Prompting Assessment Module of the WATA system (GPAM-WATA)'. The core architecture of the WATA system is the Triple-A model, which includes Assembling, Administering, and Appraising. Teachers are allowed to construct items online, manage the online item bank database, assemble examination papers online, and administer Web-based assessment. Functions such as item analysis and test analysis are also included for teachers to immediately perform analysis after the examination. The analysis results provide information to help teachers perform further appraisals (Wang et al., 2004, 2008).

GPAM-WATA is a Web-based dynamic assessment system. By performing dynamic assessment, a process of interactive assessment, teachers can provide learners with teaching assistance in the process of assessment, in turn improving their learning performance (Wang, 2010). GPAM-WATA is developed using the idea of cake format dynamic assessment proposed by Sternberg and Grigorenko (2001), and the graduated prompt approach of Campione and Brown (1985, 1987, pp. 92–95). The cake format dynamic assessment is an individualized approach to administering dynamic assessment. Examinees receive instruction by answering items one after another. If students answer correctly, they proceed to answer the next item. If they answer incorrectly, they are given a graded series of hints. These successive hints are designed to gradually reveal the correct answer, guiding examinees to find the correct answer step by step. The graduated prompt approach means that hints are delivered in a pre-arranged order. They start with general hints and gradually turn into specific hints. General hints are unspecific and less related to correct answers, while specific hints are more specific and can provide learners with a complete blueprint to solve questions (Campione & Brown, 1985, 1987, pp. 92–95). GPAM-WATA provides an instructional prompt (Wang, 2010, 2011b) when learners have difficulty answering an item. The instructional prompts are designed based on the graduated prompt approach and delivered in a pre-arranged order enabling learners to successively obtain instructional prompts when answering items incorrectly.

GPAM-WATA functions in the following way. First, the system presents items for learners to answer. If learners answer an item correctly, the system responds with the 'correct' message and lets them proceed to answer the next item. They do not need to answer the item repeatedly. If learners answer it incorrectly, the system provides an instructional prompt. After reading the instructional prompt, learners will first proceed to answer other items, and then randomly return to answer again the item they answered incorrectly. If they fail to answer it correctly after all three instructional prompts are delivered, the item does not appear again. This procedure will repeat until learners finish answering all items.

In Wang (2010, 2011b), GPAM-WATA is a general Web-based dynamic assessment system. Each and every learner must answer the same set of items upon logging into the system for self-assessment. The findings showed that GPAM-WATA was significantly effective in facilitating student learning, especially for students with low-level prior knowledge (Wang, 2010). In addition, the instructional prompts (IP) given to students when they have difficulty answering items are effective in remedying student weaknesses in learning Mathematics (Wang, 2011b). This research further adds the designs of personalized dynamic assessment and personalized e-Learning material adaptive annotation to enhance the teaching and learning strategies in GPAM-WATA. The newly developed system, the GPAM-WATA e-Learning system, is assessment-centered. When learners log into this e-Learning system, the items administered as their self-assessment, along with the e-Learning materials they read, will be personalized (see Section 3.3.1).

2.2. Two-tier diagnostic assessment

Two-tier diagnostic assessment has been widely applied as a research instrument to understand student misconceptions (Treagust, 1995). Applied to education, two-tier diagnostic assessment can be used to perform pre-test before instruction. Based on the test results, teachers can understand what the students know before instruction and design the course based on that understanding. In this way, two-tier diagnostic assessment not only helps diagnose student misconceptions but assists teaching (Treaguest, 1995). According to Treaguest, each item in the two-tier diagnostic assessment is answered in two tiers. In the first tier, examinees select the option they take as the correct answer. In the second tier, examinees choose a reason to explain why they selected the option as the correct answer in the first tier. Tsai and Chou (2002) further pointed out that two-tier diagnostic assessment could deliver even better efficiency if administered on computers.

Following the suggestions of Tsai and Chou (2002) and Treagust (1995), this research administers the two-tier diagnostic assessment online. This assessment is also used to explore student misconceptions before instruction to provide guidance for subsequent instruction. This research further integrates the two-tier diagnostic assessment into GPAM-WATA_EL, making it the basis for personalized dynamic assessment and personalized e-Learning material adaptive annotation. GPAM-WATA_EL automatically administers dynamic assessment based on student understanding before instruction, elicited by the two-tier diagnostic assessment. This e-Learning system also recommends learners the e-Learning materials they need to enhance learning based on how the learners perform in the two-tier diagnostic assessment and dynamic assessment.

2.3. Adaptive navigation support and prior knowledge

Though e-Learning environments offers rich resources, they often confuse learners. Eklund and Sinclair (2000) and Brusilovsky (2003) contended that the trouble was that learners became lost or disoriented during learning. They argued that an e-Learning environment was often learner-controlled. An e-Learning environment gives learners more opportunities to learn actively, but makes them more apt to become lost, skip important contents, choose not to answer questions, look for visually stimulating rather than informative material, and use the navigational features unwisely (Eklund & Sinclair, 2000). Based on a review of the literature, Eklund and Sinclair argued that an e-Learning environment should provide some kinds of expert assistance or guidance to make the information in an e-Learning environment more structured, provide learners with more learning possibilities and help learners solve disorientation problems.

Adaptive navigation support is one approach to solving the abovementioned problems (Eklund & Sinclair, 2000). Eklund and Sinclair believed that adaptive navigation support could be provided by a dynamic user model-driven annotation, and offered a number of techniques. Adaptive navigation support includes five technologies: direct guidance, adaptive ordering, hiding, adaptive (link) annotation, and adaptive (link) generation (Brusilovsky, 1996, 2001). Brusilovsky, Sosnovsky, and Yudelson (2009) further indicated that 'adaptive annotation' is the most effective. 'Adaptive annotation' means adding comments to links to e-Learning materials. Before clicking on the links, learners are provided with their own information related to the contents of the links. The comments are presented in different icons, colors, font sizes, or font types.

According to Hailikari, Nevgi, and Lindblom-Ylanne (2007) and Dochy, De Ridjt, and Dyck (2002), prior knowledge is the prerequisite knowledge influencing how one learns new information. Prior knowledge is also one of the key factors in learning effectiveness. In an e-Learning environment, learner's prior knowledge also plays an important role (Mitchell, Chen, & Macredie, 2005). Based on Brusilovsky (2003) and Mitchell et al., learners with different levels of prior knowledge have different needs and attitudes toward functional designs in an e-Learning environment. Mitchell et al. noted that learners with a lower level of prior knowledge needed more guidance and assistance in an e-Learning environment. Brusilovsky believed that for learners with a lower level of prior knowledge, the design of adaptive navigation support should be made more restrictive.

According to Brusilovsky (2003) and Mitchell et al. (2005), the design of an e-Learning environment needs to consider requirements of learners with different levels of prior knowledge. Learners with low-level prior knowledge need to be provided with relatively greater guidance and assistance. Moreover, based on the viewpoints of Eklund and Sinclair (2000) and Brusilovsky et al. (2009), adding adaptive navigation support to an e-Learning environment can effectively provide learners with guidance and assistance during learning. Among adaptive navigation support designs, adaptive annotation is the most effective approach (Brusilovsky et al., 2009). Brusilovsky et al. pointed out that adaptive annotation could help learners acquire knowledge faster (Brusilovsky & Pesin, 1998; Masthoff, 2002) and improve learning outcomes (Davidovic, Warren, & Trichina, 2003; Specht, 1998). Brusilovsky et al. also argued that adaptive annotation was a powerful personalization technology. Hence, this research uses adaptive annotation to design the adaptive navigation support, which is the personalized e-Learning material adaptive annotation, in the GPAM-WATA_EL (see Section 3.3.1).

3. Methodology

3.1. Participants

This research invited two elementary school Mathematics teachers with Web-based instruction experience and elementary school sixth-graders from four classes to participate. There were total of 107 valid samples, 55 male and 52 female. Students from the four classes were then randomly assigned to four groups with each class as a unit. The four groups are the PD_PE group ($n = 26$), the PD_nPE group ($n = 28$), the nPD_PE group ($n = 26$) and the nPD_nPE group ($n = 27$) (see Section 3.4). All students had participated in courses on computers and the Internet and were familiar with using computers and the Internet.

3.2. Learning contents

The learning contents of this research focus on the 'Speed' unit in an elementary school Mathematics course. The contents cover five major topics:

- Concept of time: It includes topics of sense of time and calculation of time.
- Concept of distance: It includes topics of calculation of distance and relative position of location.
- Concept of speed: It includes topics of meaning of speed, sense of speed and speed unit conversion.
- Concept of motion: It includes topics of simultaneous motion, partial simultaneous motion and continuous motion.
- Calculation and application of speed: It includes topics of questions of meet and chase, questions of average rate question, questions of echo, questions of downstream and counter current and questions of the direction of walk.

The learning contents are presented in Adobe Flash™ animations and pictures with as little text as possible. All learning contents are constructed in the e-Learning Material Database of the GPAM-WATA_EL.

3.3. Instruments

3.3.1. GPAM-WATA e-Learning system (GPAM-WATA_EL)

This research develops GPAM-WATA e-Learning system based on the GPAM-WATA (Wang, 2010, 2011b). The e-Learning system improves the design of teaching and learning strategies of GPAM-WATA, and provides an environment for assessment-centered e-Learning. GPAM-WATA_EL includes an Instructional Item and Prompt Database, a Diagnostic Item Bank and an e-Learning Material Database (Fig. 1). The Instructional Item and Prompt Database is the database for dynamic assessment. It includes the instructional items (II) and instructional prompts (IP), which together deliver the instructional assessment, administered in the form of cake format dynamic assessment (Sternberg & Grigorenko, 2001). The instructional assessment is also named dynamic assessment in this research. The purpose of instructional assessment is to help learners learn more by answering instructional items and progressively receiving instructional prompts. The Diagnostic Item Bank is the database for the two-tier diagnostic assessment. The items in Diagnostic Item Bank are developed by teachers based on their understanding of student misconceptions and are meant to diagnose student weakness in conceptual learning. The e-Learning Material Database allows teachers to construct multi-media e-Learning materials to enrich the online learning resources.

The items in Diagnostic Item Bank, the instructional items and prompts in Instructional Item and Prompt Database and the e-Learning materials in e-Learning Material Database are all constructed by teachers with rich teaching experience based on the learning concept coding system (Fig. 2), a professional concept architecture developed by teachers based on their teaching experience. GPAM-WATA_EL analyzes learner ability based on the learning concept coding system and provides personalized learning opportunities.

GPAM-WATA_EL provides two designs for assessment-centered e-Learning, personalized dynamic assessment and personalized e-Learning material adaptive annotation. In the personalized dynamic assessment, the system automatically generates dynamic assessment (instructional assessment) for each learner based on the results of the pre-test of the two-tier diagnostic assessment. All items in the Item and Prompt Database and Diagnostic Item Bank are mapped to the concepts in the learning concept coding system. Referring to the results of the pre-test of the two-tier diagnostic assessment, GPAM-WATA_EL defines concepts to which the items they answer incorrectly are mapped as the concepts they need to learn more thoroughly. Moreover, the instructional items and instructional prompts of the concepts will be selected from the Instructional Item and Prompt Database and used to automatically construct the personalized dynamic assessment. In other words, the items in the personalized dynamic assessment are related to the items students fail to answer correctly in the pre-test of the two-tier diagnostic assessment. Thus, students gain additional learning opportunities by answering instructional items one after another and receiving successively more specific instructional prompts.

In the personalized e-Learning material adaptive annotation, the system annotates the e-Learning materials each learner needs to enhance learning based on the results of the pre-test of the two-tier diagnostic assessment and dynamic assessment (instructional assessment). e-Learning materials are also mapped to concepts in the learning concept coding system. If learners incorrectly answer the items of the two-tier diagnostic assessment and the instructional items to which a particular concept is mapped, GPAM-WATA_EL will require the learner to read the relevant e-Learning materials. GPAM-WATA_EL annotates the links of the e-Learning materials related to the concept, highlighting and marking them as 'recommended reading,' and actively recommends them for learners to read. In other words, learners are thus prompted about topics that require further learning when reading e-Learning materials.

GPAM-WATA_EL allows teachers to set and arrange teaching activities online. The teaching activities can be dynamic assessment (instructional assessment), two-tier diagnostic assessment and reading e-Learning materials (Fig. 3). Below are descriptions of how students learn in GPAM-WATA_EL in this research (Fig. 1).

Firstly, students log into the GPAM-WATA_EL to take the pre-test of the two-tier diagnostic assessment (Fig. 4). Students then take the dynamic assessment (instructional assessment). When students cannot answer an instructional item (II) correctly, they successively receive instructional prompts (IP) of different levels (Fig. 5). Students can thus learn more and are guided to find the correct answer through the instructional prompts. After completing the dynamic assessment, students read the e-Learning materials (Fig. 6). After finishing reading the e-Learning materials, students take the post-test of the two-tier diagnostic assessment.

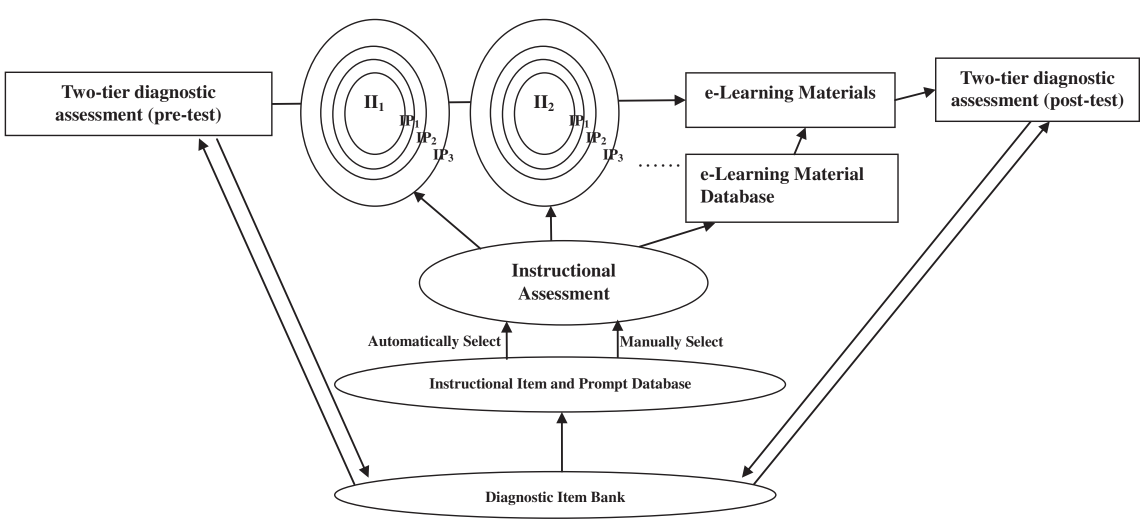

Fig. 1. GPAM-WATA_EL framework.

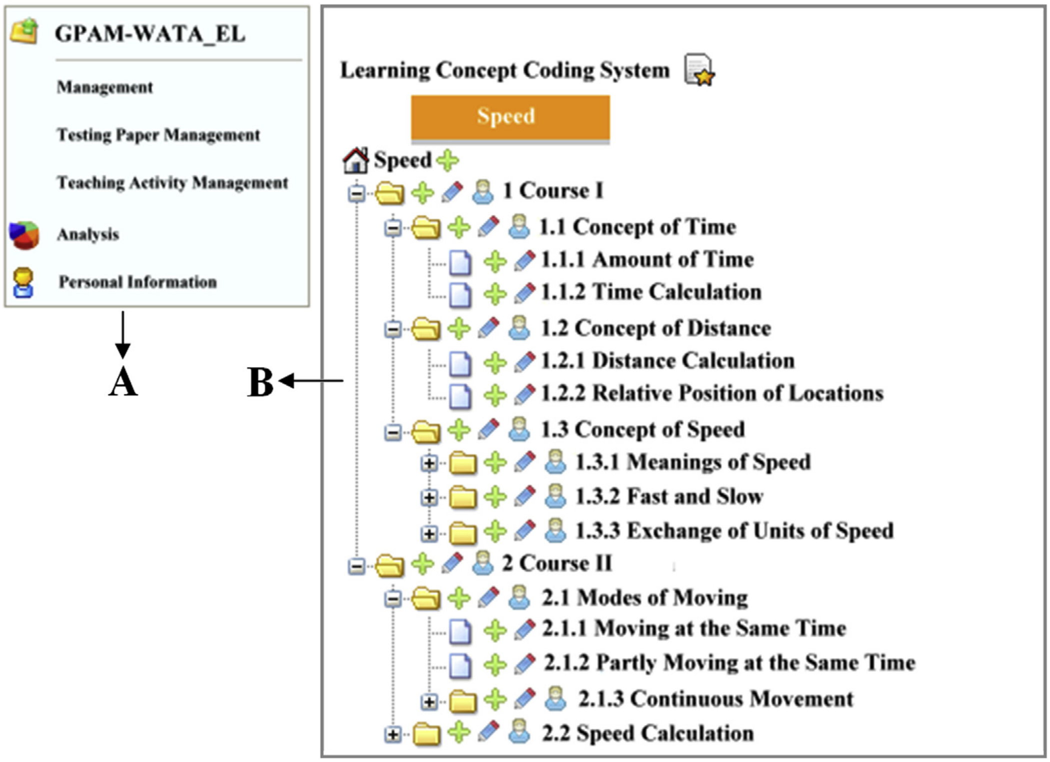

Fig. 2. (A): Function menu. (B): Learning concept coding system. Teacher can add and edit the properties of each concept code.

3.3.2. *Instructional item, instructional prompt and instructional assessment*

The instructional assessment is also named dynamic assessment. It is composed of instructional items and instructional prompts, as discussed in Sections 2.1 and 3.3.1. Currently, GPAM-WATA_EL consists of approximately 700 instructional items under different learning concepts; 46 items are relevant to the learning contents in this research, 'Speed', and will constitute the instructional assessment of this research. For learners, taking the instructional assessment in GPAM-WATA_EL means trying to answer instructional items and interactively obtaining instructional prompts. Through this process, learners will have more opportunities to learn. The instructional items and instructional prompts guide learners in learning and solving problems.

This research constructs instructional prompts based on the mathematical problem-solving theory of Mayer (1992, pp. 458–460). Mayer divides the mathematical problem-solving process into two steps, problem representation and problem solution, and four sub-processes, translation, integration, planning and monitoring, and execution. In each step and sub-process, learners need the relevant mathematical problem-solving knowledge to successfully solve problems. This research adopts this theory in developing the prompt content.

Based on the graduated prompt approach (Campione & Brown, 1985, 1987, pp. 92–95), specific hints provide a more specific problem-solving blueprint. Wang (2011b) contented that the specific hints were quite similar to the second step of the mathematical problem-solving process, the problem solution step, proposed by Mayer (1992, pp. 458–460). Based on Wang, this research reduces the four sub-processes of the two steps of mathematical problem-solving process described by Mayer to three phases by combining the last two sub-processes, planning and monitoring and execution (Table 1). The three instructional prompts for each instructional item are constructed based on the three phases, defined in Table 1.

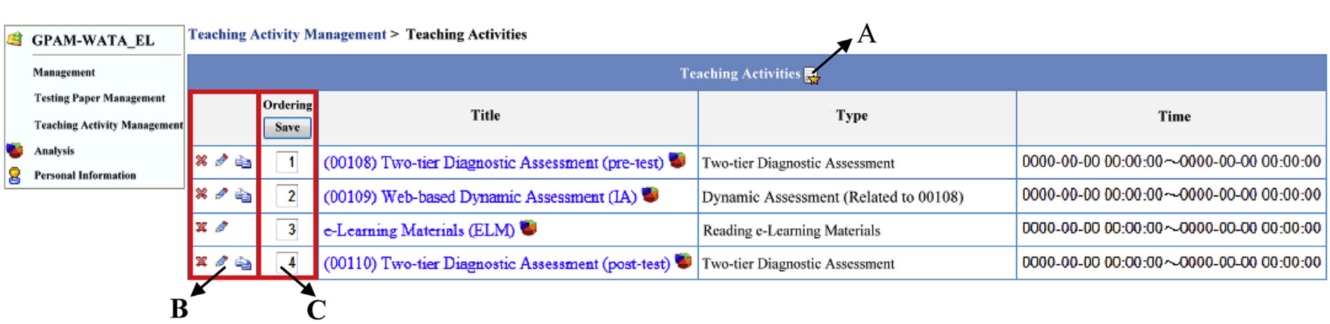

Fig. 3. Teaching activity management. (A): Teacher can click this icon to add a new teaching activity. (B): Teacher can click the icons here to manage the teaching activities, including deleting, editing and duplicating. (C): Teacher can arrange the order of teaching activities.

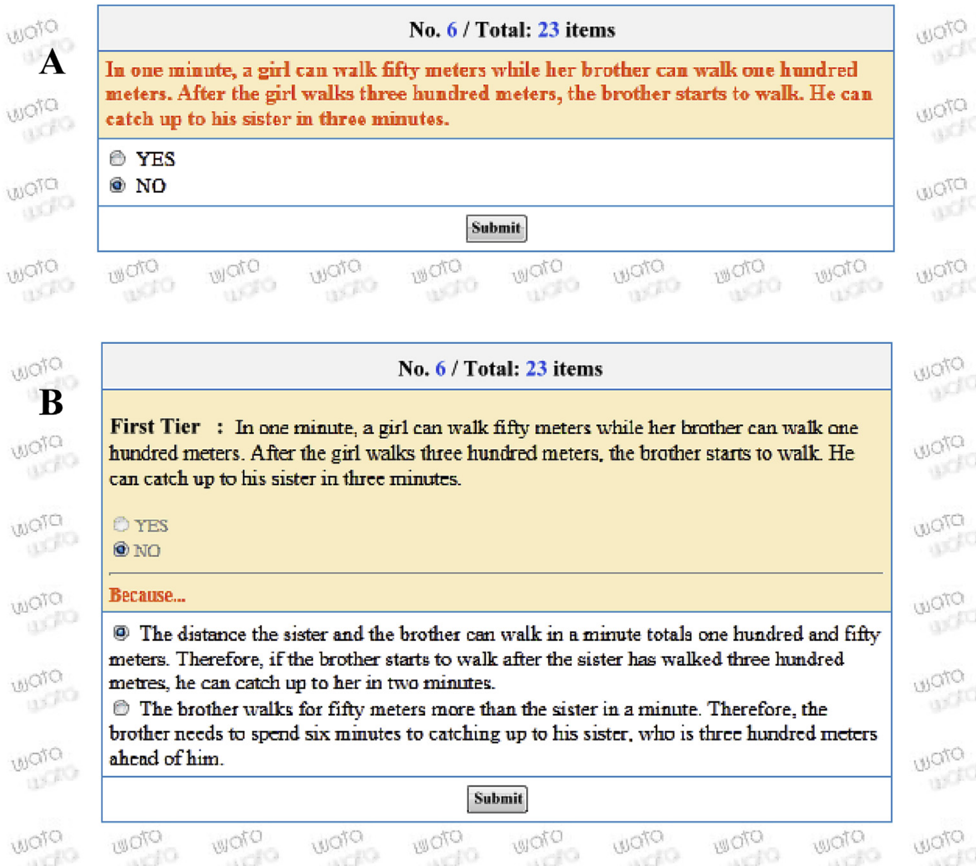

Fig. 4. Student is answering an item in the two-tier diagnostic assessment. (A) First tier. (B) Second tier.

3.3.3. Prior knowledge assessment

The prior knowledge assessment consists of 25 multiple choice items, primarily used to understand students' prior knowledge influencing how they learn the concepts of 'Speed.' It includes concepts about priority in time, time conversion, comparison and conversion of length, units of length, fractions, decimal fractions, ratios and ratio values, figure and chart comprehension. These concepts are the prior knowledge learners need before learning the contents of the topic of 'Speed' included in this research. These items are constructed by elementary school teachers based on the learning contents. All items are reviewed by experts in elementary school Mathematics teaching and assessment to ensure their validity. A two-way chart is also developed to ensure a complete and reasonable distribution of the items. The Cronbach's α of the prior knowledge assessment is 0.816 and its average difficulty is 0.796.

3.3.4. Two-tier diagnostic assessment

According to Treaguest (1995), the two-tier diagnostic assessment helps diagnose examinee misconceptions. This assessment can be used to explore student conceptions before instruction, with the results useful as a reference for subsequent instruction (Treaguest, 1995). In this research, the pre-test of the two-tier diagnostic assessment seeks primarily to understand the concepts learners should learn more thoroughly. In GPAM-WATA_EL, the results of the pre-test of the two-tier diagnostic assessment serve the basis for the personalized dynamic assessment and personalized e-Learning material adaptive annotation (see Section 3.3.1). It is also used as a post-test to understand how learner misconceptions are improved.

Currently, GPAM-WATA_EL contains nearly 300 two-tier diagnostic assessment items that belong to different learning concepts; 23 items are pertinent to the learning contents – 'Speed' embedded in this research, and will compose the two-tier diagnostic assessment of this research to understand students' learning condition of the 'Speed' concept. The results of the pre-test of the two-tier diagnostic assessment will be used as the foundation of the ensuing personalized dynamic assessment and personalized e-Learning material adaptive annotation, while the results of the post-test of the two-tier diagnostic assessment will be utilized, along with the pre-test results, to grasp the improvement of students' misconceptions. For validity, to ensure that the number and distribution of items are reasonable, a two-way chart is developed by elementary school Mathematics teachers based on the misconceptions learners often have when learning the 'Speed' concept. The 23 items are constructed using the two-way chart, and all items are reviewed by experts in elementary school Mathematics teaching and assessment. For reliability, the test-retest reliability of the two-tier diagnostic assessment is 0.856. All the items are collected in the Diagnostic Item Bank of the GPAM-WATA_EL.

3.3.5. Summative assessment

The items in the summative assessment are all multiple choice items (see Appendix for sample items). The summative assessment is used as the pre-test and post-test to evaluate student learning achievement. The pre-test shows students' understanding of the 'Speed'

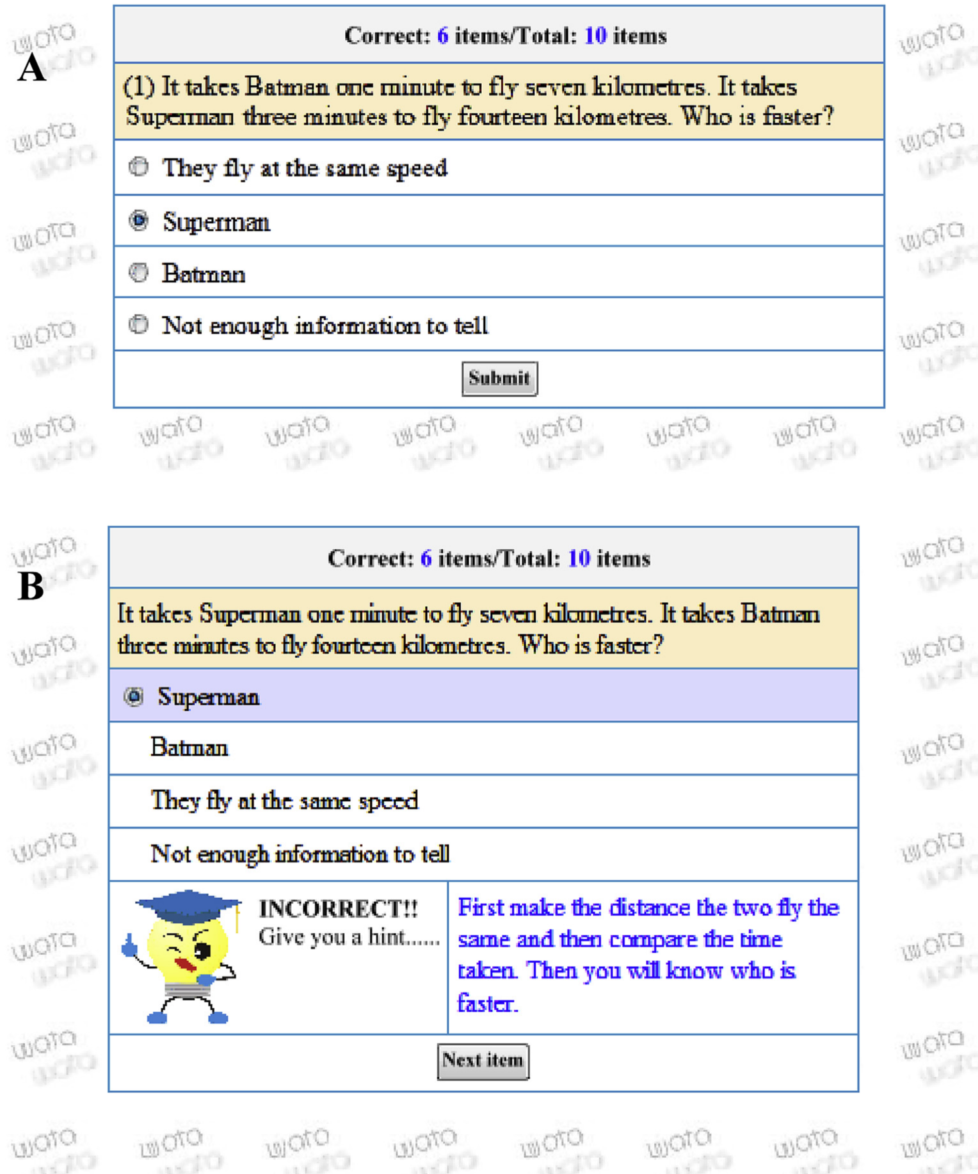

Fig. 5. (A) Student is answering an instructional item in the instructional assessment (dynamic assessment). (B) Student receives an instructional prompt.

concept before remedial teaching conducted in this research while the post-test represents their understanding after receiving remedial teaching. There are 25 items in the summative assessment. The items are designed by elementary school Mathematics teachers based on the learning contents. Items in the summative assessment do not appear in the two-tier diagnostic assessment and dynamic assessment (instructional assessment). For validity, all items are reviewed by experts of elementary school Mathematics teaching and assessment. A

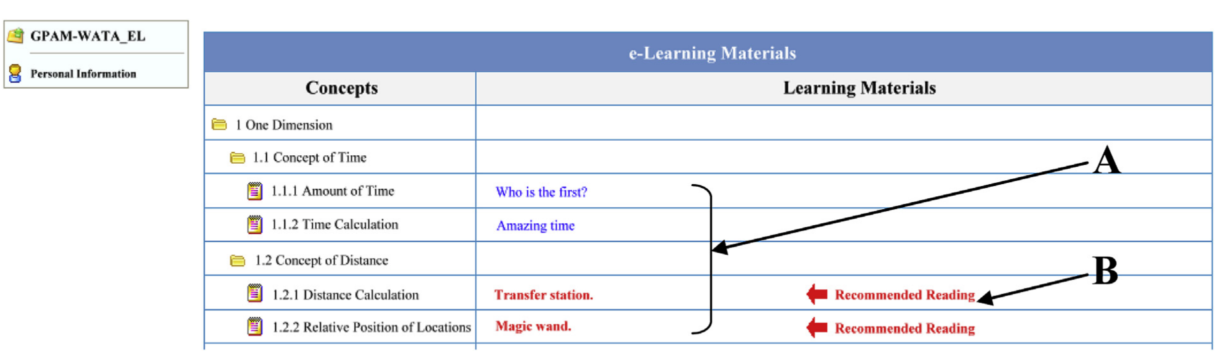

Fig. 6. Annotated e-Learning materials. (A) Student can click the titles to link to the learning materials. (B) Adaptive annotation.

Table 1
Correspondence between the design of IPs and Mayer (1992, pp. 458–460)'s mathematical problem-solving theory (Wang, 2011b).

Phases (IP order)	Design principle	Mayer's mathematical problem-solving steps	Mayer's mathematical problem-solving sub-processes	Mayer's required knowledge for each mathematical problem-solving sub-process
IP_1	Explanations of problems, helping learners to clarify conditions	Step 1. Problem representation	Sub-process 1. Translation	Linguistic knowledge, semantic knowledge
IP_2	Key concepts		Sub-process 2. Integration	Schematic knowledge
IP_3	Demonstrating how to solve a similar problem with simplified numbers or performing direct instruction	Step 2. Problem solution	Sub-process 3. Planning and monitoring	Strategic knowledge
			Sub-process 4. Execution	Procedural knowledge

two-way chart is also used to ensure that the number and distribution of items is complete and reasonable. For reliability, the Cronbach's α of the summative assessment is 0.883, while the average difficulty of the summative assessment is 0.782.

3.4. Research design and research procedure

There are two major designs in GPAM-WATA_EL, 'personalized dynamic assessment' and 'personalized e-Learning material adaptive annotation,' used to facilitate students in performing assessment-centered e-Learning. This research adopts a quasi-experimental design to investigate how the four e-Learning models constructed from the two designs facilitate student learning effectiveness, and further investigates the learning effectiveness of students with different levels of prior knowledge in the four e-Learning models. Four classes of students participating in this research are randomly divided into four groups with each class as a unit. The four groups learn in different e-Learning models. The four groups are: personalized dynamic assessment and personalized e-Learning material adaptive annotation group (PD_PE group), personalized dynamic assessment and non-personalized e-Learning material adaptive annotation group (PD_nPE group), non-personalized dynamic assessment and personalized e-Learning material adaptive annotation group (nPD_PE group) and non-personalized dynamic assessment and non-personalized e-Learning material adaptive annotation group (nPD_nPE group). Each group of students takes the same prior knowledge assessment, summative assessment, two-tier diagnostic assessment and studies the same e-Learning materials. The main difference between the four e-Learning models is whether the dynamic assessment (instructional assessment) is performed as a personalized dynamic assessment and whether personalized e-Learning material adaptive annotation is performed. For the non-personalized dynamic assessment, each student must answer all 46 instructional items in the dynamic assessment (instructional assessment). For the personalized dynamic assessment, the instructional items in dynamic assessment (instructional assessment) are selected on the basis of the results of the pre-test of the two-tier diagnostic assessment. Table 2 summarizes the research design of each e-Learning model.

The research procedure is as follows. First, all students complete learning of the 'Speed' unit in a conventional teaching environment using paper textbooks and take the prior knowledge assessment. Then all learners take the pre-test of the summative assessment to evaluate their entry ability, and log into GPAM-WATA_EL to familiarize themselves with the functions of the system. Each learner learns with a computer in the computer classroom. Students can also read the online user manual to know more about GPAM-WATA_EL. Students then receive three classes of remedial teaching in GPAM-WATA_EL and complete all teaching activities, including the pre-test of the two-tier diagnostic assessment, the dynamic assessment (instructional assessment), reading the e-Learning materials, and the post-test of the two-tier diagnostic assessment, in that order. The post-test of the two-tier diagnostic assessment is used to understand how student misconceptions are improved after the remedial teaching. In addition, students also take the post-test of the summative assessment to evaluate their learning achievement after finishing all teaching activities. Since this research is remedial Mathematics teaching, after finishing the post-test of the summative assessment, all students were allowed to log into the e-Learning models of other groups to learn. This approach can reduce the possible negative influences of the research design on students.

3.5. Data collection and analysis

The quantitative data collected in this research includes the scores of the prior knowledge assessment, scores of the pre-test and post-test of the summative assessment and scores of the pre-test and post-test of the two-tier diagnostic assessment. This research studies student learning effectiveness from two perspectives: learning achievement and improvement of misconceptions. The former is evaluated via the scores of the pre-test and post-test of the summative assessment, and the latter is evaluated via the scores of the pre-test and post-test of the two-tier diagnostic assessment. In addition, GPAM-WATA_EL also collects the reading time students spend on the e-Learning materials of each concept. To evaluate the effectiveness of personalized dynamic assessment and personalized e-Learning material adaptive annotation of GPAM-WATA_EL in facilitating student learning achievement, SPSS Version 18.0 is used to perform one-way ANCOVA. During one-way ANCOVA, to regulate the effects of pre-test and entry behavior on the post-test scores, the scores of the pre-test of the summative assessment are treated as the covariate, the scores of the post-test of the summative assessment are treated as the dependent variable, and

Table 2
Research design.

E-Learning models	Personalized dynamic assessment	Personalized e-Learning material adaptive annotation
PD_PE group	V	V
PD_nPE group	V	
nPD_PE group		V
nPD_nPE group		

the e-Learning models are treated as the fixed factor to test the effectiveness of the four different e-Learning models in performing remedial Mathematics teaching to improve students' learning effectiveness. Moreover, to understand how the personalized dynamic assessment and personalized e-Learning material adaptive annotation help students improve their misconceptions about 'Speed,' one-way ANCOVA is also used to perform an analysis. During one-way ANCOVA, to regulate the effects of pre-test and entry behavior on the post-test scores, the scores of the pre-test of the two-tier diagnostic assessment are treated as the covariate, the scores of the post-test of the two-tier diagnostic assessment are treated as the dependent variable, and the e-Learning models are treated as the fixed factor to test the effectiveness of the four different e-Learning models in performing remedial Mathematics teaching to improve students' misconceptions about 'Speed'. After one-way ANCOVA, the least significant difference (LSD) post hoc test is also adopted for pairwise comparison.

This research also investigates the performance of the students from the viewpoint of prior knowledge. It categorizes all students into two groups based on their scores on the prior knowledge assessment. The highest 50% of scores are placed in the high-level prior knowledge group, and the lowest 50% of scores are in the low-level prior knowledge group. The learning achievement and improvement of misconceptions of students with high-level and low-level prior knowledge receiving remedial Mathematics teaching in the four different e-Learning models are also analyzed using one-way ANCOVA as discussed above. In addition, to understand how the personalized e-Learning material adaptive annotation influences student reading time of e-Learning materials, this research also performs independent sample *t*-test on the percentage of reading time students spend on the e-Learning materials they need to enhance learning.

4. Results

4.1. Effectiveness of remedial Mathematics teaching using four different e-Learning models

To understand the effectiveness of the four different e-Learning models in performing remedial Mathematics teaching, this research investigates how learning achievement and misconceptions are improved across all students and students with different levels of prior knowledge.

To evaluate the learning achievement and improvement of misconceptions, one-way ANCOVA is used on the scores of the pre-test and post-test of the summative assessment and the scores of the pre-test and post-test of the two-tier diagnostic assessment of students in the four different e-Learning models. Before analysis, the assumption of homogeneity of regression coefficients was tested (summative assessment: $F_{3,99} = 2.611$, $p = 0.056$; two-tier diagnostic assessment: $F_{3,99} = 1.031$, $p = 0.382$). The results indicated that the homogeneity assumption was not violated. One-way ANCOVA results are shown in Tables 3 and 4:

Tables 3 and 4 show that the pre-test scores have a significant impact on the post-test scores (summative assessment: $F_{1,102} = 224.050$, $p < 0.01$; two-tier diagnostic assessment: $F_{1,102} = 280.275$, $p < 0.01$), as does the EL (e-Learning models) (summative assessment: $F_{3,102} = 6.720$, $p < 0.01$; two-tier diagnostic assessment: $F_{3,102} = 3.952$, $p < 0.05$). The results of the LSD post hoc test show that students in the PD_PE and PD_nPE groups are not significantly different in their learning achievement and improvement of misconceptions. Students in the PD_PE and PD_nPE groups both exhibit significantly better learning achievement and improvement of misconceptions than students in the nPD_PE and nPD_nPE groups. Tables 3 and 4 also show that students in the nPD_PE and nPD_nPE groups are not significantly different in their learning achievement and improvement of misconceptions.

Further investigation is performed to understand the learning achievement and improvement of misconceptions of students with high-level and low-level prior knowledge in the four different e-Learning models. One-way ANCOVA is then performed on the scores of the pre-test and post-test of the summative assessment and the scores of the pre-test and post-test of the two-tier diagnostic assessment of students with high-level and low-level prior knowledge in the four different e-Learning models. Before analysis, the assumption of homogeneity of regression coefficients was tested (summative assessment scores of low-level prior knowledge group: $F_{3,45} = 1.320$, $p = 0.280$, summative assessment scores of high-level prior knowledge group: $F_{3,46} = 0.556$, $p = 0.647$; two-tier diagnostic assessment scores of low-level prior knowledge group: $F_{3,45} = 2.611$, $p = 0.063$, two-tier diagnostic assessment scores of high-level prior knowledge group: $F_{3,46} = 2.691$, $p = 0.057$). The results indicate that the homogeneity assumption was not violated. One-way ANCOVA results are shown in Tables 5 and 6:

Tables 5 and 6 show that for students with low-level prior knowledge, the pre-test scores have a significant impact on the post-test scores (summative assessment: $F_{1,48} = 45.948$, $p < 0.01$; two-tier diagnostic assessment: $F_{1,48} = 40.390$, $p < 0.01$), as does the EL (e-Learning models) (summative assessment: $F_{3,48} = 4.880$, $p < 0.01$; two-tier diagnostic assessment: $F_{3,48} = 5.582$, $p < 0.01$). The results of LSD post hoc test show that students with low-level prior knowledge in the PD_PE and PD_nPE groups are not significantly different in their learning achievement and improvement of misconceptions. Students with low-level prior knowledge in the PD_PE and PD_nPE groups both have significantly better learning achievement and improvement of misconceptions than those in the nPD_PE and nPD_nPE groups. Tables 5 and 6 also show that students with low-level prior knowledge in the nPD_PE group are not significantly different from those in the nPD_nPE

Table 3
One-way ANCOVA analysis on learning achievement of students in four different e-Learning models ($n = 107$).

Source	SS	df	MS	*F* value	Post *hoc*[c]
Pre-Test[a]	20187.738	1	20187.738	224.050**	
EL[b]	1816.376	3	605.459	6.720**	PD_PE > nPD_PE PD_PE > nPD_nPE PD_nPE > nPD_PE PD_nPE > nPD_nPE
Error	9190.581	102	90.104		
Corrected total	34496.822	106			

**$p < 0.01$.
[a] Pre-test scores of the summative assessment.
[b] e-Learning models.
[c] Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table 4
One-way ANCOVA analysis on improvement of misconceptions of students in four different e-Learning models ($n = 107$).

Source	SS	df	MS	*F* value	Post hoc[c]
Pre-Test[a]	2400.401	1	2400.401	280.275**	
EL[b]	101.531	3	33.844	3.952*	PD_PE > nPD_PE PD_PE > nPD_nPE PD_nPE > nPD_PE PD_nPE > nPD_nPE
Error	873.573	102	8.564		
Corrected total	3594.43	106			

$^{**}p < 0.01$; $^{*}p < 0.05$.
[a] Pre-test scores of the two-tier diagnostic assessment.
[b] e-Learning models.
[c] Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table 5
One-way ANCOVA analysis on learning achievement of students with different levels of prior knowledge in four different e-Learning models.

Prior knowledge	Variable	Level	Mean[c,d] (std. error)	*F* value	Post hoc[e]
Low-level ($n = 53$)	Pre-Test[a]			45.948**	
	EL[b]	PD_PE	77.261 (3.423)	4.880**	PD_PE > nPD_PE
		PD_nPE	74.977 (2.973)		PD_PE > nPD_nPE
		nPD_PE	65.910 (3.118)		PD_nPE > nPD_PE
		nPD_nPE	62.311 (3.015)		PD_nPE > nPD_nPE
High-level ($n = 54$)	Pre-Test[a]			38.181**	
	EL[b]	PD_PE	94.559 (1.953)	2.274	
		PD_nPE	92.656 (1.956)		
		nPD_PE	89.577 (2.061)		
		nPD_nPE	87.884 (2.033)		

$^{**}p < 0.01$.
[a] Pre-test scores of the summative assessment.
[b] e-Learning models.
[c] Covariates appearing in the model are evaluated at the following values: The Pre-Test of students with low-level prior knowledge = 59.090. The Pre-Test of students with high-level prior knowledge = 84.670.
[d] Full score of the summative assessment is 100.
[e] Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

group in their learning achievement and improvement of misconceptions. For students with high-level prior knowledge, the pre-test scores have a significant impact on the post-test scores (summative assessment: $F_{1,49} = 38.181$, $p < 0.01$; two-tier diagnostic assessment: $F_{1,49} = 152.095$, $p < 0.01$). However, the EL (e-Learning models) has no significant impact on the post-test scores (summative assessment: $F_{3,49} = 2.274$, $p > 0.05$; two-tier diagnostic assessment: $F_{3,49} = 0.846$, $p > 0.05$). Thus, students with high-level prior knowledge display no significant difference in their learning achievement and improvement of misconceptions across the four different e-Learning models.

Tables 3 and 4 show that students exhibits significantly better learning achievement and improvement of misconceptions in the environment providing personalized dynamic assessment, while the design of personalized e-Learning material adaptive annotation does not significantly affect their learning achievement and improvement of misconceptions. Tables 5 and 6 show that students with low-level prior knowledge have significantly better learning achievement and improvement of misconceptions in the environment with personalized dynamic assessment, while personalized e-Learning material adaptive annotation does not significantly affect their learning achievement and improvement of misconceptions. However, for students with high-level prior knowledge, neither personalized dynamic assessment nor personalized e-Learning material adaptive annotation has a significant effect on their learning achievement and improvement of

Table 6
One-way ANCOVA analysis on improvement of misconceptions of students with different levels of prior knowledge in four different e-Learning models.

Prior knowledge	Variable	Level	Mean[c,d] (std. error)	*F* value	Post hoc[e]
Low-level ($n = 53$)	Pre-Test[a]			40.390**	
	EL[b]	PD_PE	12.871 (0.932)	5.582**	PD_PE > nPD_PE
		PD_nPE	12.052 (0.855)		PD_PE > nPD_nPE
		nPD_PE	8.704 (0.881)		PD_nPE > nPD_PE
		nPD_nPE	9.119 (0.851)		PD_nPE > nPD_nPE
High-level ($n = 54$)	Pre-Test[a]			152.095**	
	EL[b]	PD_PE	18.188 (0.577)	0.846	
		PD_nPE	18.098 (0.584)		
		nPD_PE	18.201 (0.599)		
		nPD_nPE	17.029 (0.618)		

$^{**}p < 0.01$.
[a] Pre-test scores of the two-tier diagnostic assessment.
[b] e-Learning models.
[c] Covariates appearing in the model are evaluated at the following values: the Pre-Test of students with low-level prior knowledge = 10.360. The Pre-Test of students with high-level prior knowledge = 16.300.
[d] Full score of the two-tier diagnostic assessment is 23.
[e] Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table 7
Analysis on students' reading time of e-Learning materials.

Group[a]	Average reading time students spend on the e-Learning materials they need to enhance learning (sec)	Average reading time students spend on all e-Learning materials (sec)	Average percentage of reading time students spend on the e-Learning materials they need to enhance learning	t value
Non-adaptive annotation ($n = 55$)	1020.782	2632.200	40.6	2.887**
Adaptive annotation ($n = 52$)	677.596	1272.404	58.2	

**$p < 0.01$.
[a] Non-adaptive annotation: Not having the design of 'personalized e-Learning material adaptive annotation' (PD_nPE group and nPD_nPE group); Adaptive annotation: Having the design of 'personalized e-Learning material adaptive annotation'(PD_PE group and nPD_PE group).

misconceptions. The results thus show that personalized e-Learning material adaptive annotation does not predict student learning achievement and improvement of misconceptions, while personalized dynamic assessment has a significant impact, especially for students with low-level prior knowledge.

4.2. Effectiveness of personalized e-Learning material adaptive annotation on student reading time of e-Learning materials

To understand how the design of personalized e-Learning material adaptive annotation in GPAM-WATA_EL affects student reading time of e-Learning materials, this research uses an independent sample t-test to analyze the percentage of reading time students spend on the e-Learning materials they need to enhance their learning. The results are shown in Table 7:

Table 7 shows that among the reading time students spend reading all e-Learning materials, students in the e-Learning models with personalized e-Learning material adaptive annotation (the PD_PE and nPD_PE groups) spend a significantly greater percentage of reading time on the e-Learning materials that they need to enhance their learning ($t = 2.887$, $p < 0.01$) than those in the e-Learning models without personalized e-Learning adaptive annotation (the PD_nPE and nPD_nPE groups). The results above show that the design of personalized e-Learning material adaptive annotation may affect student reading time of e-Learning materials, since it allows students to focus on learning about areas they need to understand better.

5. Concluding remarks

This research used Web-based two-tier diagnostic assessment and Web-based dynamic assessment to construct an assessment-centered e-Learning system, the GPAM-WATA e-Learning system (GPAM-WATA_EL). This e-Learning system allows teachers to manage two-tier diagnostic assessment, dynamic assessment (instructional assessment) and e-Learning materials online. Based on the results of the pre-test of the two-tier diagnostic assessment, GPAM-WATA_EL can construct the personalized dynamic assessment (instructional assessment) by automatically selecting the instructional items students need to answer and the instructional prompts they will receive. Thus, items in the personalized dynamic assessment are all related to the items students fail to answer correctly in the two-tier diagnostic assessment. Students can learn more by answering these selected instructional items and receiving prompts. Moreover, based on the results of the pre-test of the two-tier diagnostic assessment and dynamic assessment (instructional assessment), GPAM-WATA_EL automatically annotates and recommends e-Learning materials for students to read using the personalized e-Learning material adaptive annotation. This research finds that compared to non-personalized dynamic assessment in which all students answer the same instructional items, performing GPAM-WATA_EL in the personalized dynamic assessment model and applying it to remedial teaching of the 'Speed' unit in an elementary school Mathematics course is more effective in improving student learning achievement and misconceptions, especially for students with low-level prior knowledge. However, for students with high-level prior knowledge, no significant difference in learning effectiveness is found. These findings can be explained by Chen (2003), Cook, Beckman, Thomas, and Thompson (2008), Eklund and Sinclair (2000), Hammer (1996), Mitchell et al., (2005) and Treaguest (1995). The rich learning resources contained in an e-Learning environment tend to cause 'disorientation' (Eklund & Sinclair, 2000), making learners unable to focus on learning key points and therefore leading to low learning effectiveness. While learners with high-level prior knowledge can handle rich online learning resources using their rich prior knowledge, learners with low-level prior knowledge become more disoriented (Chen, 2003; Mitchell et al., 2005). Some researchers indicate that a personalized e-Learning environment customized to the level of prior knowledge would help prevent learners with low-level prior knowledge from coming across the disorientation problem in an e-Learning environment, and also advance learning effectiveness (Cook et al., 2008). Student weakness in conceptual learning and prior knowledge can be understood from the results of the two-tier diagnostic assessment (Hammer, 1996; Treaguest, 1995). In this research, the instructional items in the personalized dynamic assessment (instructional assessment) are automatically selected from the Instructional Item and Prompt Database based on the results of the pre-test of the two-tier diagnostic assessment. This personalized dynamic assessment helps learners focus on interacting with the instructional items and instructional prompts they need to enhance learning while reducing disorientation. In this way, learners can have better learning achievement and improvement of misconceptions. The effectiveness of this approach is especially evident on learners with low-level prior knowledge. As noted above, learners with high-level prior knowledge are less prone to disorientation. As a result, they display similar learning effectiveness statistically in both non-personalized dynamic assessment and personalized dynamic assessment.

The findings also show that the results of this research do not support the assumption that adaptive annotation leads to performance improvement, including student learning achievement and improvement of misconceptions. These findings differ from those of Brusilovsky and Pesin (1998), Davidovic et al. (2003), Masthoff (2002) and Specht (1998). This can be explained by Brusilovsky (2003). Brusilovsky contended that in some contexts, adaptive annotation did not necessarily deliver the expected effects because the effectiveness of adaptive annotation in facilitating learning effectiveness was also affected by the design of the overall e-Learning environment. The four e-Learning models in this research all include Web-based dynamic assessment. This is different from the designs used in Brusilovsky and Pesin (1998), Masthoff (2002), Specht (1998) and Davidovic et al. (2003). That may be one of the reasons why the research findings are different. However, further research into this phenomenon remains necessary.

Moreover, this research finds that the design of personalized e-Learning material adaptive annotation has a significant impact on student reading time of e-Learning materials. Students in the e-Learning environment providing personalized e-Learning material adaptive annotation spend a significantly greater percentage of reading time on the e-Learning materials they need to enhance their learning. This finding is consistent with the findings of Brusilovsky and Pesin (1998), Brusilovsky et al. (2009) and Masthoff (2002). They argued that adaptive annotation could enable learners to reduce navigational overhead. GPAM-WATA_EL uses the results of the pre-test of the two-tier diagnostic assessment and dynamic assessment (instructional assessment) to recommend and annotate e-Learning materials for students. The annotations remind students to read specific e-Learning materials they need to enhance learning. However, in an e-Learning environment without personalized e-Learning material adaptive annotation, students spend so much time on the navigation of e-Learning materials that they cannot effectively focus on the e-Learning materials they need to enhance their learning.

According to Bransford et al. (2000), the assessment-centered learning environment efficiently advances learning effectiveness; however, in the traditional learning environment, due to constraints of teaching schedule and the large number of simultaneous learners, it is difficult to create a successful assessment-centered learning environment by having learners carry out effective self-assessment. Web-based assessment, on the other hand, can assist in overcoming such constraints (Wang, 2007, 2011a). This research finds that instructional assessment, by providing opportunities of self-assessment and offering feedback (i.e. instructional prompt) when learners give incorrect answers, facilitates learners to understand their own learning conditions and promote self-regulated learning (Paris & Paris, 2001). The feedback learners receive when giving incorrect answers are instructional prompts developed by teachers to guide learners to learn, and is important in their interactive feature. The feedback, via the graduated prompt approach, provokes learners' potentials (i.e. ZPD) and leads learners to think and attain correct answers step by step, and to obtain learning during this process (Campione & Brown, 1985, 1987, pp. 92–95). In other words, contrast to other assessment strategies, learners are likely to experience better learning effectiveness when they perform self-assessment via Web-based dynamic assessment (Wang, 2010, 2011b). This research further finds that the personalized dynamic assessment of GPAM-WATA_EL, developed by applying two-tier diagnostic assessment and dynamic assessment, allows learners to perform effective self-assessment in the e-Learning environment for remedial teaching of the elementary school Mathematics course. This research also finds that, the effectiveness of personalized dynamic assessment in assisting learning is evidently better than non-personalized dynamic assessment. It helps implement successful assessment-centered e-Learning and delivers significant beneficial results on student learning achievement and improvement of misconceptions. Furthermore, the effectiveness is the most underlined with learners with low-level prior knowledge. Thus this research recommends adopting two-tier diagnostic assessment along with dynamic assessment to build personalized dynamic assessment that permit learners to execute effective online self-assessment. Based on this, an effective assessment-centered e-Learning environment can be constructed. Such an environment can be used for remedial teaching, and appears to offset shortcomings resulting from teachers' inability to interact with each individual learner and provide effective feedback to learners in traditional teaching. By means of this, learners may achieve better learning effectiveness. Teachers can also consider this model to apply to learners with low learning achievement or with low prior knowledge for improving their learning effectiveness. Guided by the assessment built into the system, students can understand key learning points and focus on learning more effectively, perform self-assessment to understand their own learning weaknesses, and make improvements via timely prompts.

Furthermore, this research faces several limitations. The research findings and implications are therefore constrained by those limitations, and not necessarily apt for other research context. First, constrained by research resources, this research consists of four treatment groups, with each group composed of 26–28 participants. The statistical power of the findings can be advanced if the number of participants in each group is increased. In addition, this research investigates the effectiveness of GPAM-WATA_EL on remedial teaching of the 'Speed' unit in elementary school Mathematics course. The participants have acquired a certain level of understanding for such concept and the concept's prerequisite knowledge; this may have contributed to higher scores on the pre-test of the summative assessment and prior knowledge assessment, particularly for learners in the high-level prior knowledge group. In addition, the items in the pre-test and post-test are identical. This may raise the issue of testing effect. As the phenomenon might bear on the research outcome, this research suggests future researches examine learners of various grades and their learning of different concepts, to further understand the effectiveness of GPAM-WATA_EL in learning. Further, students should be interviewed, and the learning paths taken by students in reading e-Learning materials should also be carefully recorded and analyzed. This will enable more detailed evaluation of GPAM-WATA_EL. According to Pintrich and Schrauben (1992), learners' engagement in learning tasks also plays a role in affecting learning effectiveness. With regards to learners' effort invested in learning tasks, this research advocates more analyses of learning behaviors in future researches. The learning behaviors of learners in this research include participation in dynamic assessment and reception of feedback and their involvement in two-tier diagnostic assessment and reading e-Learning materials. As a result of limitations in current system design, only the time learners spend reading e-Learning materials is recorded, but not the time they spend taking dynamic assessment and reading feedback, and answering the two-tier diagnostic assessment. Learning behaviors in those teaching activities may influence their learning effectiveness. This research recommends integrating these recording functions in GPAM-WATA_EL. Moreover, this research also recommends further researches on factors that may affect learners' efforts in learning tasks, such as the nature of learning tasks and students' motivational beliefs, in order to gain a deeper understanding of the interaction between those factors and each design of GPAM-WATA_EL. For the design of personalized e-Learning material adaptive annotation, this research only annotates and highlights the e-Learning materials students should read to enhance their learning as 'recommended reading.' Based on the findings of this research, this design allows students to spend a significantly greater percentage of reading time on the e-Learning materials they need to enhance their learning. However, this design does not appear to predict student learning achievement and improvement of misconceptions. Therefore, this research suggests that in addition to annotating the e-Learning materials for enhanced reading as 'recommended reading' based on the results of the pre-test of the two-tier diagnostic assessment and dynamic assessment, key concepts in the learning contents should also be highlighted and the reasons for the reading recommendation should be stated in annotations. For example, if students need to read e-Learning materials about a certain concept, before they click on the links to the e-Learning materials, they should have an overview of the materials and be given an explanation of why the materials are needed to enhance learning in advance. In this way, when reading the recommend e-Learning materials, students can have more information to help them focus on the concepts they need to enhance learning. The design of personalized e-Learning material adaptive annotation may thus become more effective.

Acknowledgment

The author deeply appreciates the National Science Council in Taiwan for the financial support and encouragement under Grant No. 98-2511-S-134-003-MY2. The author is grateful for the insightful comments from the editor and referees.

Appendix

Samples items of the summative assessment

1 A racing horse runs 50 miles a second, and a leopard runs 80 miles a second. If the two animals run in the same direction at the same time from the same place, what is the distance between them after 10 seconds?
 A 300 miles
 B 13 miles
 C 130 miles
 D 1300 miles
2 The rates of speed for skateboarding and in-line skating are 100 miles and 200 miles per minute, respectively. If a skateboarder departs first and opens a lead of 600 miles before an in-line skater pursues him, how many minutes will the in-line skater take to catch up with the skateboarder?
 A 2 minutes
 B 3 minutes
 C 6 minutes
 D cannot catch up

References

Bransford, J. D., Brown, A., & Cocking, R. (2000). *How people learn: Mind, brain, experience and school* (Expanded ed.). Washington, DC: National Academy Press.
Bransford, J. C., Delclos, J. R., Vye, N. J., Burns, M., & Hasselbring, T. S. (1987). State of the art and future directions. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 479–496). New York: Guilford Press.
Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-adapted Interaction, 6*, 87–129.
Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User Adapted Interaction, 11*, 87–110.
Brusilovsky, P. (2003). Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation. *British Journal of Educational Technology, 34*, 487–497.
Brusilovsky, P., & Pesin, L. (1998). Adaptive navigation support in educational hypermedia: An evaluation of the ISIS-Tutor. *Journal of Computing and Information Technology, 6*, 27–38.
Brusilovsky, P., Sosnovsky, S., & Yudelson, M. (2009). Addictive links: The motivational value of adaptive link annotation. *New Review of Hypermedia and Multimedia, 15*, 97–118.
Buchanan, T. (2000). The efficacy of a world-wideweb mediated formative assessment. *Journal of Computer Assisted Learning, 16*(3), 193–200.
Campione, J. C., & Brown, A. L. (1985). *Dynamic assessment: One approach and some initial data*. (Technical report No. 361). Bethesda, MD: National Institute of Child Health and Human Development; National Institute of Education. (ERIC ED269735).
Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An international approach to evaluating learning potential* (pp. 82–115). New York: The Guilford Press.
Chen, S. Y. (2003). Editorial: Individual differences in web-based instruction—an overview. *British Journal of Educational Technology, 34*, 385–392.
Cook, D. A., Beckman, T. J., Thomas, K. G., & Thompson, W. G. (2008). Adapting web-based instruction to residents' knowledge improves learning efficiency. *Journal of General Internal Medicine, 23*, 985–990.
Davidovic, A., Warren, J., & Trichina, E. (2003). Learning benefits of structural example-based adaptive tutoring systems. *IEEE Transactions on Education, 46*, 241–251.
Deutsch, T., Herrmann, K., Frese, T., & Sandholzer, H. (2012). Implementing computer-based assessment – a web-based mock examination changes attitudes. *Computers & Education, 58*(4), 1068–1075.
Dochy, F. J. R. C., De Ridjt, C., & Dyck, W. (2002). Cognitive prerequisites and learning. How far have we progressed since bloom? Implications for educational practice and teaching. *Active Learning in Higher Education, 3*, 265–284.
Eklund, J., & Sinclair, K. (2000). An empirical appraisal of the effectiveness of adaptive interfaces for instructional systems. *Educational Technology & Society, 3*, 165–177.
Elliott, J. G. (2003). Dynamic assessment in educational settings: Realizing potential. *Educational Review, 55*(1), 15–32.
Gordon, S. C., Dembo, M. H., & Hocevar, D. (2007). Do teachers' own learning behaviors influence their classroom goal orientation and control ideology? *Teaching and Teacher Education, 23*(1), 36–46.
Hailikari, T., Nevgi, A., & Lindblom-Ylänne, S. (2007). Exploring alternative ways of assessing prior knowledge, its components and their relation to student achievement: A mathematics based case study. *Studies in Educational Evaluation, 33*, 320–337.
Hammer, D. (1996). Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions? *Journal of the Learning Sciences, 5*(2), 97–127.
Haywood, H. C., Brown, A. L., & Wingenfeld, S. (1990). Dynamic approaches to psycho-educational assessment. *School Psychology Review, 19*(4), 411–422.
Hwang, G.-J., & Chang, H.-F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*(4), 1023–1031.
Jonassen, D., Davidson, M., Collins, M., Campbell, J., & Haag, B. B. (1995). Constructivism and computer-mediated communication in distance education. *The American Journal of Distance Education, 9*(2), 7–26.
Justham, D., & Timmons, S. (2005). An evaluation of using a web-based statistics test to teach statistics to postregistration nursing students. *Nurse Education Today, 25*(2), 156–163.
Kauffman, D. F. (2004). Self-regulated learning in web-based environments: Instructional tools designed to facilitate cognitive strategy use, metacognitive processing, and motivational beliefs. *Journal of Educational Computing Research, 30*(1&2), 139–161.
King, F. B., Harner, M., & Brown, S. W. (2000). Self-regulatory behavior influences in distance learning. *International Journal of Instructional Media, 27*(2), 147–155.
Masthoff, J. (2002). *Design and evaluation of a navigation agent with a mixed locus of control*. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *Intelligent Tutoring systems. Lecture Notes in Computer Science* (Vol. 2363); (pp. 982–991). Berlin: Springer-Verlag.
Mayer, R. E. (1992). *Thinking, problem solving, cognition*. New York, NY: W.H. Freeman and Company.
Mitchell, T. J. F., Chen, S. Y., & Macredie, R. D. (2005). Hypermedia learning and prior knowledge: Domain expertise vs. system expertise. *Journal of Computer Assisted Learning, 21*, 53–64.
Moore-Brown, B., Huerta, M., Uranga-Hernandez, Y., & Peña, E. D. (2006). Using dynamic assessment to evaluate children with suspected learning disabilities. *Intervention in School and Clinic, 41*, 209–217.
Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89–101.
Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In D. Schunk, & J. Meece (Eds.), *Students perceptions in the classroom: Causes and consequences* (pp. 149–183). Hillsdale, NJ: Erlbaum.
Puzziferro, M. (2008). Online technologies self-efficacy and self-regulated learning as predictors of final grade and satisfaction in college-level online courses. *The American Journal of Distance Education, 22*(2), 72–89.
Specht, M. (1998). Empirical evaluation of adaptive annotation in hypermedia. In T. Ottman, & I. Tomek (Eds.), *Proceedings of ED-MEDIA & ED-TELECOM98* (pp. 1327–1332). Charlottesville, VA: AACE.
Sternberg, R. J., & Grigorenko, E. L. (2001). All testing is dynamic testing. *Issues in Education, 7*, 137–170.

Treagust, D. F. (1995). Diagnostic assessment of student's science concepts. In S. Glynn, & R. Duit (Eds.), *Learning science in the schools: Research reforming practice* (pp. 327–346). Hillsdale, NJ: Erlbaum.
Tsai, C.-C., & Chou, C. (2002). Diagnosing students' alternative conceptions in science. *Journal of Computer Assisted Learning, 18*, 157–165.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
Wang, T. H. (2007). What strategies are effective for formative assessment in an e-Learning environment? *Journal of Computer Assisted Learning, 23*(3), 171–186.
Wang, T. H. (2010). Web-based dynamic assessment: Taking assessment as teaching and learning strategy for improving students' e-Learning effectiveness. *Computers & Education, 54*(4), 1157–1166.
Wang, T. H. (2011a). Developing web-based assessment strategies for facilitating junior high school students to perform self-regulated learning in an e-Learning environment. *Computers & Education, 57*(2), 1801–1812.
Wang, T. H. (2011b). Implementation of web-based dynamic assessment in facilitating junior high school students to learn mathematics. *Computers & Education, 56*(4), 1062–1071.
Wang, T. H., Wang, K. H., & Huang, S. C. (2008). Designing a web-based assessment environment for improving pre-service teacher assessment literacy. *Computers & Education, 51*(1), 448–462.
Wang, T. H., Wang, K. H., Wang, W. L., Huang, S. C., & Chen, S. Y. (2004). Web-based assessment and test analyses (WATA) system: Development and evaluation. *Journal of Computer Assisted Learning, 20*, 59–71.