Computers & Education 57 (2011) 1727–1740

Contents lists available at ScienceDirect

Computers & Education

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

A Web-based model for developing assessment literacy of secondary in-service teachers

Ya-Ching Fan [a], Tzu-Hua Wang [b], Kuo-Hua Wang [a,*]

[a] Graduate Institute of Science Education, National Changhua University of Education, No. 1, Jinde Rd., Changhua City 500, Taiwan, ROC
[b] Department of Education, National Hsinchu University of Education, No. 521, Nanda Rd., Hsinchu City 300, Taiwan, ROC

ARTICLE INFO

Article history:
Received 4 October 2010
Received in revised form
4 March 2011
Accepted 8 March 2011

Keywords:
Improving classroom teaching
Teaching/learning strategies
Teacher assessment literacy
Teacher education
WATA system
P2R-WATA Assessment Literacy Development Model

ABSTRACT

This research investigates the effect of a web-based model, named 'Practicing, Reflecting, and Revising with Web-based Assessment and Test Analysis system (P2R-WATA) Assessment Literacy Development Model,' on enhancing assessment knowledge and perspectives of secondary in-service teachers, and adopts a single group experimental research design. The WATA system provides teachers with personalized learning resources and situated environment to practice assembling, administering tests on-line, and appraising test-related statistical information to reflect and revise test items. The sample consisted of forty-seven secondary in-service mathematics and science teachers in a summer program for 36 h within six weeks. This research collects and analyzes quantitative data, including the pre-test and post-test of teachers' knowledge and perspectives on assessment. The major results generally confirm the effectiveness of P2R-WATA model. Firstly, the assessment knowledge of the participants has improved after training, especially for teachers with low-level prior knowledge. Secondly, the findings also reveal that there is a significant improvement on teachers' assessment perspectives.

1. Introduction

Assessing student performance in classroom instruction is one of the most pivotal tasks for today's teachers. Assessment literacy empowers teachers to make data-driven decisions aimed at teaching and learning. Teachers' inadequate knowledge and capacity can cripple the quality of education (Popham, 2009). The *National Science Education Standards* (National Research Council [NRC], 1996) of the United States emphasizes the importance of assessment as a communicating tool in the instructional process. Gronlund and Linn (1990) pointed out that teachers who assess instruction could initially realize their students' needs, monitor learning and instructional processes, diagnose student learning difficulty, and confirm learning achievement. Recent researchers have focused on proper assessment to improve student learning (Campbell & Collins, 2007; McMunn, McColskey, & Butler, 2004; Popham, 2009; Wang, 2011; Wang, 2011; Wang, Wang, Wang, & Huang, 2006). Teachers with assessment capacity can realize students' thoughts, beliefs, and reasoning to improve student learning. Therefore, assessment literacy is seen as an important professional knowledge and capacity a teacher should possess.

Regrettably, researchers have consistently shown that teachers lack assessment literacy to administer proper classroom assessment (Arter, 2001; Mertler, 2004; Mertler & Cambell, 2005; Popham, 2006; Wang, Wang, & Huang, 2008). The major reasons caused teachers lack assessment literacy including 'faults in teacher education program (Lukin, Bandalos, Eckhout, & Mickelson, 2004; Mertler, 2004; Stiggins, 2002; Wang et al., 2008)' and 'faults in teacher certification policy (Stiggins, 2002; Wang et al., 2008)' both in the United States and Taiwan.' Teachers are not required to learn about educational assessment when they complete their teacher education programs. These teachers might learn the concepts and practices of educational assessment only in a few sessions within educational psychology courses or a unit in a methods class (La Marca, 2006; Stiggins, 2006; Popham, 2009). In Taiwan, assessment-related courses are optional and sometimes lacking in teacher education programs. The P2R-WATA development model is designed to solve the problem of 'faults in teacher education programs' for pre-service teachers to enhance their assessment literacy in Taiwan (Wang et al., 2008). The model based on the key elements

* Corresponding author.
E-mail addresses: elva05@gmail.com (Y.C. Fan), tzuhuawang@gmail.com (T.H. Wang), sukhua@gmail.com (K.H. Wang).

0360-1315/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2011.03.006

of four successful training models including the Plake and Impara model (Plake & Impara, 1993), the Nebraska Assessment Cohort (NAC) model (Lukin et al., 2004), the Assessment Literacy Learning Team (ALLT) model (Arter, 2001; Stiggins, 2001), the Pre-service Assessment Learning Study group (PALS) and In-service and Pre-service Assessment Learning Study group (IPALS) model (Lukin et al., 2004) to constructs a Web-based environment training model. To investigate how to design a Web-based assessment system to help develop teacher assessment literacy, scholars have offered a wealth of advice on the design of computer-assisted and web-based assessment systems. Two essential components of P2R-WATA were the personalized and situated design. The personalized design was based on the suggestions of Egbert and Thomas (2001) and Passig (2001). The situated design was based on the most important characteristic in the four assessment literacy development models, offering opportunities for teachers to apply what they learned to realistic classroom situations. Participants 'practice' assembling, administering tests on-line, and appraising test-related statistical information generated by the WATA system. Using statistical information, teachers 'reflect' on problematic test items, student learning and instructions, 'revise' certain test items, and clarify instructions. Many scholars have emphasized the importance of teacher assessment literacy and believe that assessment-related courses to development of teacher assessment literacy should be integrated into pre-service and in-service teacher education (Lukin et al., 2004; Mertler, 2004; Stiggins, 1995, 1999; Stiggins, Arter, Chappuis, & Chappuis, 2004).

In this research, to rectify this omission on teacher education program, training in-service teachers with high-quality assessment literacy and building professional capacities for improving student learning are essential.

Additionally, the recent development of Information and communications technology has made ICT a component in many education reforms worldwide. Bodzin and Cates (2003) suggested that e-Learning environments provide resources that are more abundant for learners in educational fields than traditional classroom environments. Teachers and students can immediately interact, share knowledge simultaneously, and teachers can encourage students to learn independently. Moreover, the errorless high-speed data processing feature has made computers an accepted popular assessment tool in education. In other words, Web-based Testing (WBT) saves more time than paper-and-pencil tests, teachers can administer test items, and the immediate results enhance learning effectiveness (e.g., correct tests, analysis, and record test scores) (Mills, 2002; Sheader, Gouldsborough, & Grady, 2006; Wang, 2010). Using information technology has become a popular form of professional development in teacher education. However, to date, very few literatures have reported the development and evaluation of web-based assessment tools for teacher education. Therefore, this research attempts to investigate the effectiveness of an assessment literacy development model, named Practicing, Reflecting, and Revising with Web-based Assessment and Test Analysis (P2R-WATA) model, developed by Wang et al., (2008) for assessment literacy of secondary in-service mathematics and science teachers. This research is a follow-up research and describes the model in detail in a later section. The current investigation addresses the following research questions:

1.1 What is the effectiveness of the P2R-WATA Assessment Literacy Development model for improving secondary in-service teachers' assessment knowledge?
1.2 What is the effectiveness of the P2R-WATA Assessment Literacy Development model for improving secondary in-service teachers' assessment perspective?

2. Literature review

2.1. Assessment literacy for teachers

Assessment literacy is an important professional knowledge a teacher should possess. As an element of the teaching knowledge base, pedagogical content knowledge (PCK) contains assessment literacy as a component (Shulman, 1986). The Code of Professional Responsibilities in Educational Measurement (National Council on Measurement in Education [NCME], 1995) mentioned that teachers should take professional responsibility for educational assessment. Teachers are obligated to administer professional assessment and strictly obey the ethical principles of assessment to fulfill their professional responsibilities. Stiggins (1991, p. 7; 1999, p. 23) expressed that 'teachers spend a third to a half of their professional time involved in assessment-related activities.' The present research concurs that assessment-related activities consume considerable time from classroom instruction. However, many studies have consistently shown that teachers' assessment literacy is inadequate (Plake & Impara, 1993; Schafer, 1993; Mertler, 2004). Schafer (1993) investigated teachers' education programs in the Association of the American College of Teacher Education (AACTE) member institutions and indicated that half of the total number of participating schools did not require pre-service teachers to take assessment-related courses before they graduated. Furthermore, Stiggins (1999) pointed out that only half of the states required an examination of teachers' assessment abilities in the United States.

The majority of programs have failed for decades to provide the assessment literacy required to enable teachers to engage in assessment for learning (Stiggins, 2002). For example, Stiggins observes that only approximately a dozen states in the United States explicitly requiring assessment competence as a prerequisite for teaching licensure. This problem has existed in many countries other than the United States. DeLuca and McEwen (2007) revealed that only three of ten bachelors of education programs had a mandatory assessment course or module in Ontario, Canada. Out of ten Canadian provinces and 50 U.S. states, only Hawaii and Nebraska invest a significant amount of money to improve assessment and evaluation practices within schools (Volante & Fazio, 2007). Apparently, teachers are not required to learn about educational assessment when they complete their teacher education programs. As a consequence, these teachers might learn the concepts and practices of educational assessment only in a few sessions within educational psychology courses or a unit in a methods class (La Marca, 2006; Stiggins, 2006; Popham, 2009).

To deal with this problem, the American Federation of Teachers (AFT), NCME, and National Education Association (NEA) collaborated to develop the 'Standards for Teacher Competence in Education Assessment of Students' or STCEAS to guide pre-service teachers' and educators' learning and course assessment. This standard remains an important authority in the field of teacher assessment literacy.

Considering the prominence of assessment practices in today's classroom that increasingly emphasize the importance of improving teacher assessment literacy for student learning in the U.S. via statewide accountability projects mandated under the 'No Child Left Behind' act, some education reform programs have been proposed. For instance, professional development devised by the state of Nebraska in 2000,

Table 1
Comparing functions of seven WBT system (Wang et al., 2008).

WBT systems	Assembling				Administering	Appraising
	Step 1 & 2	Step 3	Step 4	Step 5	Step 6	Step 7
Gateway Testing System	X	X	V	V	V	□
Lon-CAPA	X	X	V	V	V	□
Mallard	X	X	V	V	V	□
Question Mark	X	X	V	V	V	□
Top-Class	X	X	V	V	V	□
Web@ssessor	X	X	V	V	V	□
WATA	V	V	V	V	V	V

X = not available; V = available; □ = partially available; Step 1: Determining the purpose of testing; Step 2: Constructing a two-way chart; Step 3: Selecting appropriate items according to the two-way chart; Step 4: Preparing relevant items; Step 5: Assembling the test; Step 6: Administering the test; Step 7: Appraising the test.

known as Nebraska's School-based Teacher-led Assessment and Reporting System (STARS), had a positive impact on teacher confidence, knowledge, and skill in classroom assessment (Bandalos, 2004; Lukin et al., 2004).

Some assessment researchers and academic organizations have interpreted assessment literacy from their own point of view. Stiggins (1995) provided a similar description when he stated that, 'Assessment literates know the difference between sound and unsound assessments. knowing what they are assessing, why they are doing it, how best to assess the skill/knowledge of interest, how to generate sound examples of student performance, what can potentially go wrong with the assessment, and how to prevent those problems before occurring. Paterno (2001) and Mertler (2004) defined assessment literacy as 'the possession of knowledge about the basic principles of sound assessment practices, including terminology, development and the use of assessment methodologies, techniques, and familiarity with alternatives to traditional measurements of learning. Assessment literacy has been defined as what a teacher must know and to understand the principles of sound assessment (Stiggins, 2006; Stiggins et al., 2004). Five essential competencies of sound classroom assessment practice are described as follows. 1. Why assess? Assessment procedures and results serve clear and appropriate purposes. 2. Assess what? Assessments reflect clear and valued learning targets. 3. Assess how? Learning targets are translated into assessments that yield accurate results. 4. Communicate how? Assessment results are managed well and communicated effectively. 5. Involve students how? Students are involved in their own assessment (Stiggins, 2006, p. 18; Stiggins et al., 2004, p. 27). The Center for School Improvement and Policy Studies (2007) defined assessment literacy as being able to recognize sound assessment, evaluation, assessments, and communication practices; realize which methods of assessment to use to gather reliable information and students' learning achievement; communicate assessment results effectively, whether using school report cards, grades, test scores, portfolios, or conferences; and optimize students' motivation and learning by encouraging students to conduct assessment, keep records, and communicate with each other. The North Central Regional Educational Laboratory provided a simpler definition, which states that assessment literacy is the readiness of an educator to design, implement, and discuss assessment strategies (Wang et al., 2008). Assessment literacy involves the understanding and appropriate use of assessment practices combined with the knowledge of the theoretical and philosophical underpinnings involved in the measuring of students' learning (DeLuca & Klinger, 2010).

Giving teachers opportunities for enhancing their assessment literacy should be a goal of teacher education (Mertler, 2004). Thus, in Taiwan, there is a need to investigate and improve teachers' assessment literacy. This research adopts the definition of assessment literacy by Wang et al., (2008), including the three aspects of 'assessment knowledge,' 'assessment perspective,' and 'assessment skill.' However, the investigation of assessment skill is quite limited. Assessment skill relates to content knowledge, and the samples in this research consist of a limited number of participating teachers in different subjects ranging from mathematics, physics, chemistry, and biology. This research analyzes the components of assessment literacy as two aspects, teacher assessment knowledge and assessment perspectives.

2.2. Assessment literacy development model

Many researchers show that assessment training in teacher education is important to improving teachers' assessment literacy (Plake & Impara, 1993; Schafer, 1993; Stiggins, 2001). Recent studies have proposed an increasing number of effective assessment literacy development models to build teacher capacity in assessment literacy, an introduction of four effective models is as follows.

Plake and Impara (1993) Model: This model selects the topic of 'Interpreting and Communicating Test Results' for developing a training prototype. The training module design performs in either one or two days of an in-service teacher program or through a series of 2-h

Table 2
The Triple-A model functions in WATA system.

Triple-A model	Complete Assessment Process (Gronlund & Linn, 1990)	Figures	Description
Assembling	Step 1: Determining the purpose of testing	Fig. 4	Teachers follow Step 1 to 5 of the basic steps in classroom testing to assemble tests in WATA system.
	Step 2: Constructing a two-way chart	Fig. 5	
	Step 3: Selecting appropriate items according to the two-way chart	Fig. 6	
	Step 4: Preparing relevant items	Fig. 7	
	Step 5: Assembling the test	Fig. 8	
Administering	Step 6: Administering the test	Fig. 9	Teachers administer a multi-examination schedule.
Appraising	Step 7: Appraising the test	Fig. 10	After tests are taken, teachers perform test and item analysis.

1. Through item analysis, the difficulty (P) and discrimination (D) analysis for the four questions are as follows. According to the Norm-referenced test, which one is the best question? (second question in AKT)
   a) P=1; D=1 b) P=.4; D=.6 c) P=.6; D=.4 d) P=.9; D=.1
2. According to Bloom's taxonomy principles, which kind of statement does the following question belong to? (third question in AKT)

A boy bought a biscuit and found the Nutrition Facts label on the outside of the package, which contain the following nutrients:

Calories..... 150 kcal	
Carbohydrate.....15g	Vitamin.....0.4mg
Protein.....2g	Folic acid.....98microgram (μg)
Fat..... 5g	Calcium.....400mg
Minerals.....1.5g	Iron.....5mg

Where do the Calories come from?
I. Carbohydrate, Protein, and Fat II. Minerals and Vitamin
III. Folic acid IV. Calcium and Iron

a) Comprehension b) Synthesis c) Evaluations d) Analysis

Fig. 1. Sample questions measuring teachers' assessment knowledge concepts in AKT.

training sessions, including a 'parent-teacher vignette,' 'course training materials,' 'practice exercises,' 'questions for a small group discussion,' and 'assignment problems,' as components. The vignette contains three main segments: Introduction: a brief description of the purpose of the training modules, and a focus on the vignette to interpret and communicate the results of student academic achievement assessment. Simulated parent-teacher conference: teachers simulate how to interpret the achievement information to parents. Conclusion: identifying the important features of the parent-teacher conference and introducing the instruction components. This model provides teachers a simulated environment to interpret assessment scores to parents, and provides an opportunity for teachers to practice assessment knowledge.

The Nebraska Assessment Cohort Model: The University of Nebraska-Lincoln proposed the NAC program in 2001. The program requires 18 credits, which contain a 6-h course in two consecutive summers and a 6-h practicum course in an academic year. The course in the first summer semester covers main and related basic assessment concepts of regional assessment and the content includes, 'how to construct a high-quality assessment,' and 'how to use assessment results to make instructive choices.' The course in the second year focuses on analysis and interpreting assessment data and making data-driven decisions.

In addition, the 6-h practicum course requires participants to learn from the Internet, including meeting professors during the intervals. All professors and teachers share their thoughts with others on the Internet, receive feedback, and discuss issues. The course includes five other features: encourage teachers to discover and use materials that relate to their own interests and needs; participants exchange their thoughts and feedbacks, and apply technology adequately to assist in instruction. The course demonstrates each step of technology that participants perform and introduces statistical software, requiring teachers to show the outcome; constructs an instruction platform to provide participants the latest information for contacting peers and professors, and all teachers need to make and exchange their own electric portfolios, containing their knowledge/thoughts of the practicum course. The program aims at providing teachers opportunities to practice their assessment knowledge and skill.

Assessment Literacy Learning Team Model (Arter, 2001; Stiggins, 2001): The ALLT developed a set of assessment literacy courses in 1999. The basic premise of the course is to invite in-service teachers and their principal to participate voluntarily in a small professional development group by meeting regularly and sharing professional experiences. The ALLT model is a framework of a learning team. The model features are as follows: provide high-quality learning sources; emphasize classroom practice; work meaningfully with peers and share successful and failed experiences; and make learning meaningful.

Pre-service Assessment Learning Study group and In-service and Pre-service Assessment Learning Study group Models (Lukin et al., 2004): The PALS and IPALS models developed from ALLT. Participants in the PALS model consist of pre-service teachers as learners and in-service teachers as educators. This model emphasizes practical classroom application of pre-service teachers. Participants meet once a week in the first semester to study and discuss the edited Stiggins contents, and design practical research for the second semester by following Kemmis and McTaggart (1990) proposed plan-act-observe-reflect model, which helps pre-service teachers implement assessment concepts into a realistic practice environment and evaluate the influences of activities to students' learning. Participants in the IPALS model consist of in-service and pre-service teachers as learners. The PALS and IPALS model arrange pre-service teachers and in-service teachers in pairs and combine assessment-related knowledge and techniques with classroom experiences, which are pivotal and challenging for pre-service teachers in the teacher education program.

Table 3
The MTMM results in AKT and SAP.

	AKT	SAP
AKT	0.993[a]	
SAP	0.441[b]	0.660[a]

[a] the Cronbach alpha of AKT and SAP.
[b] the Pearson's correlation coefficient between AKT and SAP.

Table 4
Paired *t*-test result of in-teachers' assessment knowledge ($n = 47$).

AKT	Mean	SD	t value	Effect size
Pre-test	61.17	8.37	2.82**	0.42
Post-test	64.89	9.34		

**$p < 0.01$.

Each of the above four models has important characteristics and effectiveness in developing teachers' assessment literacy. All models emphasize the importance of integrating classroom experiences into assessing literacy development. The models offer opportunities for teachers to practice and apply assessment knowledge and techniques in realistic classroom environments, and teach participants assessment-related content in a traditional way. However, this research not only focus on practicing and sharing assessment knowledge but quickly providing opportunities for teachers to revise self-made tests and alter instruction based on students' responses. From a constructivism point of view, the change in the teachers is equal to their learning (Polettini, 2002). Teacher learning is a self-regulation process in internal cognitive conflict as well as being personal, reflective, and transformative (Sandholtz, 2002). Learning is regarded as a continuous process; individuals and groups construct meaningful knowledge through interactions in the social environment, experience sharing, communication with others, and revision (Zaslavsky & Leikin, 2004). From this perspective, the constructivism theory underpinning it states that reflection, revision, and feedback on the learning process are essential components. Reflecting and revising based on students' responses, especially from web-based educational statistic information, gives teachers opportunities to change assessment methods, enhance instruction, and improve students' learning. To reach this goal, this research takes validated components (i.e., share, interaction, and realistic classroom experiences) of the four models as references and applies the advantages of information technology to confirm the effectiveness of in-service teachers' assessment literacy development in an e-Learning environment.

2.3. P2R-WATA assessment literacy development with web-based assessment and test analysis system

With the rapid development of Internet and communications technology, ICT is emerging as a useful supplement to traditional methods. Web-based testing has recently become a popular and effective assessment tool in the educational field around the world (Costa, Mullan, Kothe, & Butow, 2010; Sheader et al., 2006; Wang, 2007). The Web-based assessment tool is growing and allows teachers to administer test items efficiently to students, provides flexibility to assess anytime and anywhere and delivers immediate feedback to test takers on-line, including correcting and recording personalized test scores. Gardner, Sheridan, and White (2002) indicated that a good assessment system should be able to build a test item pools and score bulletin board. These designs help students to take tests on-line and gather their own learning scores rapidly.

Table 5
Assessment concepts of average incorrect-answer rate in AKT.

Summary of assessment conceptions	Pre-test (%)	Post-test (%)	Improving rate (%)
1. Constructing items, assembling test papers and administering tests			
1.1 Principles of constructing a multiple-choice item	19.15	19.15	0.00
1.2 Characteristics of multiple-choices item	42.55	31.91	10.64
1.3 Bloom's Taxonomy	55.32	50.35	4.97
1.4 Difference between summative assessment and formative assessment	32.98	26.60	6.38
1.5 Functions of summative assessment and formative assessment in teaching activity	37.24	30.85	6.39
1.6 General steps of administering tests	7.80	4.26	3.54
1.7 Administering tests	12.76	10.64	2.12
2. Analysis and appraising of testing data			
2.1 Test analysis			
2.1.1 Variance	51.06	35.11	15.95
2.1.2 Average	2.13	2.13	0.00
2.1.3 Standard deviation	38.30	38.30	0.00
2.1.4 KR20 and Cronbach's alpha reliability	63.12	58.16	4.96
2.1.5 Test difficulty	4.25	0	4.25
2.1.6 Validity (content validity - Specification Chart)	20.57	12.77	7.80
2.1.7 Analysis of In-services' scores distribution	80.85	74.47	6.38
2.1.8 T-scores	76.53	73.47	3.06
2.1.9 Z-scores	40.43	36.17	4.26
2.1.10 Normal Distribution	21.27	12.77	8.50
2.2 Item analysis			
2.2.1 Options distracters power analysis	59.57	57.45	2.12
2.2.2 Item discrimination analysis	56.38	46.81	9.57
2.2.3 In-services' error-concepts analysis	39.36	34.04	5.32
2.2.4 Item difficulty analysis	46.81	29.79	17.02

Table 6
Paired *t*-test result on assessment knowledge of participants with different prior knowledge.

Group	AKT	Mean	SD	*t* value	Effect size
High score ($n = 14$)	Pre-test	70.89	5.87	0.07	0.02
	Post-test	71.07	10.03		
Low score ($n = 33$)	Pre-test	57.05	5.09	3.43**	0.81
	Post-test	62.27	7.47		

**$p < 0.01$.

Gronlund and Linn (1990) posited that the 'basic steps in classroom testing' include the following seven steps. Step 1: Determining the purpose of testing; Step 2: Constructing a two-way chart; Step 3: Selecting appropriate items according to the two-way chart; Step 4: Preparing relevant items; Step 5: Assembling the test; Step 6: Administering the test; Step 7: Appraising the test. Wang et al., (2008) indicated differences between the Web-based Assessment and Test Analysis system (WATA system) (Wang, Wang, Wang, Huang, & Chen, 2004) and other WBT system functions using the perspective of Gronlund and Linn (1990). The result reveals that the WATA system is the only system that allows teachers to determine the purpose of testing and to construct a Two-Way Chart in step 1 to 3 of the basic steps in classroom testing. At the 'appraising the test' step, other assessment systems perform original scoring data, statistic charts, and test results without item analysis, such as the Gateway Testing System, Lon-CAPA, Mallard, Question Mark, Top-Class and the Web@ssessor system. Table 1 show the seven different WBT systems functions. According to Bransford, Brown, and Cocking (2000) scaffolding is a temporary support framework provided learners to extend their competencies. Hannafin, Land, and Oliver (1999) differentiated four types of scaffolding by functions, including conceptual, metacognitive, procedural, and strategic. Procedural scaffolding is the guidance on how to utilize available resources, materials, and tools to help learners learn to do a task, action, or process (Hannafin, Land, & Oliver, 1999; Luke, Freebody, Cazden & Lin, 2005). The WATA system plays the role of procedural scaffolding due to this system equipped with the framework of 'Complete Assessment Process' based on the seven steps proposed by Gronlund and Linn (1990). Moreover, hyperlinks for the reference materials of assessment content were embedded in each step to guide and help learners understand assessment-related terminologies and the meaning of each step. If necessary, learners can click the hyperlink then reference materials for assessment content was shown, as shown in Fig. 7. Learners can obtain assessment-related knowledge and practice in a realistic context through the scaffolding provided by WATA system. The webpage interface of the seven steps in WATA system is as shown in Figs. 4–10. The WATA system is equipped with the framework of 'Complete Assessment Process' proposed by Gronlund and Linn (1990) and offers teachers the Triple-A Model (Wang et al., 2004), which plays a pivotal role in procedural scaffolding of 'basic steps in classroom testing.' Table 2 shows the components of the Triple-A Model include:

_ assembling (teachers follow Step 1 to 5 of the basic steps in classroom testing to assemble tests);
_ administering (teachers administer a multi-examination schedule in Step 6);
_ appraising (after tests are taken, teachers perform test and item analysis in Step 7);

A review of e-Learning environment-related papers indicates that personalized design is important to teacher education (Egbert & Thomas, 2001; Passig, 2001). The WATA system takes the suggestion as a reference to construct an individual interface for learners and provides electronic learning resources in each page of the Triple-A Model. The situated design is important in teacher education (Wiske, Sick, & Wirsig, 2001; Clarke & Hollingsworth, 2002). Because of IT advantages, the e-Learning environment provides a simulated instructional setting for teachers. The Triple-A Model provides strategic scaffolding to support teachers in applying assessment knowledge, principles, and experiences to new contextualization situations.

Wang et al., (2008) developed the P2R-WATA Assessment Literacy Development Model (P2R-WATA) to improve the assessment literacy of pre-service teachers. All participating pre-service teachers simulate assembling tests according to the assessment process in the WATA system, and 'practice' administering and appraising tests on-line, to 'reflect' the statistical data from students' learning, and to 'revise'

Table 7
Cochran's Q result on teachers' assessment perspectives.

Items	High Score Group	Low Score Group	Totals
To assess perspectives on assessment functions,			
1. Discriminate student's learning effects	N/A	3.000	3.000
2. Realize teaching material concepts which students easily confuse	N/A	1.000	1.000
3. Analyze different alternative concepts of individual students	1.000	2.000	0.333
4. Analyze and improve teachers' instruction strategy	N/A	1.000	1.000
5. Realize advantages and disadvantages of question assignments	N/A	N/A	N/A
6. Realize advantages and disadvantages of each assigned item	7.000**	14.222**	21.160**
7. Realize advantages and disadvantages of the complete testing	6.000**	3.600	9.000**
To assess perspectives on administering classroom assessment,			
1. Confirm the purpose of testing	N/A	2.000	2.000
2. Establish a Two-way Chart	2.000	3.000	4.571*
3. Use a Two-way Chart to administer the test	4.000*	14.222**	18.182**
4. Select proper question patterns and questions	1.000	8.000**	9.000**
5. Administer a test	1.000	1.000	2.000
6. Take a test	N/A	N/A	N/A
7. Assess a test score	N/A	1.000	1.000
8. Perform a test analysis and apply the result to your instruction	1.000	6.000*	7.000**

*$p < 0.05$; **$p < 0.01$; N/A: not available.

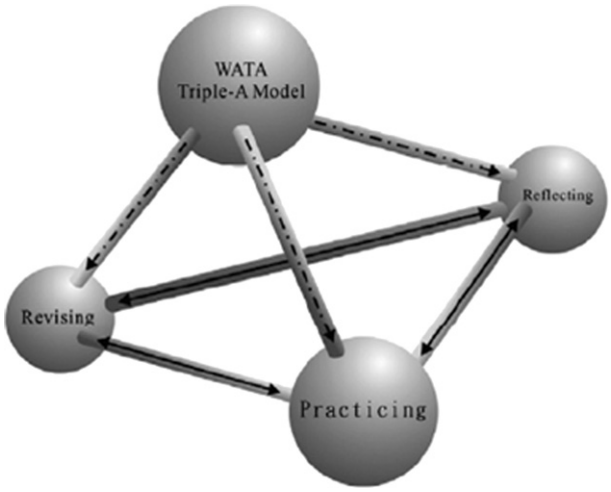

Fig. 2. P2R-WATA Assessment Literacy Development Model (Wang et al., 2008).

problematic test items. The result indicates that P2R-WATA helps pre-service teachers improve their assessment knowledge and perspectives. Participants showed positive perspectives toward the WATA system. This research adopts the P2R-WATA Model to confirm development and effectiveness of assessment literacy for in-service teachers.

3. Methodology

3.1. Participants

Participants in this research were forty-seven of secondary in-service mathematics and natural science teachers enrolled in the summer program entitled, 'Test and Assessment in Mathematics and Natural Science' in the Educational University, Taiwan. The research duration was two years. The samples in the first year consisted of twenty-three teachers and twenty-four teachers for the second year. Classroom instruction was conducted by the same professor. During the 36-h, six-week period, all in-service teachers received an equal amount of instructional time and were provided with the same materials, assignments, and e-Learning environment.

3.2. Instruments and materials

3.2.1. Assessment knowledge test (AKT) (Wang et al., 2008)

AKT was mainly used to assess teachers' understanding of assessment concepts (e.g., principles of constructing multiple-choice items, Bloom's taxonomy, reliability, validity, scores distribution, distracter ability, discrimination, general schedule of administering a test, and

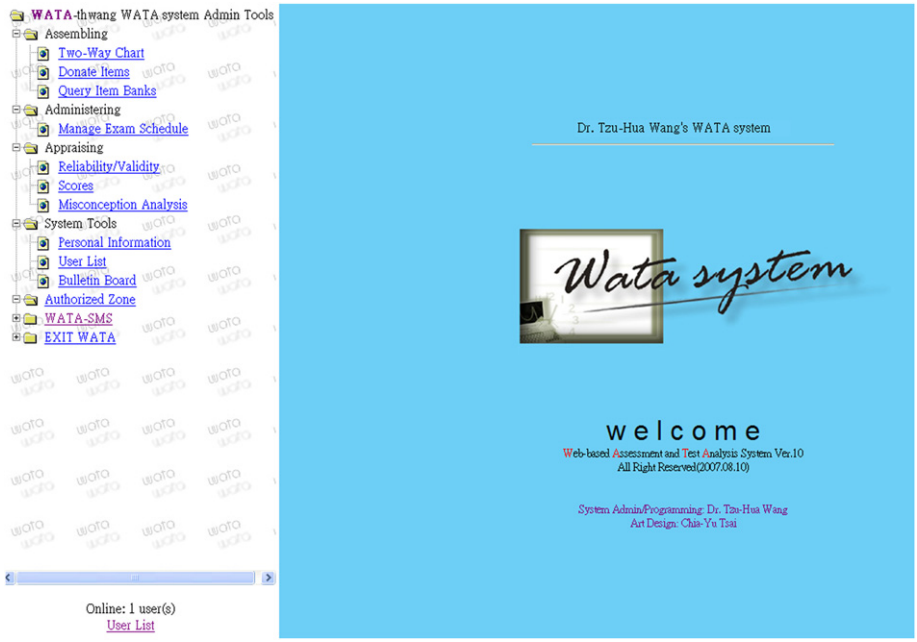

Fig. 3. Personalized Triple-A Model of the WATA system (Wang et al., 2008).

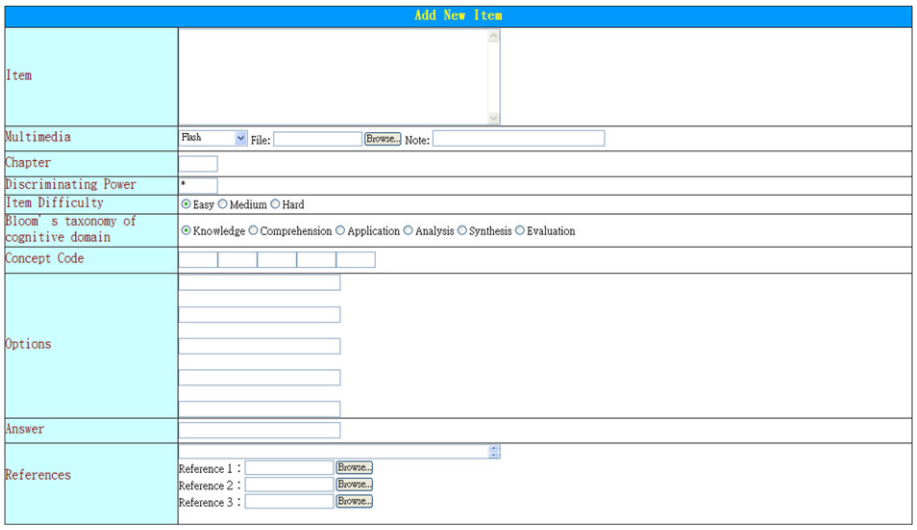

Fig. 4. Personalized construct test item.

item difficulty index). Fig. 1 presents two examples of test questions. Table 5 shows a list of all assessment concepts. Those in the AKT were established based on the ideas of STCEAS (AFT, NCME, & NEA, 1990), the basic assessment concepts proposed by Gronlund and Linn (1990) and three education and assessment experts chose the most important assessment concepts for teachers. This research selected twenty-one assessment concepts to form forty multiple-choice questions, administered with the WATA system. The test items included one correct answer and three possible responses. The same concept was tested using two different types of questions. All questions were piloted and required modifications were made prior to administering the test. Cronbach alpha for the AKT was 0.993 and its average difficulty was 0.505 (Wang et al., 2008). This research adopted the Multitrait-Multimethod matrix to analyze the data acquired by AKT and SAP (Survey of Assessment Perspectives). Table 3 show that the diagonal values (i.e., the Cronbach alpha reliability coefficients of AKT and SAP) are larger than others (i.e., the Pearson's correlation coefficients between AKT and SAP). Therefore, the data acquired by AKT and SAP have fair discriminant validity (Campbell & Fiske, 1959).

3.2.2. *Survey of assessment perspectives (SAP) (Wang et al., 2008)*

SAP measures teachers' perspectives toward assessment functions and procedures. The SAP was established based on Gronlund and Linn (1990) assessment ideas. Three educational experts carried out the content validity of the test items. The SAP format is a binary categorical scale based on participants' expression of agreement or disagreement (Yes/No questions). The scale included fifteen items with two subscales, including 'perspectives about assessment functions,' consisting of seven items (Cronbach's alpha = 0.71), and 'perspectives about assessment steps,' consisting of eight items (Cronbach's alpha = 0.78) (Wang et al., 2008). Table 7 presents the test questions.

3.2.3. *Practicing, reflecting, and revising with web-based assessment and test analysis*

This research used the WATA system to construct an assessment literacy development model called the 'P2R-WATA' model (Fig. 2). The WATA system belongs to a regular WBT response system equipped with a personalized Triple-A (Assembling, Administering, and Appraising) Model (Fig. 3) (Wang et al., 2004). The Triple-A Model is based on the 'basic steps in classroom testing,' proposed by Gronlund and Linn (1990) and interviews seventeen in-service teachers and assessment experts (Wang et al., 2008). The 'Assembling' function empowers teachers to construct item pools, test items, and administer tests based on the 'Two-way Chart' (Fig. 5; Fig. 6). The 'Administering' function enables teachers to manage test schedules, and monitor the testing process, including collecting and recording scores. The function provides personal identification numbers (PINs) and examination passwords for testees to take the test over the Internet. The 'Appraising' function automatically reports test analysis data with educational statistical information to teachers, including the difficulty index, discrimination index, and the optional distracter power of the items (Fig. 10).

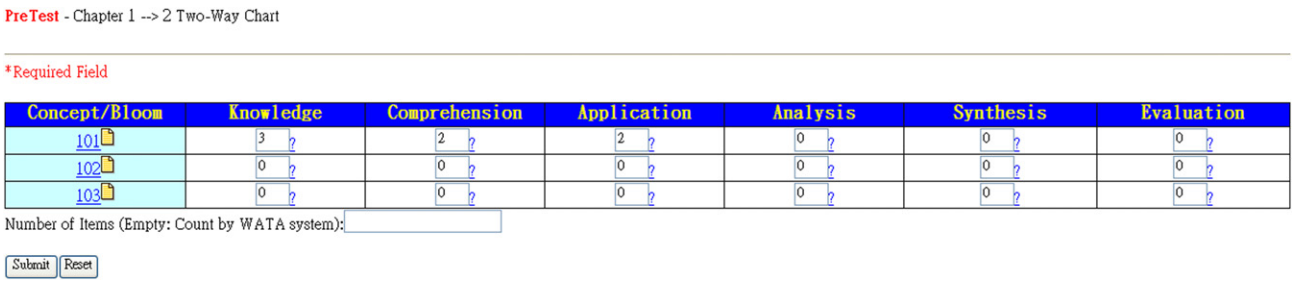

Fig. 5. Assemble a test according to the Two-Way Chart.

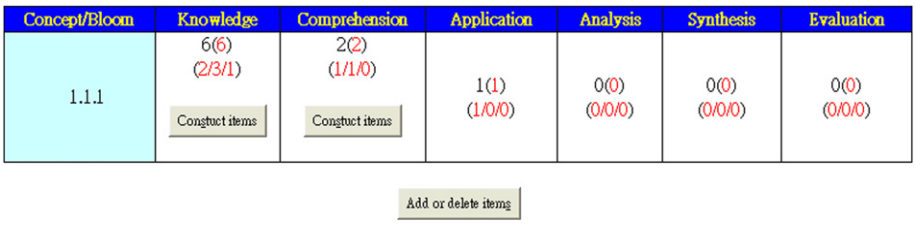

Fig. 6. Selecting appropriate items according to the two-way chart.

3.2.4. *Design of e-learning contents for assessment literacy training*

The e-learning design assists teachers in learning to administer assessment. This research adopted all the on-line reference material of assessment content by Wang et al., (2008). The on-line learning materials were taken from a contemporary educational assessment literature based on the ideas of STCEAS (AFT et al., 1990), the basic assessment concepts proposed by Gronlund and Linn (1990) and three experts in science education and assessment chose the most important assessment concepts for teachers to use in the summer program.

The on-line reference material of assessment content contained two main topics, including 'Constructing items, assembling test papers, and administering tests' and 'Analysis and appraisal of testing data.' The topic of constructing items, assembling test papers and administering tests included principles of constructing a multiple-choice item, characteristics of multiple-choice items, Bloom's Taxonomy, the difference between summative assessment and formative assessment, functions of summative assessment and formative assessment in teaching activity, general steps of administering tests, and administering tests. The topic of appraising and applying test results included variance, average, standard deviation, KR20 reliability, average difficulty of the test, validity, score distributions, optional distracter power analysis, ID, *T* score, *z* score, and *Z* score, incorrect answers of students for each item, and item discrimination and difficulty analysis.

3.3. *Research design*

This research adopted a single-group experimental design. The participants used the P2R-WATA Assessment Literacy Development Model (P2R-WATA), based on the WATA system with a Triple-A Model framework. The Triple-A Model provides teachers with the 'framework of complete assessment steps,' which acts as a pivotal scaffold in an assessment literacy training process. Two essential components of P2R-WATA were the personalized and situated design. The personalized design was based on the suggestions of Egbert and Thomas (2001) and Passig (2001). The personalized interface in an e-Learning environment for teacher education should adapt to individual needs of each teacher. The WATA system provides teachers with personalized learning resources in the e-Learning environment, enabling participants to enhance their assessment knowledge and perspectives using the Triple-A Model in their personalized learning process.

The situated design was based on the most important characteristic in the four assessment literacy development models, offering opportunities for teachers to apply what they learned to realistic classroom situations. Participants 'practice' assembling, administering tests on-line, and appraising test-related statistical information generated by the WATA system. Using statistical information, teachers 'reflect' on problematic test items, student learning and instructions, 'revise' certain test items, and clarify instructions.

3.4. *Research procedures*

The research procedures were as follows. The summer program began with the AKT and SAP pre-test administered to the forty-seven in-service teachers to ascertain their entry behavior of assessment knowledge and perspectives. Second, the participants were asked to administer a test through the three-stages, including practicing, reflecting, and revising the P2R-WATA model process.

During the practice stage, teachers read e-learning materials on assessment built in the WATA system and 'practice' administering a test with five multiple-choice questions based on their own teaching subjects and the Two-Way Chart. After constructing the items and assembling tests based on the WATA system, the tests were sent to some cooperative middle schools and some suitable classes were invited to test students. All students had been taught these mathematic and science subjects previously. Students took the test on-line via the WATA system in a computer classroom.

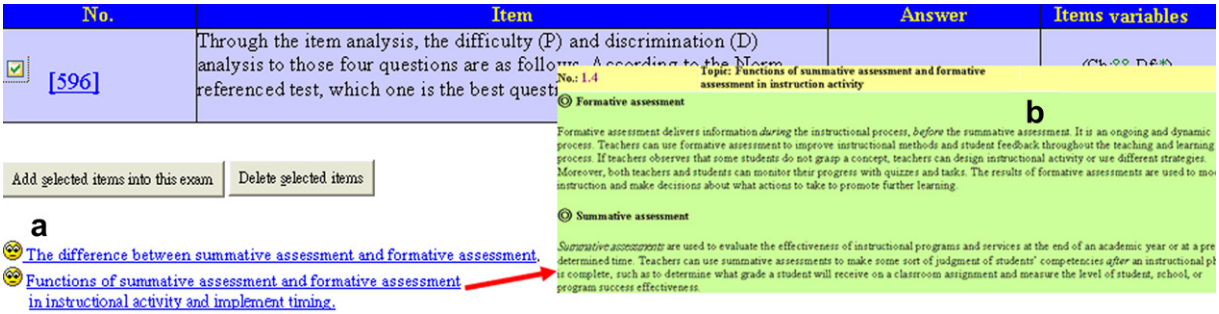

Fig. 7. Teachers enable to prepare relevant items, when click on the hyperlink (a) then the reference materials of assessment content will be shown (b).

Variables	Settings and values
> Setting type	◉ New exam ○ Modify
> Exam explanation	
> Exam name	AKT_final
> Exam coverage	
> Present correct answers	◉ Yes ○ No
> Email students' exam records	◉ Yes ○ No
> Query peers' scores	◉ Yes ○ No
> Take this exam repeatedly	◉ Yes ○ No
> Number of items(Totally items)	40
> Primary E-mail address to receive students' exam records and questions	
> Secondary E-mail address to receive students' exam records and questions	
> Time limitation for this exam (mins)	

Fig. 8. Teachers enable to assemble the test in step 5 (Wang, 2008).

During the reflecting stage, the teachers were requested to 'reflect' on the results of the tests with the individual statistics information provided automatically by the WATA system after students completed the tests. After reflecting, the teachers revised the test items directly in the WATA system.

During the revising stage, teachers were requested to 'revise' problematic test items, such as the poor difficulty index, the distraction index, and the discrimination index. Following the P2R-WATA procedure, in-service teachers took the AKT and SAP post-test at the end of the program to evaluate their learning effectiveness. Through the process of the P2R-WATA model, teachers hopefully improved their understanding of the complete testing process and enhanced their assessment literacy.

3.5. Data collection and analysis

All data collected were quantitative, comprising the AKT and SAP pre-test and post-test scores. Each question was equivalent to 2.5 points and the total score was 100 points in the AKT. The results of the pre-test and post-test were analyzed by the paired *t*-test and effect size. According to the definition of Cohen (1988), the general conventions for labeling effect sizes are as follows: 0.2 is small, 0.5 is medium, and 0.8 is large. Meanwhile, the SAP was a binary categorization scale, including fifteen questions with a total score of 15 points. Each question with a 'Yes' answer was equivalent to one point, while a 'No' answer was equivalent to zero. The Cochran Q statistic method was used to determine changes in the participants' assessment perspectives.

4. Results

4.1. Development of assessment knowledge

To measure the development of in-service teachers' assessment knowledge in the P2R-WATA model training, we tested participants with the AKT pre-test and post-test. Table 4 shows the results of the paired *t*-test. A significant mean difference existed between AKT pre-test and post-test scores ($t = 2.82$, $p < 0.01$) and the effect sizes was small to medium (Cohen's $d = 0.42$). The results confirm that in-service teachers in the summer program of 'Test and Assessment of Mathematics and Natural Science' using the P2R-WATA Assessment Literacy Development model performed better and significantly improved assessment knowledge.

4.2. Development of assessment knowledge concepts

The AKT was a 40-item multiple-choice test that included twenty-one educational assessment concepts. To assess participants' understanding of assessment concepts, Table 5 shows the results of the pre-test and post-test incorrect-answer responses of AKT. Before treatment, the average percentage of incorrect responses from participants was 38.50% in the pre-test AKT. After treatment, the average percentage of incorrect responses was 32.63% in the post-test AKT. The average percentage of three assessment concepts on the pre-test and post-test incorrect-answer responses did not change. However, the concepts in principles of constructing a multiple-choice item, average,

Examiner and Exam name	Exam variable setting	Exam coverage	Time, password and testee account setting
Teacher 1	> Set exam variables > Composing exam from Two-way Chart > List the items in this exam > Print the items in the form of paper-and-pencil test > Manage the answering history of testees > Manage the correct answers of items > Duplicate this exam	1 -- 1	2011 Year 02 Month 08 Day 11 Hour 57 Minute Hours (can login) PIN code Modify/optimize Manage testee accounts

Fig. 9. Teachers enable to administer the test in step 6 (Wang, 2008).

General Test Analysis

Testee	47
Variance	82.217
Standard Deviation	9.339
Average	61.17
Average Difficulty	61.17%

Score Analysis

Ranking	Testee	Scores
1	p2r0501	87.5
2	p2r0511	85
3	p2r0537	77.5
4	p2r0513	70
4	p2r0522	70

Item Analysis & Distracter Analysis

Items	A	B	C	D	N/A	Num	DP	ID
Through the item analysis, the difficulty (P) and discrimination (D) analysis to those four questions are as follows. According to the Norm-referenced test, which one is the best question? (A). P=1 ; D=1 (B). P=.4 ; D=.6 (C). P=.6 ; D=.4 (D). P=.9 ; D=.1	0% H:0% L:0%	55.56% H:100% L:50%	44.44% H:0% L:50%	0% H:0% L:0%	0% H:0% L:0%	47	0.51	56%
According to Bloom's taxonomy principles, which kind of statement does the following question belong to? A boy bought a biscuit and found the Nutrition Facts label on the outside of the package, which contain the following nutrients? Calories…… 150 kcal Carbohydrate…15g / Vitamin…0.4mg Protein…2g / Folic acid… 98 microgram (μg) Fat…5g / Calcium…400mg Minerals…1.5g / Iron…5mg Where do the Calories come from? I. Carbohydrate, Protein, and Fat II. Minerals and Vitamin III. Folic acid IV. Calcium and Iron (A). Comprehension (B). Synthesis (C). Evaluations (D). Analysis	12.5% H:0% L:50%	37.5% H:0% L:0%	0% H:0% L:0%	50% H:100% L:50%	0% H:0% L:0%	47	0.5	50%

Fig. 10. Test and Item analysis.

and standard deviation need reinforcing and further discussion with qualitative data analysis to better train in-service teachers. After the model training, the other eighteen assessment concepts showed positive development.

4.3. Assessment knowledge development of participants with different prior knowledge

This research used a paired *t*-test to examine effectiveness of the P2R-WATA model on in-service teachers with different prior knowledge. We divided participants into two levels using the median of AKT pre-test scores. Out of the participating forty-seven in-service teachers, this work selected fourteen teachers from the high-level prior knowledge group and the other thirty-three from the low-level prior knowledge group to compare their AKT pre-test and post-test scores. The results indicated that the low-level prior knowledge group showed a significant difference in their assessment knowledge after training ($t = 3.43$ $p < 0.01$, see Table 6) and the effect sizes was large (Cohen's $d = 0.81$). Fig. 11 shows that the low-level prior knowledge group significantly converged toward that of the high-level-prior knowledge

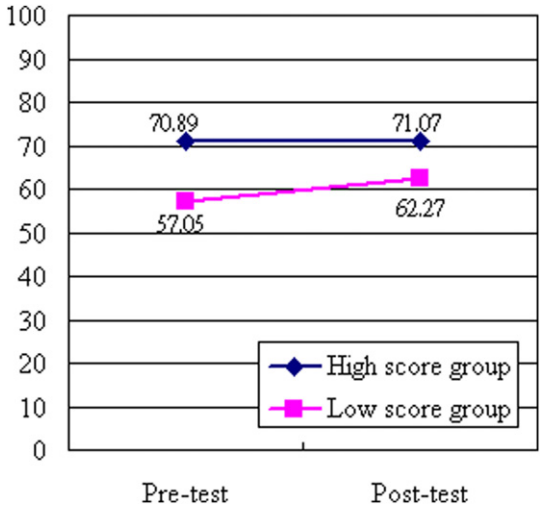

Fig. 11. Graph of interaction between Groups and AKT scores.

group. The P2R-WATA model had a positive effect on improving in-service teachers' assessment knowledge, especially for teachers with low-level prior knowledge.

4.4. Changes in assessment perspectives of teachers

To measure the development of in-service teachers' assessment perspectives in the P2R-WATA model training, we tested participants with the SAP pre-test and post-test. This research presented an analysis of their SAP pre-test and post-test for each question using a nonparametric Cochran's Q test. Table 7 indicates that teachers' assessment perspectives changed significantly in the six items, including realizing advantages and disadvantages of each assigned item (Cochran's $Q = 21.16$, $p < 0.01$), realizing advantages and disadvantages of the complete testing (Cochran's $Q = 9.00$, $p < 0.01$), establishing a Two-way chart (Cochran's $Q = 4.57$, $p < 0.05$), using a Two-way Chart to administer a test (Cochran's $Q = 18.18$, $p < 0.01$), selecting proper question patterns and questions (Cochran's $Q = 9.00$, $p < 0.01$), and performing a test analysis and applying it to instruction (Cochran's $Q = 7.00$, $p < 0.01$).

The high-level prior knowledge group exhibited positive changes in eight items of a fifteen-item questionnaire (53.3%). Significant changes among the three items included realizing advantages and disadvantages of each assigned item, realizing advantages, and disadvantages of the complete testing, and using a Two-way Chart to administer the test. The low-level prior knowledge group exhibited positive changes in thirteen items (86.7%). Significant changes in four items included realizing advantages and disadvantages of each assigned item, using the Two-way Chart to administer the test, selecting proper question patterns and questions, and performing test analysis and applying the result to their instructions. In general, the P2R-WATA model had a positive change in in-service teachers' assessment perspective. This model had significant effect in convincing in-service teachers that the Two-way Chart is essential and in the effect of each assigned item for assessment.

5. Conclusion

This research determines the effectiveness of the P2R-WATA model in facilitating the assessment knowledge and perspectives of secondary in-service mathematics and natural sciences teachers. The findings reveal that the P2R-WATA model has significant effectiveness in improving the participating in-service teachers' assessment knowledge and perspectives, especially those in the low-level prior knowledge group. These results are consistent with the findings of previous studies that acknowledge the assessment literacy training model as necessary and suitable for teachers, and explain the reason for the P2R-WATA design (Wang et al., 2008). The characteristics of the P2R-WATA are based on the suggestions of the personalized e-Learning environment design by Passig (2001) and the situated design by Wiske et al., (2001). The personalized design allows participating teachers to determine their own learning sequences, and use the educational statistic information provided by the WATA system for reflection and revision. The situated design provides teachers with opportunities to put assessment knowledge into realistic practice situations to enhance their assessment literacy.

The assessment perspectives of teachers changed significantly in performing and establishing a Two-way chart for administering and analyzing tests, and applying the result to their own instruction. According to those results above, it can be explained by the reason that P2R-WATA model requires teachers to construct a Two-way chart to administer tests under scaffolding in the WATA system and offers hands-on classroom experience in an instructional setting.

Additionally, based on the findings, this research suggests that using a well-designed web-based assessment system is an important part of the teacher education experience in improving teacher assessment literacy. The design of the Triple-A Model as a scaffolding for a temporary support framework helps teachers complete assessment steps in developing assessment literacy. Attention must focus on offering teachers opportunities to practice assembling, administering, and appraising tests on-line, and applying assessment knowledge to a realistic classroom environment. The educational statistic information from the WATA system requests teachers to reflect, report on test results, and to revise problematic test items. This suggestion is similar to the results of Wang et al., (2008). They assigned pre-service teachers to an experimental group (with an e-Learning environment using the WATA system and personalized electronic learning

resources), and a control group (with a traditional learning environment and printed rather than personalized electronic learning resources), in which the improvement trend of the experimental group was significantly diverse from that of the control group. The Web-based assessment system has the potential to expand and enrich learning opportunities for educators. Therefore, pre-service and in-service teacher education programs should consider the P2R-WATA model for developing assessment literacy.

Future research is required to explore whether other factors might have an impact on assessment literacy for in-service teachers. For example, researchers could collect and analyze information about participating teachers' instructional experiences and their background information. Besides, this research focused on a single group design with the e-Learning environment. Future research could adopt a quasi-experimental design. Moreover, investigating how teachers use assessment results to make decisions, which in turn influences student learning effectiveness is also pivotal to future research. Performance assessment is playing a pivotal role in present education reform; future research we will direct toward teacher professional development in performance assessment in WATA system. Finally, more in-service teachers should be enrolled in this assessment literacy development program, could help teachers with different subject matter backgrounds have the same findings to develop proper assessment training program and curricula.

Acknowledgement

The authors deeply appreciate the National Science Council in Taiwan for the financial support and encouragement under Grant No. 96-2511-S-018-006. The authors are grateful for the insightful comments from the referees.

References

American Federation of Teachers, National Council on Measurement in Education, & National Education Association (AFT, NCME, & NEA). (1990). The standards for teacher competence in the educational assessment of students. *Educational Measurement: Issues and Practice, 9*(4), 30–32.
Arter, J. (2001). Learning teams for classroom assessment literacy. *NASSP Bulletin, 85*(621), 53–65.
Bandalos, D. (2004). Introduction to the special issue on the Nebraska Standards-based, Teacher-led assessment and Reporting System (STARS). *Educational Assessment: Issues and Practice, 23*(2), 4–6.
Bodzin, A., & Cates, W. (2003). Enhancing preservice teachers' understanding of Web-based scientific inquiry. *Journal of Science Teacher Education, 14*(4), 237–257.
Bransford, J., Brown, A., & Cocking, R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
Campbell, C., & Collins, V. L. (2007). Identifying essential topics in general and special education introductory assessment textbooks. *Educational Measurement, Issues and Practice, 26*(1), 9–18.
Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.
Center for School Improvement and Policy Studies, Boise State University. (n.d.). What is assessment literacy? Assessment Literacy. Retrieved August 30, 2010, from http://csi.boisestate.edu/al/.
Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*(8), 913–1059.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
Costa, D. S. J., Mullan, B. A., Kothe, E., & Butow, P. (2010). A web-based formative assessment tool for master's students: a pilot study. *Computers & Education, 54*(4), 1248–1253.
DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice, 17*(4), 419–438.
DeLuca, C., & McEwen, L. (2007, April). Evaluating assessment curriculum in teacher education programs. In *Paper presented at the annual meeting of the Edward F. Kelly evaluation conference, Ottawa, Canada*.
Egbert, J., & Thomas, M. (2001). The new frontier: a case study in applying instructional design for distance teacher education. *Journal of Technology and Teacher Education, 9*(3), 391–405.
Gardner, L., Sheridan, D., & White, D. (2002). A web-based learning and assessment system to support flexible education. *Journal of Computer Assisted Learning, 18*(2), 125–136.
Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan Publishing Company.
Hannafin, M., Land, S., & Oliver, K. (1999). Open learning environments: Foundations, methods and models. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory, Vol. 2* (pp. 115–140). Mahaw, NJ: Lawrence Erlbaum Associates.
Kemmis, S., & McTaggart, R. (Eds.). (1990). *The action research planner*. Victoria, Australia: Deakin University Press.
La Marca, P. (2006, June). Assessment literacy: Building capacity for improving student learning. In *Paper presented at the Council of Chief State School Officers (CCSSO) annual conference on large-scale assessment, San Francisco, CA*.
Luke, A, Freebody, P., Cazden, C., & Lin, A. (2005). *A Coding Scheme for the Analysis of Classroom Discourse in Singapore Schools*. Singapore: Center for Research in Pedagogy and Practice, National Institute of Education.
Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice, 23*(2), 26–32.
McMunn, N., McColskey, W., & Butler, S. (2004). Building teacher capacity in classroom assessment to improve student learning. *International Journal of Educational Policy, Research, & Practice, 4*(4), 25–48.
Mertler, G. A. (2004). Secondary teachers' assessment literacy: does classroom experience make a difference? *American Secondary Education, 33*(1), 49–64.
Mertler, C. A., & Campbell, C. (2005, April). Measuring teachers' knowledge & application of classroom assessment concepts: Development of the assessment literacy inventory. In *Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada*.
Mills, C. N. (Ed.). (2002). *Computer-based testing: Building the foundation for future assessments*. New Jersey: Lawrence Erlbaum.
National Council on Measurement in Education Ad Hoc Committee on the Development of a Code of Ethics. (1995). *Code of professional responsibilities in educational measurement*. Retrieved August 30, 2010, from. http://www.ncme.org.
National Research Council (NRC). (1996). *National science education standards*. Washington, DC: National Academy Press.
North Central Regional Educational Laboratory (n.d.). Indicator assessment. Retrieved August 30, 2010, from http://www.ncrel.org/engauge/framework/pro/literacy/prolitin.htm.
Passig, D. (2001). Future online teachers' scaffolding: what kind of advanced technological innovations would teachers like to see in future distance training projects? *Journal of Technology and Teacher Education, 9*(4), 599–606.
Paterno, J. (2001). *Measuring success: a glossary of assessment terms. Building Cathedrals: Compassion for the 21st Century*. Retrieved August 30, 2010, from. http://www.tcsdk12.org/math/Documents/measuringsuccess.html.
Plake, B. S., & Impara, J. C. (1993). *Teacher assessment literacy: development of training modules*. (ERIC Document Reproduction Service No. ED358131).
Polettini, A. F. F. (2002). Mathematics teaching life histories in the study of teachers' perceptions of change. *Teaching and Teacher Education, 16*(7), 765–783.
Popham, W. J. (2006). Needed: a dose of assessment literacy. *Educational Leadership, 63*(6), 84–85.
Popham, W. J. (2009). Assessment literacy for teachers: faddish or fundamental? *Theory Into Practice, 48*(1), 4–11.
Sandholtz, J. H. (2002). Inservice Training or Professional Development: contrasting opportunities in a school/university partnership. *Teaching and Teacher Education, 18*(7), 815–830.
Schafer, W. D. (1993). Assessment literacy for teachers. *Theory Into Practice, 32*, 118–126.
Sheader, E., Gouldsborough, I., & Grady, R. (2006). Staff and student perceptions of computer-assisted assessment for physiology practical classes. *American Journal of Physiology-Advances in Physiology Education, 30*(4), 174–180.
Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher, 15*(1), 4–14.
Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice, 10*(1), 7–12.
Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*(3), 238–245.

Stiggins, R. J. (1999). Are you assessment literate? *The High School Journal, 6*(5), 20–23.
Stiggins, R. J. (2001). *Student-involved classroom assessment* (3rd ed.). New Jersey: Prentice Hall.
Stiggins, R. (2002). Assessment crisis: the absence of assessment FOR learning. *Phi Delta Kappan, 83*(10), 758–765.
Stiggins, R. (2006). Assessment for learning: a key to motivation and achievement. *Phi Delta Kappan, 2*(2), 1–19.
Stiggins, R., Arter, J., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right–using it well*. Portland, OR: Assessment Training Institute.
Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education, 30*(3), 749–770.
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. (2011). Developing Web-based assessment strategies for facilitating junior high school students to perform self-regulated learning in an e-Learning environment. *Computers & Education, 31*(3), 1801–1812.
Wang, T. H. (2011). Implementation of Web-based dynamic assessment in facilitating junior high school students to learn mathematics. *Computers & Education, 56*(4), 1062–1071.
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. (2008). Web-based formative assessment: Issues and trends. In A. R. Lipshitz, & S. P. Parsons (Eds.), *E-Learning: 21st Century Issues and Challenges* (pp. 1–28). New York: Nova Science Publishers.
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, K. H., Wang, T. H., Wang, W. L., & Huang, S. C. (2006). Learning styles and formative assessment strategy: enhancing student achievement in Web-based learning. *Journal of Computer Assisted Learning, 22*(3), 207–217.
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*(1), 59–71.
Wiske, M. S., Sick, M., & Wirsig, S. (2001). New technologies to support teaching for understanding. *International Journal of Educational Research, 35*(5), 483–501.
Zaslavsky, O., & Leikin, R. (2004). Professional development of mathematics teacher educators: growth though practice. *Journal of Mathematics Teacher Education, 7*(1), 5–32.