

OPASS: AN ONLINE PORTFOLIO ASSESSMENT AND DIAGNOSIS SCHEME TO SUPPORT WEB-BASED SCIENTIFIC INQUIRY EXPERIMENTS

Jun-Ming SU,

Department of Information and Learning Technology, National University of Tainan, Taiwan, R.O.C
jmsu@mail.nutn.edu.tw

Huan-Yu LIN,

Department of Computer Science, National Chiao Tung University, Taiwan, R.O.C
linhuanyu@gmail.com

Shian-Shyong TSENG

Department of Applied Informatics and Multimedia, Asia University, Taiwan, R.O.C
sstseng@asia.edu.tw

Chia-Jung LU

Department of Computer Science, National Chiao Tung University, Taiwan, R.O.C
rebeccalu0926@gmail.com

ABSTRACT

Promoting the development of students' scientific inquiry capabilities is a major learning objective in science education. As a result, teachers require effective assessment approaches to evaluate students' scientific inquiry-related performance. Teachers must also be able to offer appropriate supplementary instructions, as needed, to students. Scientific inquiry capabilities should be assessed by evaluating students' scientific inquiry portfolios in actual hands-on experiments. Although virtual laboratory systems can reduce the cost of conducting scientific inquiry experiments, the manual portfolio assessment approach is still difficult and time-consuming for teachers. Therefore, in this paper, in order to provide students with personalized learning guidance concerning not only the conceptual knowledge, but also the high-order, integrative abilities of scientific inquiry, an Online Portfolio Assessment and Diagnosis Scheme, called OPASS, was proposed to assist teachers in automatically assessing and diagnosing students' abilities as they relate to scientific inquiry performance. Personalized diagnostic reports were generated by employing the rule-based inference approach, which diagnosed learning problems and provided corresponding reasons and remedial suggestions based on teacher-defined assessment knowledge of the scientific inquiry experiment. For the evaluation, experimental results showed that the OPASS was helpful and beneficial for both students and teachers.

INTRODUCTION

Today, Scientific Inquiry (SI)-based learning receives widespread attention. The purpose of such learning is to promote students' knowledge and understanding of scientific ideas as well as how scientists study the natural world (National Research Council [NRC], 1996). If students possess scientific inquiry skills, they are capable of conducting an investigation, collecting evidence from a variety of sources, developing an explanation from the data, and communicating and defending their conclusions (National Science Teacher Association [NSTA], 2004; Handelsman, et al., 2004). Educators should teach students to learn and acquire not only conceptual knowledge, but also scientific inquiry skills. Consequently, the assessment concerning scientific inquiry is necessary and required to foster knowledge and skills of inquiry-based learning. In general, the traditional paper-and-pencil test is a suitable approach to measure students' knowledge of science concepts and scientific inquiry, e.g., Substantive Knowledge. However, it is not easy to assess and evaluate learning problems and performance of higher-order capabilities related to scientific inquiry, e.g., Procedural Knowledge, and Problem Solving and Integrative Abilities (Wenning, 2007; Jacobs-Sera, Hatfull, & Hanauer, 2009; Bennett, Persky, Weiss, & Jenkins, 2007, 2010).

Furthermore, scientific inquiry can be considered as a set of process skills that consists of questioning, hypothesis-making, experimenting, recording, analyzing, and concluding, which can be regarded as "hands-on" learning (NRC, 1996, NSTA, 2004; Ketelhut, Dede, & Clarke, 2010). Nevertheless, learning and assessing in the physical laboratory are inconvenient and time-consuming for both teachers and students (Hanauer, Hatfull, & Jacobs-Sera, 2009, pp. 117-118). With this in mind, a significant amount of research has been dedicated to develop the virtual and Web-based interactive learning systems to support online scientific inquiry learning (Yaron et al., 2008; Hsu, Wu, & Hwang, 2008; Dalgarno, Bishop, Adlong, & Bedgood, 2009; Yaron, Karabinos, Davenport, Leinhardt, & Greeno, 2009; Yaron, Karabinos, Lange, Greeno, & Leinhardt, 2010; Ketelhut et al., 2010). Through this type of learning, students can efficiently improve and foster their experiences and skills based on scientific inquiry learning activities, and their portfolios can be collected for the further analysis. Using

students' portfolios collected from inquiry-based learning activities to manually assess scientific inquiry is an ideal approach (Zachos, Hick, Doanne, & Sargent, 2000; Lunsford & Melear, 2004), but it is not easy to perform and it is time-consuming for teachers (Zachos, 2004; Jacobs-Sera et al., 2009, Hanauer et al., 2009, pp. 117-118, Bennett et al., 2007, 2010).

Moreover, many articles also argue that students should be provided with not only the score of test, but also the individual learning guidance for improving their learning performance. For this reason, several analysis and diagnosis approaches have also been proposed to assess the learning portfolio of students and then offer them the personalized learning guidance related to misconceived notions of a given subject (Hwang, 2003; Kosba, Dimitroca, & Boyle, 2007; Chu, Hwang, & Huang, 2010; Panjaburee, Hwang, Triampo, & Shih, 2010) and scientific inquiry skill scores (Ting, Zadeh, & Chong, 2006; Ting, Phon-Amnuaisuk, & Chong, 2008; Bennett et al., 2007, 2010). Nevertheless, an analysis of learning problems related to scientific inquiry skills needs to diagnose the operational and procedural portfolios so students can understand their learning status and problems in relation to not only scores and concepts, but also to operations and skills of scientific inquiry.

Therefore, in this paper, to provide students with personalized learning guidance concerning not only the conceptual knowledge, but also the high-order, integrative abilities of scientific inquiry, an Online Portfolio Assessment and Diagnosis Scheme, called OPASS, has been proposed. The OPASS is able to efficiently evaluate students' assessment portfolios collected from the Web-based scientific inquiry experiment. It employs the rule-based inference approach to automatically diagnose learning problems related to concepts, cause and effect operations, and skills of scientific inquiry according to teacher-defined assessment knowledge of the scientific inquiry experiment. Consequently, students can be provided with personalized scientific inquiry diagnostic reports to improve not only subject concepts, but scientific inquiry skills as well.

RELATED WORKS

Assessments of Scientific Inquiry

The knowledge and capabilities of scientific inquiry are multidimensional (NRC, 1996; Wenning, 2007; Hanauer et al., 2009, pp. 11-21) and can be divided into three types: (1) Substantive Knowledge, e.g., scientific concepts, facts, and processes; (2) Procedural Knowledge, e.g., procedural aspects of conducting a scientific inquiry; and (3) Problem Solving and Integrative Abilities, e.g., the ability to solve problems, pose solutions, conceptualize results, and reach conclusions (Jacobs-Sera et al., 2009, p. 36). Therefore, the assessment concerning scientific inquiry is necessary and required to foster inquiry-based learning. Hence, in order to assess the scientific inquiry levels of students, Zachos et al. (2000) proposed critical "*scientific inquiry capabilities*" as assessment measures, whereby a series of structured performance tasks were designed to investigate students' competence in conducting scientific inquiry. Zachos (2004) then proposed that the students' responses, presented with the structured performance tasks, should be recorded and assessed based on scientific inquiry capabilities (Zachos et al., 2000) because the direct observation of performance is not feasible within educational systems and it is time-consuming for both teachers and students (Hanauer et al., 2009, pp. 117-118). A paper-and-pencil, 35-item Scientific Inquiry Literacy Test (ScInqLiT), developed by Wenning (2007), is a diagnostic multiple choice test of knowledge relevant for scientific inquiry based on a defined form of scientific literacy. This test can be used to measure students' scientific inquiry knowledge and it is ideal for the pre- and post-testing measures. However, Wenning (2007) also suggested that ScInqLiT should be regarded as an indicator of students' abilities only because procedural knowledge should be assessed by means of performance tests. Furthermore, based on the concept of avoiding direct assessment of students' scientific inquiry process knowledge, Lunsford and Melear (2004) used the final product of scientific inquiry activity (e.g., portfolios, laboratory practices, and student demonstrations) to assess and infer the learning status and performance concerning scientific inquiry capabilities of students.

To assess scientific inquiry performance, Hanauer et al. (2009, pp. 39-42) defined the characteristics of the Authentic Scientific Inquiry Assessment (ASIA) and thus proposed the active assessment development procedure, which consists of five stages: (1) Empirical Description of Scientific Inquiry; (2) Definition of Educational Aims; (3) Assessment Tool Development; (4) Scoring Rubric Development; and (5) Assessment Piloting. Based on this development procedure, a case study referred to as the Phage Hunters Integrating Research and Education (PHIRE) program, was proposed to address how specific assessment strategies and tools were constructed and implemented (Hanauer et al., 2009, pp. 55-113). The PHIRE program aims to introduce students to the scientific process, and to emphasize the involvement of students who have little scientific training, but are curious about science and the natural world in which we live. Therefore, it was designed as a 10-step program, consisting of the following: (1) Phage Isolation; (2) Phage Purification; (3) Phage Amplification; (4) Electron Microscopy; (5) Nucleic Acid Extraction and Restriction Analysis; (6) DNA Sequencing; (7) Genome Annotation; (8) Comparison of the DNA Sequence to Known Genome; (9) Comparative Genome Analysis; and

(10) Publication. These steps are used to train and assess participating students. The PHIRE assessment strategy covers formative diagnostic and summative aims of an scientific inquiry education pertaining to the bacteriophage subject. The strategy includes five assessment tools to assess and evaluate the performance of students' scientific inquiry skills: (1) the Substantive Knowledge Test; (2) the Physical Checklist; (3) the Visual Literacy Test; (4) the Notebook Assessment Tool; and (5) the Knowledge Presentation Performance Test. Each test consisted of either multiple choice questions, open-ended questions, or observations. However, the practical issues of space, time, and money become significant problems to perform the PHIRE program, although it can offer students individual assessment and diagnostic reports of scientific inquiry (Hanauer et al., 2009, pp. 117-118). Conducting the scientific inquiry assessment by means of inquiry-based learning activities related to definitions of scientific inquiry capabilities appears to be an ideal approach (Zachos et al., 2000; Lunsford & Melear, 2004), but it is not easy to perform and it is time-consuming to manually assess the portfolio (Zachos, 2004; Hanauer et al., 2009, pp. 117-118). In addition, it can also be difficult to evaluate learning problems and performance of higher-order capabilities related to scientific inquiry through the use of traditional paper-and-pencil tests (Wenning, 2007; Bennett et al., 2007, 2010, Jacobs-Sera et al., 2009).

Virtual and Web-Based Interactive Learning Environments

Scientific inquiry, as a set of process skills, which consists of questioning, hypothesis-making, experimenting, recording, analyzing, and concluding, can be regarded as "hands-on" learning (NRC, 1996, NSTA, 2004; Ketelhut et al. 2010). Therefore, students need to experience and practice the scientific inquiry-based activity in the physical laboratory in order to efficiently foster and acquire the skills of scientific inquiry. However, practicing in the physical laboratory is not convenient and it is time-consuming for both teachers and students (Zachos, 2004; Jacobs-Sera et al., 2009, Hanauer et al., 2009, pp. 117-118, Bennett et al., 2007, 2010). A significant amount of research has been dedicated to the development of virtual and Web-based interactive learning systems to support online scientific inquiry learning. A virtual laboratory, called ChemCollective (2010; Yaron et al., 2008, 2009), was developed to allow students to design and carry out their own experiments. Therefore, Yaron et al. (2010) created activities, which enable students to use their chemistry knowledge to practice and resolve problems. According to their results, homework using the virtual laboratory with real-world scenarios contributes significantly to learning. In addition, the virtual laboratory can record all student interactions for the further analysis. Dalgarno et al. (2009) also apply the 3D simulated virtual environment, called the Virtual Chemistry Laboratory, which can be used by distance university chemistry students for familiarization with the laboratory. Teaching students to efficiently learn and acquire scientific inquiry skills is not easy for teachers. Therefore, Ketelhut et al. (2010) proposed a novel pedagogy to infuse inquiry into a standards-based science curriculum by means of a Multi-User Virtual Environment (MUVE), called River City, in order to enhance students' motivation and improve their overall learning performance of scientific inquiry. In this MUVE, students can make observations, pose questions, access information, gather and analyze data, plan investigations, propose answers and explanations, and communicate the results. The experimental results also show that students were able to conduct inquiries in the virtual worlds and were motivated by that process. To improve learning effectiveness, computer simulations, animations, and Web-based interactive content have also been used in many courses and curriculums (Hameed, Hackling, & Garnett, 1993; Windschitl & Andre 1998; Salajan et al., 2009). Hsu et al. (2008) proposed a Technology-Enhanced Learning (TEL) environment to support science learning related to the causes of the seasons, where a Web-based interactive simulation tool was applied to support students' explorations. Students can test and evaluate their hypothesis and learned concepts. Although the aforementioned virtual and Web-based interactive learning environments can enhance students' motivation, foster students' experiences, and improve students' learning performance, the assessment and diagnosis of an individual student still need to be performed and manually analyzed by teachers according to the collected data within a given student's portfolio.

Analysis and Diagnosis of the Learning Portfolio

To analyze learning portfolios, Chen, Liu, Ou, and Liu (2000; Chang, Chen, & Ou, 1998) applied decision tree and data cube techniques to analyze the learning behaviors of students and to discover pedagogical rules related to students' learning performance from Web logs. These logs include the amount of article reading/posting, question-asking, login, etc. According to their proposed approach, teachers can easily observe learning processes and analyze learning behaviors of students for pedagogical needs. However, this approach cannot provide automatic analysis. In order to automatically diagnose learning problems, Hwang (2003) proposed a Concept-Effect Relationship (CER) model to represent prerequisite relationships among concepts of a course, which can be used to evaluate a student's learning status, which may then, in turn, provide that student with the diagnostic report that not only denotes the score, but also the description of any misconceptions. Afterwards, to solve the problem that a concept might contain a hierarchical structure of knowledge with different degrees of difficulty, Chu et al. (2010) defined an Enhanced Concept-Effect Relationship (ECER) to assist teachers in identifying relationships among concepts and their multiple knowledge levels. They then proposed a learning

diagnosis algorithm to analyze a student's learning problems and personalized learning guidance was offered. Based on the concept of the ECER model, a multi-expert approach has also been proposed to integrate the opinions of multiple experts in order to obtain high quality relationships between a test item and concept in the ECER model (Panjaburee et al., 2010). To address the problem of generating automatic feedback for teachers, a Teacher ADViser (TADV) system was developed (Kosba, et al., 2007). TADV defined the knowledge model based on the concept map of a course in relation to the individual student, group, and class, and then a feedback generation algorithm using the fuzzy approach was proposed to analyze tracking data of students. Consequently, learning feedback, including conceptual learning performance and possible learning suggestions, will be automatically generated for both the teacher and student. These analytical and diagnostic approaches (Hwang, 2003; Kosba et al., 2007; Chu et al., 2010; Panjaburee et al., 2010) previously mentioned are able to automatically analyze a student's learning portfolio and generate individualized learning guidance and feedback for both teachers and students; only the diagnosis concerning the conceptual knowledge is taken into account.

Considering the automatic assessment of scientific inquiry skills, Ting et al. (2008) proposed a Dynamic Decision Network (DDN) model in the INQPRO, a scientific inquiry exploratory learning environment for learning Physics (Ting et al., 2006), to assess the mastery of two temporal, variable, scientific inquiry skills of students, i.e., Hypothesis Formulation and Variable Identification. The proposed DDN model can be generated dynamically by integrating various INQPRO Graphical User Interfaces (GUIs) in real-time. In the INQPRO system, students are first required to make a hypothesis statement to elucidate their selected scenarios. Afterwards, students can actively interact with GUIs, and an animated pedagogical agent will give them the tailored suggestions and interventions according to assessment results consisting of three mastery levels (i.e., mastery, partial mastery, non-mastery) of two scientific inquiry skills. However, the tailored interventions only considered the limited suggestions in terms of three mastery levels and incorrect GUI operations of two scientific inquiry skills. The various learning problems, concerning conceptual knowledge, cause and effect operations, and skills of scientific inquiry, with corresponding reasons and remedial suggestions, were not taken into consideration.

Additionally, to measure problem solving with technology, the National Assessment of Educational Progress (NAEP) Technology-Based Assessment Project developed a Technology-Rich Environments (TRE) in the domain of physical science surrounding helium gas balloons (Bennett et al., 2007, 2010). In the TRE search scenario, students needed to use a simulated World Wide Web environment to locate and synthesize information regarding scientific helium balloons. Students were then to answer one constructed response question and four multiple-choice questions related to the uses and science of gas-balloon flight. In the TRE simulation scenario that followed, students could use an interactive simulation tool to experiment with solving problems about relationships among buoyancy, mass, and volume. The TRE employed Evidence-Centered Design (ECD) (Mislevy, Almond, & Lukas, 2003) to develop the interpretive framework consisting of student and evidence models for translating the multiplicity of actions collected from each student into inferences. The student model represented a set of hypotheses about the components of proficiency in a domain and thus defined two primary assessment skills: scientific inquiry and computer skills. The evidence model showed how relevant student actions were connected to those assessment skills; evidence was captured by computer and consisted of student actions called "observables." The TRE used the scoring criteria called "evaluation rules" to assess the accuracy of observables, and used a modeling procedure based on Bayesian networks (Mislevy, Almond, Yan, & Steinberg, 2000) to create the summary scores of skills. Therefore, by means of the TRE assessment process, problem solving capabilities of students can be assessed and scored. Nevertheless, data collected in the assessment portfolio still needs to be manually evaluated by reviewers and detailed diagnoses concerning skill problems must be further developed. The aforementioned research and systems either provide students with limited diagnostic feedback, e.g., conceptual knowledge and summary skill level scores (Ting et al., 2008), or performed the manual assessment (Hanauer et al., 2009; Bennett et al., 2007, 2010). However, analysis of learning problems regarding scientific inquiry capabilities needs to diagnose the operational and procedural portfolio of students. With this need in mind, our main concern is how to propose a novel, online, automatic assessment and diagnosis scheme to efficiently provide students with descriptive diagnostic feedback, corresponding explanations, and remedial suggestions to correct learning problems concerning conceptual knowledge, cause and effect operations, and skills of scientific inquiry.

ONLINE PORTFOLIO ASSESSMENT AND DIAGNOSIS SCHEME

Problem Description

As stated previously, Scientific Inquiry (SI) as a set of process skills, which consists of questioning, hypothesis-making, experimenting, recording, analyzing, and concluding, can be regarded as "hands-on" learning (NRC, 1996, NSTA, 2004; Ketelhut et al. 2010). Although the virtual and Web-based interactive learning systems can be used to enhance learning performance of scientific inquiry (Yaron et al., 2008, 2009,

2010; Hsu et al., 2008; Dalgarno et al., 2009; Ketelhut et al., 2010), to manually assess scientific inquiry competencies according to the students' portfolios collected from the inquiry-based learning activities is still difficult and time-consuming for teachers (Zachos, 2004; Hanauer et al., 2009, pp. 117-118; Bennett et al., 2007, 2010). Besides, the limited diagnostic feedback, e.g., conceptual knowledge and summary scores of skills (Ting et al., 2008) cannot allow students to thoroughly understand their learning problems in terms of scientific inquiry. Therefore, in this paper, to provide students with personalized learning guidance concerning not only conceptual knowledge, but also with higher-order knowledge of scientific inquiry capabilities (Wenning, 2007; Jacobs-Sera et al., 2009), pressing issues remain about how to efficiently analyze students' assessment portfolios and automatically offer them the individual diagnostic reports related to concepts, cause and effect operations, skills of scientific inquiry, and related remedial suggestions. The following three issues must be solved:

- (1) How to model and define useful and meaningful assessment knowledge, which can be defined by teachers or domain experts, to correctly present conceptual and evaluation knowledge for the assessment of an scientific inquiry experiment.
- (2) How to efficiently analyze learning problems according to the assessment portfolio collected by a Web-based scientific inquiry experiment based on the teacher-defined assessment knowledge.
- (3) How to generate a personalized diagnostic report concerning any learning problems related to the concepts, cause and effect operations, and scientific inquiry skills to improve the overall understanding of scientific inquiry.

Framework of the Online Portfolio Assessment and Diagnosis Scheme

According to the issues mentioned in previous sections, an Online Portfolio Assessment and Diagnosis Scheme, called OPASS, has been proposed. The OPASS framework is shown in Figure 1. This scheme employ the rule-based inference approach to efficiently and automatically evaluate students' assessment portfolios of a Web-based scientific inquiry experiment and then diagnose learning problems concerning concepts, cause and effect operations, and scientific inquiry skills according to the teacher-defined assessment knowledge. It can further provide students with personalized scientific inquiry diagnostic reports to improve not only subject concepts, but also scientific inquiry capabilities.

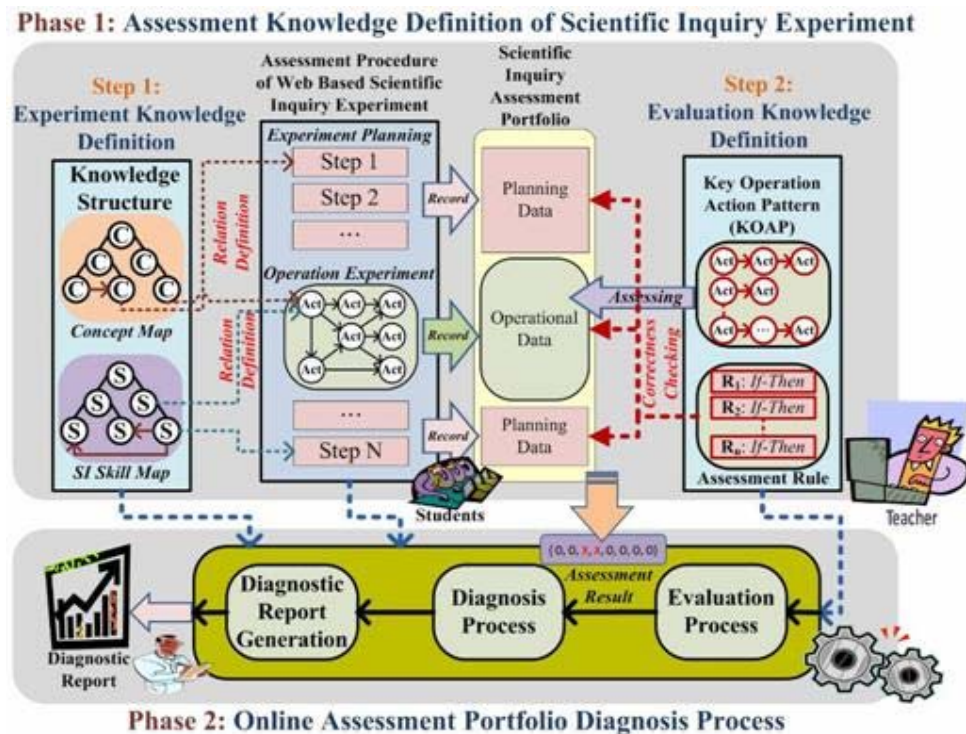


Figure 1: Framework of the OPASS

The OPASS includes two phases described as follows:

- 1. Assessment Knowledge Definition of Scientific Inquiry Experiment Phase:** In order to correctly assess a student's portfolio of the Web-based scientific inquiry experiment, the assessment knowledge consisting of (1) **Experiment Knowledge** and (2) **Evaluation Knowledge** must be defined in advance by the teacher, as shown at

the top of Figure 1. Experiment knowledge, defined as the knowledge structure, includes the concept map of a subject and the skill map of scientific inquiry, and is used to represent required concepts and skills that students need to understand and acquire in the scientific inquiry experiment. Therefore, to assess the students' capabilities of concepts and skills, steps of experiment planning and actions of the operation experiment in the assessment procedure of the Web-based scientific inquiry experiment can thus be associated with experiment knowledge. Moreover, to check the accuracy of the students' assessment portfolios, evaluation knowledge including the Key Operation Action Pattern (KOAP) and the Assessment Rule (AR) must also be defined. The KOAP is proposed to define the key operational actions and sequences, which influence the correctness of operational data in the operation experiment. Hence, based on the KOAP and the experiment knowledge, the assessment rule is proposed to evaluate the accuracy of the students' assessment portfolios and to further identify problems related to scientific inquiry.

2. Online Assessment Portfolio Diagnosis Process Phase: In order to efficiently provide students with personalized scientific inquiry diagnostic reports, this phase, which consists of three modules, has been proposed to automatically evaluate and diagnose learning problems according to students' assessment portfolios, then to generate personalized diagnostic reports. The three modules include the following:

- **Evaluation Process:** uses the teacher-defined Assessment Rule (AR) to evaluate the correctness of the student's assessment portfolio of the Web-based scientific inquiry experiment.
- **Diagnosis Process:** diagnoses learning problems of concepts, cause and effect operations, and skills of scientific inquiry by means of the proposed Diagnostic Rule (DR).
- **Diagnostic Report Generation:** generates the personalized scientific inquiry diagnostic report with descriptions, corresponding reasons, and remedy suggestions of learning problems based on the defined Description Format.

Details of each phase will be described in the following sections.

Assessment Knowledge Definition of Scientific Inquiry Experiment Phase

As mentioned above, in order to automatically assess and diagnose students' scientific inquiry learning problems according to their assessment portfolios of a Web-based scientific inquiry experiment, the assessment knowledge of the scientific inquiry experiment must be predefined by the teacher (Matthews, Pharr, Biswas, & Neelakandan, 2000; Hwang, 2003; Chu et al., 2010; Panjaburee et al. 2010). Therefore, in the OPASS, the assessment knowledge, which consists of experiment knowledge and evaluation knowledge, has been proposed. Figure 2 shows the relation model of assessment knowledge in the OPASS. In the OPASS, each **Experiment Step** of the assessment procedure in the Web-based scientific inquiry experiment and each definition of the **Key Operation Action Pattern (KOAP)** can be associated with concepts and skills of the teacher-defined knowledge structure. Based on this aforementioned relational definition, the teacher-defined **Assessment Rule (AR)**, represented by the **IF (Condition Setting) THEN (Assessment Function)** rule format, is able to evaluate the assessment portfolios and diagnose conceptual problems, cause and effect operations, and scientific inquiry skills, where the **Assessment Function** uses the **Problem** definition to check whether students have a problem or not for the corresponding experiment step. Each assessment function is also associated with the corresponding **Diagnostic Knowledge** including **Problem Description, Reason, and Suggestion**, which will be further used to generate the diagnostic reports by the proposed **Diagnostic Rule (DR)**. Each definition of the Assessment Knowledge (AK) will be described in following subsections.

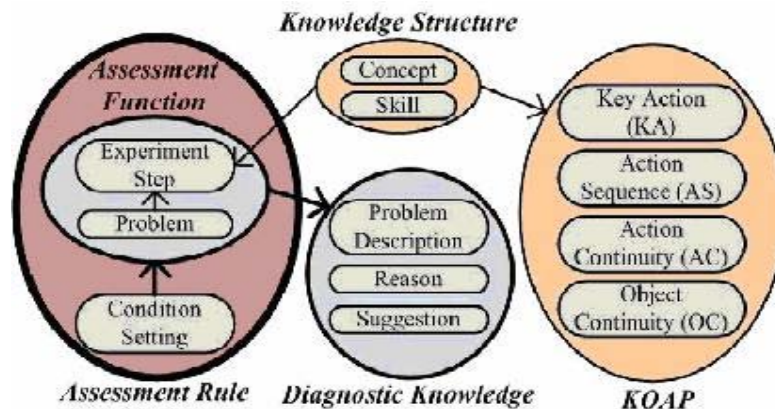


Figure 2: Relation Model of the Assessment Knowledge in the OPASS

Definitions of the Experiment Knowledge

In order to assess students' experimental portfolios, the experiment knowledge related to the scientific inquiry experiment need to be defined in advance. Therefore, in the OPASS, two kinds of knowledge structures have to be defined by the teacher: the concept map of a subject and the skill map of scientific inquiry. The former denotes necessary concepts that students need to learn and understand, and the latter denotes the required skills students need to be equipped with in this assessment experiment. The concept map and the skill map used in the OPASS are defined as follows, respectively.

Definition of the Concept Map (CM):

$CM=(C, R)$, where:

- $C = \{c_1, c_2, \dots, c_n\}$: c_i represents the main concept in a subject
- $R = \{cr_1, cr_2, \dots, cr_m\}$: cr_i represents the Relation Type between two concepts in a CM, where the Relation Type is defined as the **APO**: c_i is **A Part Of** c_j , or the **PR**: c_i is the **Prerequisite** of c_k .

Here, the CM, consisting of a set of concepts (c_i) with two types of relations, i.e., A-Part-Of relations (**APO**) and Prerequisite Relations (**PR**), is a hierarchical structure of concepts of a subject. By means of these relational definitions among concepts, learning problems related to subject concepts can thus be found and diagnosed for a student. Figure 3 depicts an example of a partial CM of a Biology Transpiration Experiment, where the concept *Phenomenon* has three sub-concepts: *Transpiration*, *Photosynthesis*, and *Capillarity*, and prerequisite concepts of *transpiration* are *water transportation* and *Capillarity*.

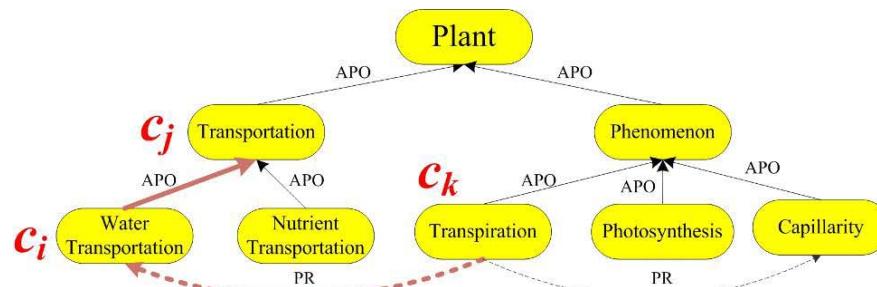


Figure 3: Example of a Partial CM of the Biology Transpiration Experiment

Definition of the Skill Map (SM) for Scientific Inquiry:

$SM=(S, R)$, where:

- $S = \{s_1, s_2, \dots, s_n\}$: s_i represents a Skill of Scientific Inquiry Skills.
- $R = \{sr_1, sr_2, \dots, sr_m\}$: sr_i represents the Relation Type between two skills in a SM, where the Relation Type is defined as the **APO**: s_i is **A Part Of** s_j , or the **D**: s_i is **Dependence on** s_k .

The structure of the SM for scientific inquiry is the same as the CM, except for cross-link relation definitions, Dependence Relations (**D**), which represent cause-and-effect relations between two skills. For example, Figure 4 illustrates an example of a partial SM for the scientific process, where the skill, *Setting Variables*, depends on the skill, *Making Hypothesis*.

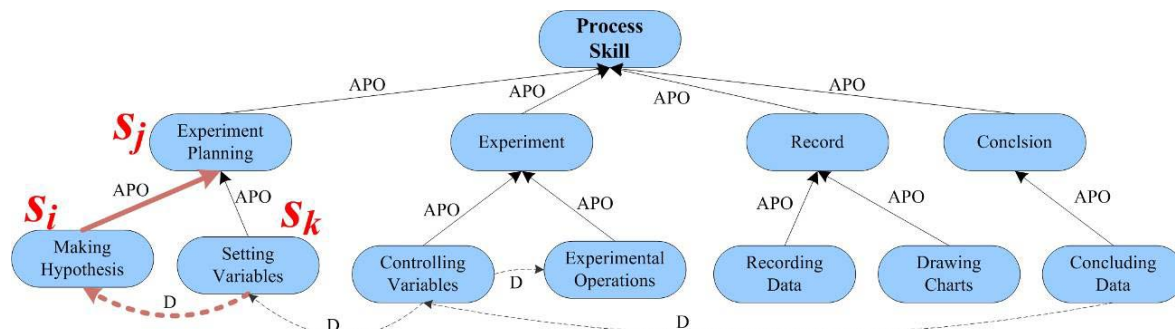


Figure 4: Example of a Partial Scientific Process Skill Map for the Scientific Inquiry Experiment

Definitions of the Evaluation Knowledge

Definitions of the Key Operation Action Patterns:

During the Web-based scientific inquiry experiment, students will be asked to operate the Web-based operation experiment tool, which emulates the actual experiment operation, and their behavior will be collected and regarded as Operational Data of the scientific inquiry assessment portfolio. However, an important problem is how to automatically assess and evaluate operational data of students. Therefore, in the OPASS, the Key Operation Action Patterns (KOAP) has been proposed to evaluate the accuracy of students' operational data. The KOAP defines key operational actions and sequences, which will influence the operational accuracy of the Web-based operation experiment tool. Accordingly, the teacher can define the necessary KOAP to observe and evaluate students' operational data. The definitions related to the Experiment Operations (EO) and KOAP in terms of the Web-based operation experiment are defined as follows:

Definitions of the EO:

$EO = \{a_1, a_2, \dots, a_n\}$: denotes all actions that a student can operate in terms of a Web-based operation experiment tool in the scientific inquiry assessment experiment.



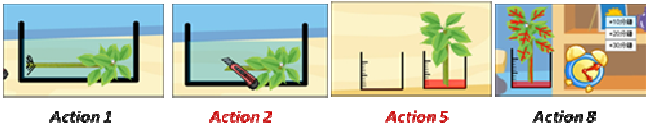
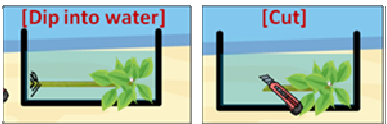
Definitions of the KOAP:

$KOAP = (KA, AC, AS, OC)$, where:

- $KA = \{a_i, a_j, \dots, a_m \mid 0 \leq m \leq n\}$: denotes the **Key Action (KA)**, each action (a_i) of which plays an **important action** of all operational actions in EO , whose accuracy will influence the accuracy of the whole EO.
- $AC = (a_i, a_{i+1}, a_{i+2}, \dots)$: denotes the **Action Continuity (AC)**, which is an action sequence with continuous actions.
- $AS = (a_i, a_{i+j}, \dots, a_{i+k} \mid i < j < k)$: denotes the **Action Sequence (AS)**, which is an action sequence, but its continuity is not necessary.
- $OC = (a_i, a_{i+1}, a_{i+2}, \dots)$: denotes the **Object Continuity (OC)**, which is a continuous action sequence for a targeted object.

Therefore, according to the definition of the KOAP, the accuracy of a student's operational portfolio of a Web-based operation experiment tool can thus be automatically assessed, analyzed, and diagnosed. Table 1 illustrates examples of the KOAP with descriptions.

Table 1: Illustration with the Description of each KOAP

Type	Illustration	Description
Key Action (KA)		[Filling] the [Red Water] in a [cup without scale] into the [Beaker with Scale] is a Key Action (KA).
Action Continuity (AC)		In order to sniff out the fire correctly, [Action 1] must be followed by [Action 2] and it's not allowable to operate other actions between them.
Action Sequence (AS)		$AS = (a_1, a_2, a_5, a_8)$ is a correct operational action sequence to finish the operation experiment, where [Action 2] must be done before [Action 5], but other actions can be operated between Action 2 and Action 5.
Object Continuity (OC)		For the targeted object, Celery, it must be [Cut] only after [Dip into water]. It will be regarded as the incorrect operation if there are other actions between them.

Definitions of the Assessment Rules (AR):

In the OPASS, the rule-based inference approach has been applied to infer the accuracy of the assessment experiment according to a student's assessment portfolio. Therefore, the teacher can define assessment rules in advance to evaluate the accuracy of a student's answer and to identify learning problems related to subject

concepts, cause and effect operations, and skills of scientific inquiry. The assessment rule can be defined by the following definition.

Definitions of the AR:

$AR = \{Ar_1, Ar_2, \dots, Ar_n\}$, where:

- **$Ar_i = \text{If (Condition Setting) Then (Assessment Function)}$** : each Ar_i of AR can be represented by the IF-THEN rule format, where:
 - **Condition Setting** $= \{Cs_1, Cs_2, \dots, Cs_m\}$: each Cs_i of the *Condition Setting* can be used to evaluate the accuracy of the student's answer in terms of the assessment portfolio consisting of planning data and operational data defined in Section: Definitions of Assessment Portfolio. If the result of the **Condition Setting** is true, the **Assessment Function** will be triggered to evaluate the student's assessment portfolio.

In the OPASS, the Predicate Function (Giarratano & Riley, 2004) has been applied to be the function used in the AR. A predicate function is defined to be any function that returns TRUE or FALSE. Therefore, any value other than FALSE is considered as TRUE. The predicate function always returns a Boolean value. The **Assessment Function** used in the AR is defined as follows.

Definitions of the Assessment Function in AR:

WrongStep(Step_i, Problem_i): checks the experiment Step_i of the assessment procedure, which was executed correctly or not during the Web-based scientific inquiry experiment, where:

- **Step_i**: the name of an experiment step in the scientific inquiry assessment experiment.
- **Problem_i**: denotes a checking predicate function, which can check whether a student made this kind of problem at an executed experiment Step_i. Therefore, each Problem_i has its corresponding checking predicate function definition, which can be extended and defined by the teacher according to requirements of the assessment, such as:
 - **ObjectContinuity_Error(obj_k, ActionSequence_m, WrongPattern_n)**: checks the accuracy of the continuity of the object (obj_k) defined in the KOAP according to the comparison between the correct Object Continuity (OC) (ActionSequence_m) and the student-made action pattern, which will be regarded as WrongPattern_n if it is not the correct experimental operation.
 - **IndependentVariable_Error(obj_k, IF-Statement_n, Then-Statement_n)**: checks the accuracy of the independent variable of the object (obj_k) according to the hypothesis setting (IF-Statement_n and Then-Statement_n), defined in Section: Definitions of Assessment Portfolio, that the student made.

Example 1:

If a student dipped a stalk of celery into water and then used a knife to cut its root during the virtual operation experiment, the accuracy of this experimental operation the student made can thus be checked by defined Assessment Functions, **WrongStep**("Action Operation", *ObjectContinuity_Error*([celery], [dip in water] [cut root] [put into tank] [waiting], [dip in water] [cut root])). Therefore, students' operational actions, i.e., [dip in water] [cut root], are not correct because the correct object continuity definition (OC) of Key Operation Action Patterns (KOAP) was defined as [dip in water] [cut root] [put into tank] [waiting]. Moreover, the accuracy of the hypothesis setting can also be checked by the **WrongStep**("Operational Experiment", *IndependentVariable_Error*([celery], [cross section area of celery stem], [the decreasing quantity of the red water])).

Condition Setting Function in AR:

In addition to the assessment function, the condition setting of the AR can also use the predicate function to check the condition of a rule. Therefore, in the OPASS, the **Condition Setting** $= \{Cs_1, Cs_2, \dots, Cs_m\}$, where, for instance, the $Cs_i = \text{NotMatch}(\text{ObjectContinuity}(\text{targeted object}, \text{correct OC definition})$: evaluates the accuracy between the correct OC definition and the student's operational actions in terms of the targeted object, or $Cs_j = (\text{TargetObject}(\text{obj}) \ \& \ \text{IndependentVariable}(X) \ \& \ \text{CorrectIndependentVariable}(Y) \ \& \ (X \neq Y))$: evaluates the accuracy between the correct independent variable (Y) and the actual one that the student set (X) in terms of the targeted object (obj) and the condition will be true if the $(X \neq Y)$ is true.

Example 2:

Assume there are $Ar_1 = \text{If} (\text{NotMatch}(\text{ObjectContinuity}([\text{celery}], \{ [\text{dip in water}], [\text{cut root}], [\text{put into tank}] [\text{waiting}]])) \ \text{Then} \ \text{WrongStep}(\text{"Action Operation"}, \text{ObjectContinuity_Error}([\text{celery}], \{[\text{dip in water}], [\text{cut root}], [\text{put into tank}], [\text{waiting}]\}, \{[\text{dip in water}], [\text{cut root}]\}))$, and $Ar_2 = \text{If} (\text{TargetObject}([\text{celery}]) \ \& \ \text{IndependentVariable}([\text{length of stem}]) \ \& \ \text{CorrectIndependentVariable}([\text{amount of leaves}]) \ \& \ ([\text{length of stem}] \neq [\text{amount of leaves}])) \ \text{Then} \ \text{WrongStep}(\text{"Operational Experiment"}, \text{IndependentVariable_Error}([\text{celery}],$

[cross section area of celery stem], [the decreasing quantity of the red water]). Therefore, the Assessment Function, *WrongStep()*, will be triggered if the *Condition Setting* of the Ar_1 or the Ar_2 is true.

Definitions of the Assessment Portfolio

As seen in Figure 1, the assessment portfolio of scientific inquiry consists of planning data and operational data. Before the assessment process, the log of the Web-based experiment system must be transformed into the defined format in the OPASS. Logs of planning data, as shown in Table 2, are the set of attribute-value pairs. For example, in an experiment of biology transpiration, students defined a hypothesis: *If the [celery]'s [leaves] are [more], the [decreasing quantity] of the [red water] is [more]*. Then, logs recorded six attributes, including objects, attributes, and their changes in the condition and effect parts of the hypothesis.

Table 2: Example Logs of Planning Data

Attribute	Value	Attribute	Value
Hypothesis-IF-Object	Celery	Hypothesis-THEN-Object	Red water
Hypothesis-IF-Attribute	Leaves	Hypothesis-THEN-Attribute	Decreasing quantity
Hypothesis-IF-Value	More	Hypothesis-THEN-Value	More

Logs of operational data, as shown in Table 3, were a sequence of operations, which consists of an action name, a used object, an object of target, and a set of environmental attribute-value pairs. For example, the action sequence in Table 3 described that a student [fill] a [beaker with scale] with [red water]. Then, the student [dip] a [head of celery] into a [tank] and use a [knife] to [cut] the [stem of the celery]. Afterward, this student [put] the [celery] into the [beaker with scale] and [waited].

Table 3: Example Logs of Operational Data

Action	Used Object	Target Object	Environmental Status
Fill	Red water	Beaker with scale	Temperature: 25 °C, Light: Yes, Humidity: 60%
Dip	Celery	Tank	Temperature: 25 °C, Light: Yes, Humidity: 60%
Cut	Knife	Celery	Temperature: 25 °C, Light: Yes, Humidity: 60%
Put	Celery	Beaker with scale	Temperature: 25 °C, Light: Yes, Humidity: 60%
Wait			Temperature: 25 °C, Light: Yes, Humidity: 60%

Online Assessment Portfolio Diagnosis Process Phase

By means the teacher-defined assessment knowledge related to the scientific inquiry experiment described in the previous section, the student's assessment portfolio can thus be automatically evaluated and diagnosed by the Online Assessment Portfolio Diagnosis Process (OAPDP) in phase 2 of the OPASS. The details will be described in this section.

Procedure of the Online Assessment Portfolio Diagnosis Process

Figure 5 shows the flowchart of the OAPDP, which consists of three modules: (1) Evaluation Process; (2) Diagnosis Process; and (3) Diagnostic Report Generation. In the Evaluation Process, the OAPDP uses the teacher-defined Assessment Rule (AR) to evaluate the accuracy of the students' scientific inquiry assessment portfolio and then finds the **Wrong Experiment Step** from the assessment result according to the inference results of the Rule Inference Process. Afterwards, in the Diagnosis Process, the OAPDP first diagnoses the mis-concept/skill with the corresponding reason for each wrong experiment step by means of the Diagnosis Rule (DR) based on the relation model of assessment knowledge as seen in Figure 2. The OAPDP further analyzes the Remedial Path according to relational definitions of the experiment knowledge, i.e., the prerequisite (*PR*) in the CM and the Dependence (*D*) in the SM of scientific inquiry. Consequently, the Major mis-concept/skill with the corresponding wrong experiment step can be discovered. Finally, the Diagnostic Report Generation module is able to generate the personalized scientific inquiry diagnostic report consisting of descriptions, corresponding reasons, and related remedial suggestions to correct learning problems based on the defined Description Format.

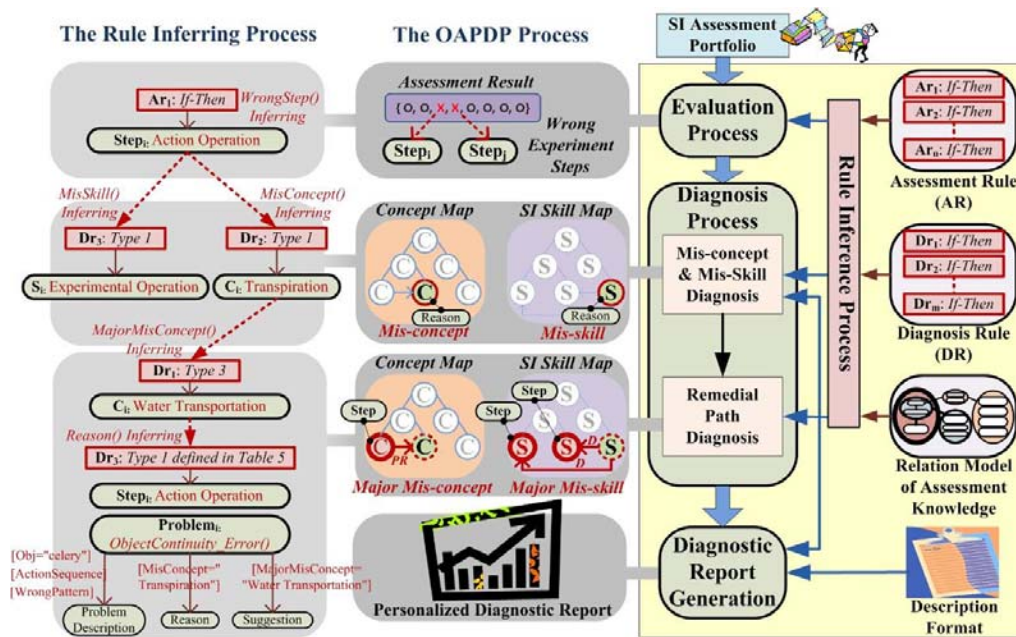


Figure 5: Flowchart of the OAPDP

Diagnosis Process in the OAPDP

As mentioned above, the Diagnosis Process module in the OAPDP uses the Diagnosis Rule (DR) based on the relation model of assessment knowledge to diagnose the mis-concept/skill with the corresponding reason for each wrong experiment step. In the OPASS, the DR has thus been proposed and defined as follows.

Definitions of the DR:

$DR = \{Dr_1, Dr_2, \dots, Dr_n\}$, where:

- $Dr_i = \text{IF (Condition Setting) Then (Diagnostic Function)}$: each Dr_i of the DR can be represented by the IF-THEN rule format, where three types of DRs are defined as follows:
 - (1) **DRs of the Mis-Concept, Mis-Skill, and Reason:**
 - **IF (WrongStep(\$S, \$P) & StepConceptRelation(WrongStep(\$S, \$P), \$Concept)) Then MisConcept(\$Concept)**: diagnoses the mis-concept (*MisConcept()*) according to the relationship between the wrong experiment step (*WrongStep()*) and the associated concept by the function *StepConceptRelation()*. The \$S and \$P denote the $Step_i$ and the $Problem_i$ of Assessment Function, *WrongStep()*, in AR.
 - **IF (WrongStep(\$S, \$P) & StepSkillRelation (WrongStep (\$S, \$P), \$Skill)) Then MisSkill(\$Skill)**: diagnoses the mis-skill according to the relationship between the wrong experiment step and associated skill of scientific inquiry by the function *StepSkillRelation()*.
 - **IF (WrongStep(\$S, \$P) & StepReasonRelation(WrongStep(\$S, \$P), \$Type, \$Desc)) Then Reason(\$Type, \$Desc)**: diagnoses the corresponding reason of occurred mis-concept or mis-skill according to the relationship between the wrong experiment step and associated reason, where **Type** is “Concept” or “Skill,” each of which has a corresponding description (**\$Desc**) to explain the reason for a problem that a student made for the wrong experiment step.
 - (2) **DRs of the Major Wrong Step of Assessment Experiment:**
 - **IF (MajorMisSkill(\$Skill) & WrongStep(\$S,\$P) & StepSkillRelation(WrongStep(\$S, \$P), \$Skill)) Then MajorWrongStep(\$S, \$P)**: diagnoses the major wrong experiment steps of a student according to the relationship between the wrong experiment and the major mis-skill.
 - (3) **DRs of the Remedial Concept and Skill of Mis-Concept and Mis-Skill:**
 - **IF (MajorMisConcept(\$Cx) & Prerequisite(\$Cy, \$Cx)) Then PRConcept(\$Cy)**: diagnoses the remedial concept of the student’s mis-concept according to the prerequisite concept relationship (*Prerequisite()*) of the major mis-concept.
 - **IF (MajorMisSkill(\$Sx) & Prerequisite(\$Sy, \$Sx)) Then PRSkill(\$Sy)** : diagnoses the remedial skill of the student’s mis-skill according to the prerequisite skill relationship (*Prerequisite()*) of the major mis-skill.

Table 4 lists examples of the DR Definition and Table 5 also presents examples of the Assessment Function Definition, *WrongStep(\$S, \$P)*, associated with the Problem Description, the Reason, and the Suggestion

Description. The learning problems related to the concepts, cause and effect operations, and skills of scientific inquiry can thus be analyzed and diagnosed by means of the proposed DR.

Table 4: Example of Three Types in the DR Definition

Type	IF (Condition Setting)		THEN
Symbol Definitions	<ul style="list-style-type: none"> ● \$S1="Operational Experiment" ● \$S2=" Action Operation " ● \$P1=<i>IndependentVariable_Error</i>([celery], [cross section area of celery stem], [the decreasing quantity of the red water])) ● \$P2= <i>ObjectContinuity_Error</i>([celery], {[dip in water], [cut root], [put into tank], [waiting]}, {[dip in water], [cut root]}) 		
Type 1	Dr ₁	$WrongStep(\$S1, \$P1)$ & $StepConceptRelation(WrongStep(\$S1, \$P1), "Transpiration")$	MisConcept("Transpiration")
	Dr ₂	$WrongStep(\$S2, \$P2)$ & $StepConceptRelation(WrongStep(\$S2, \$P2), "Transpiration")$	
	Dr ₃	$WrongStep(\$S2, \$P2)$ & $StepSkillRelation (WrongStep(\$S2, \$P2), "Transpiration")$	MisSkill("Experimental Operation ")
Type 2	Dr ₁	$MajorMisSkill("Experiment Planning")$ & $WrongStep(\$S1, \$P1)$ & $StepSkillRelation(WrongStep(\$S1, \$P1), "Experiment Planning")$	MajorWrongStep(\$S1,\$P1)
	Dr ₂	$MajorMisSkill("Experimental Operation")$ & $WrongStep(\$S2, \$P2)$ & $StepSkillRelation(WrongStep(\$S2, \$P2), "Experimental Operation")$	MajorWrongStep(\$S2, \$P2)
Type 3	Dr ₁	$MajorMisConcept("Transpiration")$ & $Prerequisite("Water Transportation", "Transpiration")$	PRConcept("Water Transportation")
	Dr ₂	$MajorMisSkill("Setting Variable")$ & $Prerequisite("Making Hypothesis", "Setting Variable")$	PRSkill("Making Hypothesis")

Table 5: Example of WrongStep(\$S, \$P) Definition Associated with Problem Description, Reason, and Suggestion Description in the OPASS

DR	Step (\$S)	Problem (\$P)	Problem Description (A), Reason (B), Suggestion Description (C)	
Dr ₁	Making Hypothesis	<i>Hypothesi_Error</i> (scene, IF-Statement, Then-Statement)	A	Because the solution that you made in the [scene] is that "IF [IF-Statement] THEN [Then-Statement]", it can not solve the problem of the experiment.
			C	Please carefully read the "Problem Description" of [scene] again and try to use another approach to solve it.
Dr ₂	Operational Experiment	<i>VariableOperation_Error</i> (Variable)	A	The [Variable] you operate is not the same variable you set in the Setting Variable Step of the experiment.
			B	Reason ("Skill", "the variable that you set in the Setting Variable Step of the experiment can not be operated in this experiment")
			C	You must operate the same variable in the Setting Variable Step and the Operational Experiment Step both.
Dr ₃	Action Operation	<i>ObjectContinuity_Error</i> (Obj, ActionSequence, WrongPattern))	A	Because the [Obj] must be operated by [ActionSequence], we guess that your operation order [WrongPattern] is wrong.
			B	Reason ("Concept", "you may not thoroughly understand the [MisConcept] ")
			C	We suggest that you should learn the [MajorMisConcept] and [MisConcept] in advance.

Example 3:

The left-hand side of Figure 5 illustrates the rule inferring process during the OAPDP process by employing the rule-based inference approach. To follow the descriptions in Examples 1 and 2, if the Ar_1 in AR come to be true, a Wrong Experiment Step, "Action Operation," can be found from the assessment portfolio of scientific inquiry in the Evaluation Process. Therefore, In the Diagnosis Process, after the Mis-concept and Mis-Skill Diagnosis, the mis-concept, "Transpiration," and the mis-skill, "Experimental Operation," at this "Action Operation" step can be inferred by using the Dr_2 of Type 1 and the Dr_3 of Type 1 in DR in Table 4, respectively. Afterwards, in the Remedial Path Diagnosis, the major mis-concept, "Water Transportation," can be found through the Dr_1 of Type 3, and according to the inferred mis-concept and the definition of the concept map in Figure 3. Finally, by using the aforementioned results, the Dr_3 in Table 5 was triggered to reason and diagnose the learning problems with Problem Description, Reason, and Suggestion Description for this wrong experiment step, "Action Operation.". Consequently, the personalized diagnostic results can be offered to the student as follows: You had the wrong experiment step at [Action Operation Step], (A) because the [Obj="celery"] must be operated by [ActionSequence="[celery], {[dip in water], [cut root], [put into tank], [waiting]}"], we guess that your operation order [WrongPattern="[dip in water], [cut root]"] is wrong. (B) The Reason is taht "you may not thoroughly understand the [MisConcept="Transpiration"] "). (C) We suggest that you should learn the [MajorMisConcept="Water Transportation"] and ["MisConcept="Transpiration"] in advance. Consequently, the various learning problems, concerning conceptual knowledge, cause and effect operations, and skills of scientific inquiry, with corresponding reasons and remedial suggestions can be automatically analyzed and diagnosed by the Diagnosis Process in the OAPDP. These diagnostic results will be further organized and syntheized into a readable and understandable resport in the Diagnostic Report Generation in the OAPDP.

Diagnostic Report Generation

After the Evaluation and Diagnosis process modules have been processed, the students' learning problems in relation to the concepts, cause and effect operations, and skills of the scientific inquiry experiment can be diagnosed, and corresponding reasons and descriptions can also be acquired. The personalized diagnostic report can thus be generated by running the Diagnostic Report Generation in the OAPDP. The proposed **Diagnostic Report Generation Algorithm (DRGalgo)** is described in Algorithm 1, and Figure 6 shows an example of the personalized diagnostic report generated by the DRGalgo.

Algorithm 1: Diagnostic Report Generation Algorithm (DRGalgo)**Symbol Definition:**

- **WrongStep_i**: the detected wrong experiment step of SI experiment for the student.
- **MisConcept**: the detected mis-concept of the student.
- **MisSkill**: the detected mis-skill of the student.
- **MajorWrongStep**: the detected major wrong step of SI experiment for the student.
- **MajorMisConcept**: the detected major mis-concept of the student.
- **MajorMisSkill**: the detected major mis-skill of the student.
- **PRConcept**: the prerequisite concept of a concept.
- **\$**: output the value of variable

Input: All detected wrong experiment steps of SI assessment experiment

Output: Personalized Diagnostic Report

Step 1: Generate the detailed description for each Wrong Step (**WrongStep_i**) of Assessment Experiment,

1.1: output the statement: "[Problem]: you made wrong action at [**WrongStep_i**] Step."

1.2: output the statement: "[Corresponding Skill]: [**MisSkill**]."

1.3: output the statement: "[Phenomenon]: [**(the Problem Description of WrongStep_i)**]"

1.4: **If** Reason.Type = "Concept"

Then output the statement: "[Possible Reason]: you may not thoroughly understand the [**MisConcept**]."

Else If Reason.Type = "Skill"

Then output the statement: "[Possible Reason]: because [**(the Reason Description of WrongStep_i)**]
for the [**MisSkill**]"

1.5: output the statement: "[Suggestion]: [**(the Suggestion Description of WrongStep_i)**]"

Step 2: Generate the overall diagnostic description for student's assessment result

2.1: **If** [conclusion is wrong]

Then

 (1) output the statement: "[Problem]: your conclusion is wrong. The possible reason may be the [**(the Problem Description of the MajorWrongStep)**]."

- (2) output the statement: "[Skill Suggestion]: [*\$(the Suggestion Description of MajorWrongStep)*]."
 - (3) output the statement: "[Concept Suggestion]: for the concept of subject in this experiment, suggest that you need to thoroughly learn and understand the concept of [*\$MajorMisConcept*]."
 - (4) output the statement: "[Prerequisite Concept Suggestion]: other than the concept [*\$MajorMisConcept*], suggest that you can also thoroughly learn and understand its prerequisite concept [*\$PRConcept*]."
- Step 3:** Output the Personalized Diagnostic Report

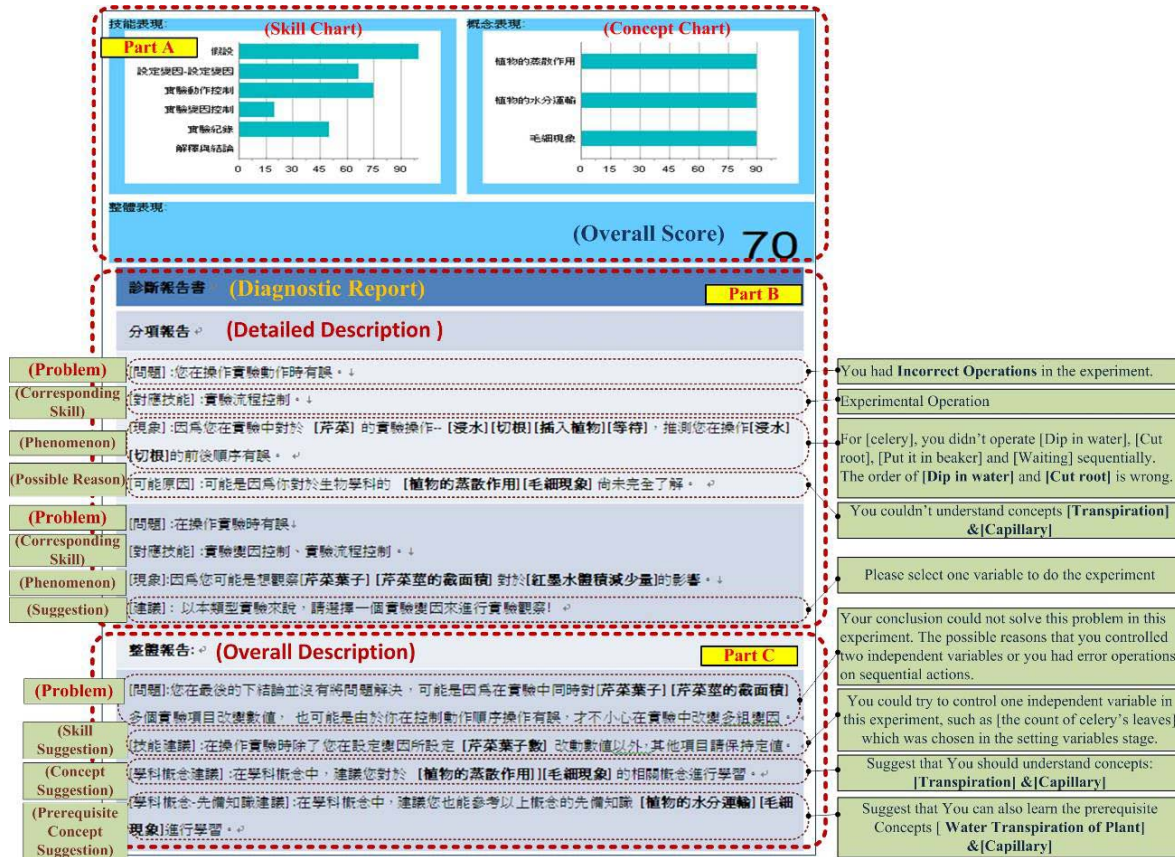


Figure 6: Example of the Personalized Diagnostic Report Generated by DRGalgo in OPASS

Implementation and Experiment Prototypical System of the OPASS

In order to evaluate the effectiveness of the OPASS, the prototypical system has been developed, as shown in Fig 7. The OPASS system consists of three databases: (1) Assessment Knowledge Base; (2) Diagnosis Rule Base; and (3) Assessment Portfolio Database. The assessment knowledge can be defined by teachers to meet the requirements of scientific inquiry assessments based on the proposed Assessment Knowledge (AK) definition. The OPASS can be integrated with the Web-based scientific inquiry experiment system based on the proposed connection protocol. Therefore, students can use the browser to take the scientific inquiry assessment and their operational behavior will be recorded into the assessment portfolio database. After students finish the assessment, the OAPDP will automatically analyze the assessment portfolio using the rule inference process according to assessment knowledge and then automatically generate personalized diagnostic reports to students according to diagnostic rules.

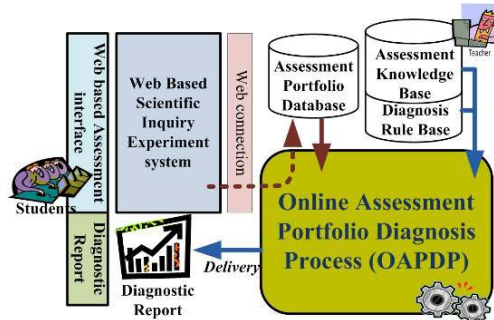


Figure 7: Architecture of the Prototypical OPASS

As seen in Figure 8, six assessment activities executed on the Web-based scientific inquiry experiment system have also been developed for the Physics (Figure 8b) and Biology (Figure 8c) experiments, respectively. In Figure 8a, each assessment was developed based on the assessment procedure consisting of six steps, where the operation experiment in step 3 offers a Web-based interactive, operational experiment tool to allow students to operate it and observe responses and reactions.

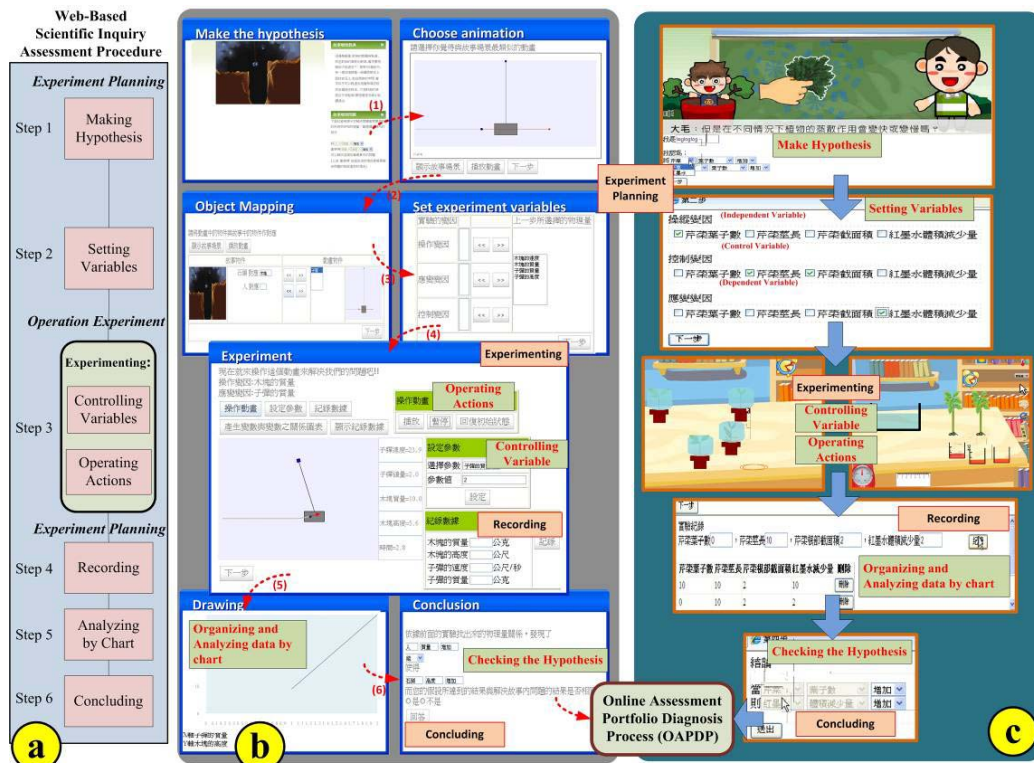


Figure 8: Assessment Activities of the Web-based Scientific Inquiry Experiment in: (b) Physics and (c) Biology Based on (a) Assessment Procedure

EXPERIMENTAL RESULTS

Experiment Plan and Execution:

In order to evaluate the performance of the prototypical OPASS system, several experiments were conducted. Two classes, from different schools in Taiwan, participated in the assessment experiments. Thirty first-grade students of high school, in the urban district, and ten third-grade students of junior high school, in the remote district, participated in the assessment experiments of scientific inquiry in Biology and Physics, respectively. First, teachers explained the purpose of the experiment and taught students how to use the Web-based scientific inquiry experiment system (OPASS). Students could practice and familiarize themselves with the system by participating in the testing experiment (Figure 9a). Following the practice test, the students took the formal assessment experiments (Figure 9b) to understand their learning problems by means of personalized diagnostic reports (Figure 9c). Finally, a questionnaire of a five-level Likert Scale, as seen in Table 6, was designed and provided to students to evaluate their degrees of satisfaction concerning the OPASS system.

Table 6: Questionnaire of Students' Degrees of Satisfaction of the OPASS System (Five-Level Likert Scale from 1 (Strongly Disagree) to 5 (Strongly Agree))

Q1: It would be helpful to provide personalized analysis and learning suggestions concerning the operation and examination after the assessment experiment.
Q2: In Part A of the diagnosis report, the bar charts of skills, concepts, and overall scores can assist you in understanding your assessment outcome.
Q3: In Part B of the diagnosis report, the descriptions consisting of the wrong plans, wrong operations, reasons, and possible remedial suggestions can assist you in understanding the problems during the experiment.
Q4: In Part C of the diagnosis reports, the descriptions concerning the overall diagnosis and suggestions can improve your learning.
Q5: This diagnosis report is useful and can improve your learning efficacy.



Figure 9: (a) Students Practicing the OPASS, (b) Taking the Examination, and (c) Reading the Diagnostic Report Regarding the Scientific Inquiry Experiment in the Physics Domain

RESULT ANALYSIS

Analysis of Student's Scores in the OPASS System

● **Correlations of OPASS Scores with Prior Knowledge Measures**

Examining the correlations of the OPASS scores with each measure of prior knowledge can help clarify meaning. For example, students with more prior knowledge tended to perform better on each score of the OPASS than students with lower levels of prior knowledge. The prior knowledge measures were intended to give an indication of the degree of student familiarity with the science and related concepts being assessed in the scientific inquiry experiments of the OPASS system (Bennett et al., 2007, 2010). In this paper, the prior knowledge measures consist of two kinds of knowledge: (1) Science Knowledge; and (2) Scientific Inquiry. The prior science knowledge measure was designed to be related to the Physics and Biology domain. Therefore, the grade of a student of Physics and Biology at school was adopted as the prior science knowledge measure.

The prior scientific inquiry knowledge measure was intended to concern skills of scientific inquiry. In order to assess prior scientific inquiry knowledge of participant students, a comprehensive Test of Integrated Science Process Skill (TIPS) was developed by Dillashaw and Okey (1980). This test included integrated science process skills (e.g., stating hypotheses, controlling variables, designing experiments, operational definition, graphing and interpreting data) and was adopted as a reference to design a Chinese version. The TIPS had a high reliability (0.89) and was non-curriculum-specific for the middle and secondary schools. Afterwards, Burn, Okey, and Wise. (1985) developed the TIPS II based on the original TIPS.

By means of the data collected from the experiments of the OPASS system, Table 7 lists the summary statistics of the Prior Science Knowledge and the OPASS Measures for the 30 first-grade high school students (Grade 10) in the Physics domain (effective sample size (N) = 24). Table 8 presents the correlations of the "Total Score" of the OPASS, consisting of "Scientific Inquiry" and "Science Knowledge", with the two prior knowledge measures, "Science Knowledge" and "Scientific Inquiry knowledge".

Table 7: Summary Statistics for Prior Knowledge and OPASS Measures – Grade 10, Physics domain

Measures Statistic	Prior Science Knowledge		OPASS		
	Science Knowledge: Grade in Physics	Scientific Inquiry: Total Score of TIPS	Total Score	Scientific Inquiry	Science Knowledge
Number of Students (N)	24	24	24	24	24
Mean Score	71.88	72.64	75.83	79.17	72.50
Standard Deviation (SD)	9.205	7.982	9.289	8.456	12.324

Table 8: Correlations of OPASS Scores with Prior Knowledge Measures in TIPS – Grade 10, Physics Domain

OPASS Score	Prior Science Knowledge: Grade in Physics	Prior Scientific Inquiry Knowledge: Total Score of TIPS
Total	-.263	.431*
Scientific Inquiry	-.156	.492*
Science Knowledge	-.290	.313

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

According to the correlations in Table 8, the “Total” score of OPASS did not correlate with the two prior knowledge measures: “Science Knowledge: Grade in Physics” and “Prior Scientific Inquiry Knowledge: Total Score of TIPS.” In addition, the “Prior Science Knowledge” did not correlate with the “Prior Scientific Inquiry Knowledge.” This indicates that the mastery levels of students’ grades in Physics may not influence the performance of OPASS and TIPS.

Besides, the “Total” score of TIPS has the significant positive correlations with the “Total” score (0.431, $p < .05$) and “Scientific Inquiry” (0.492, $p < .05$) of OPASS, respectively. This means that students with more prior scientific inquiry knowledge tend to perform better on “Total” and “Scientific Inquiry” scores of the OPASS. Furthermore, the “Scientific Inquiry” of OPASS has a significant positive correlation (0.584, $p < .01$) with the “Science Knowledge” of OPASS. The reason for this outcome is that the OPASS system integrated the scientific inquiry skills and science knowledge together with each step and action of the Web-based assessment procedure.

According to the results of Table 8, the “Total” score of the OPASS has a significant correlation with the TIPS score. In this paper, the “Scientific Inquiry” score of OPASS consists of five scales: (1) Making Hypothesis; (2) Setting Variables; (3) Experimenting; (4) Graphing; and (5) Concluding. For estimating the correlations, the TIPS scales were mapped to these five OPASS scales. Therefore, the correlations of each sub-score of OPASS with TIPS are shown in Table 9 to investigate the reliability and validity of the OPASS system.

Table 9: Correlations of OPASS Scores with Prior Knowledge Measures in TIPS – Grade 10, Physics Domain

OPASS Score	TIPS Score	Making Hypothesis	Setting Variables	Experimenting	Graphing	Concluding
Total		.031	.506*	.271	.235	.203
Scientific Inquiry		.059	.352	.237	.149	.210
Science Knowledge		.005	.509*	.240	.245	.158
(1) Making Hypothesis		.a	.a	.a	.a	.a
(2) Setting Variables		.025	.593**	.147	-.062	.166
(3) Experimenting		-.145	.254	.120	-.073	-.054
(4) Graphing		.199	.147	.303	.646**	.351
(5) Concluding		-.038	-.151	-.074	-.320	-.094

a. Cannot be computed because at least one of the variables is constant.

As Table 9 shows, the correlation values of “Making Hypothesis” (OPASS) with TIPS cannot be computed because all students correctly performed this step in OPASS. The “Concluding” (OPASS) also did not correlate with the one of TIPS because 19 out of 24 students were correct. The reason for this is that students learned concepts and skills related to “Making Hypothesis” and “Concluding” in the practice section, and such learning effects subsequently became prior knowledge when the students took the online assessment of scientific inquiry in the examination section, as depicted in Figure 9.

The “Experimenting” portion (OPASS) has no significant positive correlation with the one of TIPS. That is

because students were required to interact and operate the Web-based operation experiments at the “Experimenting” step in the OPASS system, which can be regarded as a “hands-on” assessment. The operational data of students, as shown in Table 3, were recorded and collected in the assessment portfolio and assessed according to the teacher-defined assessment knowledge definition, e.g., Key Operation Action Pattern (KOAP). On the contrary, the TIPS is a paper-and pencil test and is a suitable approach to measure students' knowledge of scientific concepts and inquiry (e.g., Substantive Knowledge, but it is not easy to assess and evaluate learning problems and performance of higher-order capabilities related to scientific inquiry).

Furthermore, the “Setting Variables” and “Graphing” in the OPASS system have significant positive correlations (0.593 and 0.646, $p < 0.01$) with the ones of TIPS, respectively. Those correlations describe that students with more prior knowledge in terms of “Setting Variables” and “Graphing” in TIPS tend to perform better on corresponding scales in the OPASS system than students with lower levels. Consequently, the significant correlations between the OPASS and the TIPS can show that the OPASS system is able to perform a reliable and valid assessment of scientific inquiry.

In addition to the evaluation for grade 9 students in the Biology domain at the urban district, the prototypical OPASS system was evaluated by 10 grade 9 students who reside in the remote district, as listed in Table 10 and 11, respectively. The results show that the performance of the OPASS has no significant correlations with the “Prior Knowledge” of students in terms of “Average of Subjects” and “Grade in Biology,” which is the same as the experiment results in Physics.

Table 10: Summary Statistics for Prior Knowledge and OPASS Measures – Grade 9, Biology Domain

Measures Statistic	Prior Knowledge		OPASS		
	Knowledge: Average of Subjects	Science Knowledge: Grade in Biology	Total Score	Scientific Inquiry	Science Knowledge
Number of Students (N)	10	10	10	10	10
Mean Score	75.27	78.83	56.83	62.00	51.67
Standard Deviation (SD)	9.536	9.425	18.316	16.633	23.107

Table 11: Correlations of OPASS Scores with Prior Knowledge Measures – Grade 9, Biology Domain

OPASS Score	Prior Knowledge: Average of Subjects	Prior Science Knowledge: Grade in Biology
Total	.065	.132
Scientific Inquiry	-.070	.056
Science Knowledge	.154	.168

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Assessment Accuracies of the OPASS System through Domain Experts

In addition to the evaluation by the correlations between the OPASS system and a comprehensive TIPS test with high reliability and validity, the evaluations of domain experts are also important for evaluating the accuracy of diagnostic reports (Ting et al, 2008). Therefore, an evaluation tool was developed to allow the domain expert to review and evaluate the accuracies of the diagnostic results of each student by checking the assessment portfolios. Three teachers as domain experts were invited to evaluate all students' experimental logs and score all statements in the diagnostic reports generated by the OPASS system. A statement's score was from 0 to 1. Figure 10 shows the statistical results in terms of different parts of the diagnostic report for three tests shown in Figure 6. According to evaluation results, the accuracies of the diagnostic reports are very high and meet the professional opinions of the teachers. In addition, the teachers also agreed that automatic, generated diagnostic reports can significantly assist teachers in understanding the status of students' inquiry abilities. This personalized diagnosis task is difficult for teachers to complete manually.



Figure 10: Statistical Results of Teachers' Evaluations for Diagnostic Report Accuracies

Analysis of Students' Feedback

Figure 11 shows the statistical results of the questionnaire (Cronbach's Alpha = 0.825) concerning students' satisfaction in terms of two classes (N=10 in Class 1 for Biology and N=24 in Class 1 for Physics), as shown in Table 6. The satisfaction degree is from 3.86 to 4.2 and the average is 4.17. This shows that most of students agreed that the diagnostic mechanism and the diagnostic report generated by the OPASS system are useful and can be expected to improve learning efficacy and assist in understanding the learning and operational problems in Web-based scientific inquiry experiments.

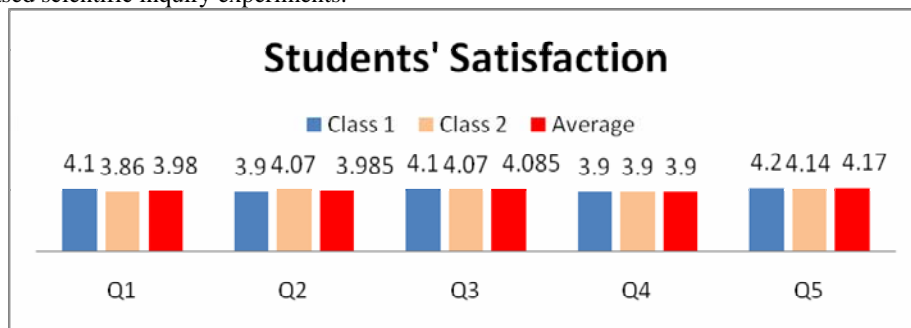


Figure 11: Statistical Results of the Questionnaire Concerning the Students' Satisfaction

LIMITATION AND DISCUSSION

Limitation of the OPASS

We discussed the limitations of our proposed OPASS approach in terms of the following three aspects.

(1) Capability of the Relation Model of Assessment Knowledge (AK)

According to the definition, the Relation Model can represent diverse requirements of scientific inquiry assessment performed by the OPASS system. However, some requirements of definitions, such as KOAP and Assessment Rules (AR), may not be wholly considered in this paper. Therefore, the Relation Model definition needs to be extended according to new requirements of scientific inquiry experiments.

(2) Capability of the Online Assessment Portfolio Diagnosis Process (OAPDP)

The accuracy of diagnostic reports depends on the correct definitions of teacher-defined AK and Diagnosis Rules (DR). Therefore, the mechanism of OAPDP and DR needs to be modified and adjusted if the definition of AK has been extended to meet various requirements of scientific inquiry experiments.

(3) Capability of the Usability of the OPASS

The OPASS can automatically analyze students' portfolios and generate personalized diagnostic reports concerning the learning problems with reasons and suggestions. However, it is still difficult and time-consuming to edit the teacher-defined AK without the support of the management tool; this will decrease the usability of the OPASS.

Discussion of the OPASS

According to evaluation results of the correlations of students' OPASS scores with their prior knowledge, including Science Knowledge and Scientific Inquiry, the students' master levels in terms of prior science knowledge in Physics and Biology subjects may not reflect similar performance in the OPASS and the TIPS scores because these testing mechanisms focus on the assessment of scientific inquiry abilities while the grade of

subjects at school may not consider this an important aspect. On the contrary, the purposes of OPASS and TIPS aim to assess the abilities of scientific inquiry, so the “Total” score of the OPASS significantly correlates 0.431 ($p < 0.05$) with the TIPS. Accordingly, this result proves that students with more prior scientific inquiry knowledge evaluated by the TIPS tend to perform better on the OPASS than students with lower levels. To investigate the correlations of sub-scores between the OPASS and the TIPS in terms of five scales: (1) Making Hypothesis; (2) Setting Variables; (3) Experimenting; (4) Graphing; and (5) Concluding, the reliability and validity of the OPASS can thus be evaluated. Results in Table 9 show that “Setting Variables” and “Graphing” of the OPASS have significant positive correlations (0.593 and 0.646, $p < 0.01$) with the ones of the TIPS, respectively. However, the “Making Hypothesis” and “Concluding” scales of the OPASS did not correlate with the ones of the TIPS due to the students’ prior knowledge attained through the practice section of the OPASS. This issue can be resolved by defining more various types of test items in these two steps of the OPASS.

Furthermore, the OPASS takes into account not only the science knowledge but also integrative abilities concerning scientific inquiry, so students were requested to operate the hands-on, Web-based operation experiments. Therefore, the learning problems concerning cause and effect operation behavior can thus be assessed by the OPASS, while the TIPS restricted to the confines of the paper-and-pencil test, cannot perform it well. This difference resulted in the “Experimenting” step of the OPASS did not obtain a significant positive correlation with the corresponding step of the TIPS. In addition, through the evaluation of three domain experts, the assessment accuracies of the diagnostic reports are very high (average is from 0.92 to 0.94) and meet the professional opinions of teachers. Therefore, the OPASS is valid and reliable for assessing the scientific inquiry abilities according to the analysis results in terms of the correlations and assessment accuracies.

In addition, during the OPASS experiments, teachers provided the following feedback:

Feedback 1: *“This system can indeed attract students’ interests and improve their motivation to understand learning problems. For example, some students tried the assessment experiment many times to correct their mistakes and actively discuss with teachers after reading their diagnosis reports.”*

Feedback 2: *“This system can help both students and teachers with scientific inquiry learning and assessment, although it is still under development.”*

According to the teachers’ aforementioned feedback, we can conclude that the OPASS is helpful and useful for both teachers and students in learning and assessing scientific inquiry capabilities.

Moreover, according to the survey of existing research described in the related work section, the methods employed as the criteria in this paper can be defined. These definitions can be based on what types of knowledge and capabilities (“substantive knowledge,” “procedural knowledge,” or “problem solving and integrated abilities”) are to be assessed to support what purposes (“general” or “scientific inquiry-based assessment”) by means of what the assessment items’ format (“selected response item,” “constructed response item,” “actual experimenting,” or “virtual and Web-based experimenting tool”), and how to evaluate performance (“manually” or “automatically” assessing) to provide students with what results (“summary score” or “diagnostic report”). Furthermore, the extensibility of proposed systems are primary concerns. Table 13 shows the comparison of our approach with existing articles in relation to assessment and diagnosis of scientific inquiry in terms of the aforementioned criteria.

The differences between our proposed approach and the existing studies are:

- (1) We define the representation of AK, consisting of Experiment and Evaluation Knowledge, considering not only Substantive Knowledge, but also Procedural Knowledge, and Problem Solving and Integrative Abilities to efficiently describe the requirements of the assessment and diagnosis for the purpose of Scientific Inquiry-based Assessment.
- (2) We define the Relation Model to integrate the teacher-defined AK and DR, which can be efficiently processed by employing the rule-based inference approach to identify the students’ learning problems.
- (3) We proposed an online automatic diagnosis scheme to efficiently analyze the assessment portfolios and generate the personalized diagnostic reports concerning not only the summary score of the assessment, but also the learning problems, with reasons and suggestions, in relation to conceptual knowledge, cause and effect operations, and skills of scientific inquiry.
- (4) We define the connection portocol of the OPASS to effeciently integrate with virtual and Web-based experiment tools/systems to enhance the extensibility of the system.

In contrast to the OPASS, existing research separately or partially considered the knowledge (Hwang, 2003; Chu et al., 2010) and abilities of scientific inquiry (Ting et al., 2008) to provide students with performance results described by either the summary score (Bennett et al., 2007, 2010) and limited suggestions (Ting et al., 2008)

only, or diagnostic reports concerning learning problems (and associated reasons and suggestions) assessed by manually evaluating students' portfolios (Hanauer et al., 2009).

Table 13: Comparison with Existing Approaches (O: Yes, ×: NO, Δ: Partial)

Method	Hwang (2003) & Chu et al. (2010)	Hanauer et al. (2009)	Ting et al. (2008)	Bennett et al. (2007, 2010)	Our approach
Scientific Inquiry-based Assessment	×	O	O	O	O
Knowledge Types of assessment	Substantive Knowledge	O	O	Δ	O
	Procedural Knowledge	×	O	O	O
	Problem Solving & Integrative Abilities	×	O	×	O
Formats of Assessment Items	Multiple choice questions (Selected Response Item)	O	O	O	O
	Open-ended questions (Constructed Response Item)	×	O	×	×
	Actual Experimenting	×	O	×	×
	Virtual & Web-based Experiment Tool	×	Δ	O	O
Performance Evaluation	Presented by (Summary) Score	O	O	O	O
Diagnostic Report about	Problem, Reason, Suggestion	O	O	Δ	×
	Cause and Effect Operation	×	Δ	×	O
Automatic Assessment Approach for Portfolio	O	×	O	×	O
Extensibility	O	Δ	O	×	O

CONCLUSION

In order to provide students with personalized learning guidance concerning not only the conceptual knowledge, but also the high-order, integrative abilities of scientific inquiry, the OPASS system was proposed to automatically generate personalized diagnostic reports by evaluating assessment portfolios collected from the Web-based scientific inquiry experiment. The diagnostic reports described the students' performance using not only the summary score of the assessment, but also learning problems with corresponding reasons and remedial suggestions. In the OPASS, students were requested to operate the hands-on, Web-based operation experiments. Therefore, learning problems concerning the cause and effect operation behavior can be assessed according to the teacher-defined assessment knowledge.

For the evaluation, experiments of the prototypical OPASS system have been conducted. The reliability and validity of the OPASS can be evaluated and proved due to significant correlations between the OPASS and a comprehensive Test of Integrated Science Process Skill (TIPS), and high accuracies evaluated by domain experts. Additionally, according to feedback from both students and teachers, the OPASS system can improve students' motivation to understand learning problems, and it can help teachers to understand students' learning status and provide more appropriate instruction.

In conclusion, the three main contributions of this paper include:

- (1) A proposal of the extensible Relation Model, which integrates teacher-defined assessment knowledge with diagnosis rule to represent relevant knowledge and abilities of scientific inquiry, processed by the rule-based inference approach.
- (2) A proposal of the Online Assessment Portfolio Diagnosis Process (OAPDP) to automatically generate personalized diagnostic reports concerning learning problems with reasons and suggestions.
- (3) A proposal of the OPASS's connection protocol to efficiently integrate with Web-based experiment systems to enhance the extensibility.

In the near future, the online diagnosis scheme in the OPASS will be enhanced, the support of the diverse assessments will be enriched, and the satisfaction of both students and teachers need to be further improved. The knowledge sharing and managing mechanism will be focused to reduce the cost of constructing the assessment knowledge. Behavior mining techniques will be applied to discover various potential experimental behaviors to assist teachers in designing assessment rules and key operation action patterns.

REFERENCES

- Bennett, R.E., Persky, H., Weiss, A.R., & Jenkins, F. (2007). *Problem solving in technology-rich environments: A report from the NAEP Technology-Based Assessment Project* (NCES 2007-466). Washington, DC: National Center for Education Statistics, US Department of Education. Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007466>
- Bennett, R.E., Persky, H., Weiss, A., & Jenkins, F. (2010). Measuring problem solving with technology: A demonstration study for NAEP. *The Journal of Technology, Learning, and Assessment*, 8(8), 1-44. Retrieved from <http://www.jtla.org>
- Burns, J.C., Okey, J.R., & Wise, K.C. (1985). Development of an integrated process skills test: TIPS II. *Journal of Research in Science Teaching*, 22(2), 169-177. Retrieved from [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1098-2736](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1098-2736)
- Chang, C.K., Chen, G.D., & Ou, K.L. (1998). Student portfolio analysis for decision support of Web based classroom teacher by data cube technology. *Journal of Educational Computing Research*, 19(3), 307-328. Retrieved from <http://baywoodjournals.com/index.php/OJS/index>
- Chen, G.D., Liu, C.C., Ou, K.L. & Liu, B.J. (2000). Discovering decision knowledge from Web log portfolio for managing classroom processes by applying decision tree and data cube technology. *Journal of Educational Computing Research*, 23(3), 305-332. Retrieved from <http://baywoodjournals.com/index.php/OJS/index>
- ChemCollective. (2011). *The ChemCollective: Online Resource for Teaching and Learning Chemistry*. Retrieved February 10, 2011, from <http://www.chemcollective.org/>
- Chu, H. C., Hwang, G. J., & Huang, Y. M. (2010). An enhanced learning diagnosis model based on concept effect relationships with multiple knowledge levels. *Innovations in Education and Teaching International*, 47(1), 53–67. doi:10.1080/14703290903525846
- Dalgarno, B., Bishop, A.G., Adlong, W., & Bedgood Jr., D.R.. (2009). Effectiveness of a virtual laboratory as a preparatory resource for distance education chemistry students. *Computers & Education*, 53, 853-865. doi:10.1016/j.compedu.2009.05.005
- Dillashaw, F.G. & Okey, J.R. (1980). Test of the integrated science process skills for secondary science students. *Science Education*, 64, 601-608. Retrieved from [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1098-237X](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1098-237X)
- Giarratano, J.C. & Riley, G.D. (2004, October). *Expert Systems: Principles and Programming* (4th ed.). North Carolina, NC: Baker & Taylor Books.
- Hameed, H., Hackling, M. W., & Garnett, P. J. (1993). Facilitating conceptual change in chemical equilibrium using a CAI strategy. *International Journal of Science Education*, 15(2), 221–230. doi:10.1080/0950069930150209
- Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., et al. (2004). Scientific Teaching. *Science*, 304, 521–522. doi:10.1126/science.1096022
- Hanauer, I.D., Hatfull, G.F., & Jacobs-Sera, D. (Eds.). (2009). *Active Assessment: Assessing Scientific Inquiry*. New York, NY: Springer. doi:10.1007/978-0-387-89649-6
- Hsu, Y.S., Wu, H.K., & Hwang, F.K. (2008). Fostering high school students' conceptual understandings about seasons: the design of a technology-enhanced learning environment. *Research in Science Education*, 38(2), 127–147. doi:10.1007/s11165-007-9041-1
- Hwang, G.J. (2003). A concept map model for developing intelligent tutoring systems. *Computers & Education*, 40(3), 217–235. doi:10.1016/S0360-1315(02)00121-5
- Jacobs-Sera, D., Hatfull, G.F., and Hanauer, D.I. (2009). Assessing scientific inquiry. In Hanauer, I.D., Hatfull, G.F., & Jacobs-Sera, D. (Eds.), *Active assessment: Assessing scientific inquiry* (pp. 31-43). New York, NY: Springer. doi:10.1007/978-0-387-89649-6_4
- Ketelhut, D.J., Dede, C., & Clarke, J. (2010). A multi-user virtual environment for building higher order inquiry skills in science. *British journal of Educational Technology*, 41(1), 56-68. doi:10.1111/j.1467-8535.2009.01036.x
- Kosba, E., Dimitroca, V., & Boyle, R. (2007). Adaptive feedback generation to support teachers in web-based distance education. *User Modeling and User-Adapted Interaction*, 17(4), 379-413. doi:10.1007/s11257-007-9031-z
- Lunsford, E., & Melear, C.T. (2004). Using scoring rubrics to evaluate inquiry. *Journal of College Science Teaching*, 34(1), 34-38. Retrieved from <http://www.nsta.org/college/>
- Matthews, M., Pharr, W., Biswas, G., & Neelakandan, H. (2000). USCSH: an active intelligent assistance system. *Artificial Intelligence Review*, 14(1–2), 121–141. doi:10.1023/A:1006508409887
- Mislevy, R.J., Almond, R.G., & Lukas, J.F. (2003). *A Brief Introduction to Evidence-Centered Design* (RR-03-16). Princeton, NJ: Educational Testing Service. Retrieved from <http://www.ets.org/Media/Research/pdf/RR-03-16.pdf>

- Mislevy, R.J., Almond, R.G., Yan, D., & Steinberg, L.S. (2000, March). *Bayes Nets in Educational Assessment: Where Do the Numbers Come From?* (CSE Technical Report 518). Retrieved from <http://www.eric.ed.gov/PDFS/ED443881.pdf>
- National Research Council (NRC). (1996). *National science education standards*. Washington, DC: National Academies Press.
- National Science Teacher Association (NSTA). (2004). Scientific inquiry. Retrieved February 10, 2011, from <http://www.nsta.org/about/positions/inquiry.aspx>
- Panjaburee, P., Hwang, G.J., Triampo, W., & Shih, B.Y. (2010). A multi-expert approach for developing testing and diagnostic systems based on the concept effect model. *Computers & Education*, 55, 527-540. doi:10.1016/j.compedu.2010.02.015
- Salajan, F.D., Perschbacher, S., Cash, M., Talwar, R., El-Badrawy, W., & Mount, G.J. (2009). Learning with Web-based interactive objects: An investigation into student perceptions of effectiveness. *Computers & Education*, 53, 632-643. doi:10.1016/j.compedu.2009.04.006
- Ting, C.Y., Phon-Amnuaisuk, S., & Chong, Y.K. (2008). Modeling and intervening across time in scientific inquiry exploratory learning environment. *Educational Technology & Society*, 11(3), 239-258. Retrieved from <http://www.ifets.info/>
- Ting, C.Y., Zadeh, M.R.B., & Chong, Y.K. (2006). A decision-theoretic approach to scientific inquiry exploratory learning environment. In Ikeda, M., Ashley, K., & Chan, T.W. (Eds.), *Intelligent tutoring systems, Lecture Notes in Computer Science* (pp. 85-94). New York, NY: Springer. doi:10.1007/11774303_9
- Wenning, C. (2007). Assessing inquiry skills as a component of scientific literacy. *Journal of Physics Education Online*, 4(2), 21-24. Retrieved from <http://www.phy.ilstu.edu/jpteo/jpteo0.html>
- Windschitl, M., & Andre, T. (1998). Using computer simulations to enhance conceptual change: The roles of constructivist instruction and student epistemological beliefs. *Journal of Research in Science Teaching*, 35(2), 145-160. Retrieved from [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1098-2736](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1098-2736)
- Yaron, D., Karabinos, M., Evans, K., Cuadros, J., Davenport, J., Leinhardt, G., & Greeno, J.G. (2008). The ChemCollective digital library. *Online Conferences on Chemistry (CONFICHEM)*. Retrieved from <http://www.chemcollective.org/about/paper/confchem08/index.html>
- Yaron, D., Karabinos, M., Davenport, J., Leinhardt, G., Greeno, J.G. (2009). The ChemCollective digital library", *Journal of Chemical Education*, 86(1), 123. doi:10.1021/ed086p123.1
- Yaron, D., Karabinos, M., Lange, D., Greeno, J.G., and Leinhardt, G., (2010). The ChemCollective: Virtual labs and online activities for introductory chemistry courses. *Science*, 328(5978), 584-585. doi: 10.1126/science.1182435
- Zachos, P. (2004). Pendulum phenomena and the assessment of scientific inquiry capabilities. *Science and Education*, 13(7-8), 743-756. doi:10.1007/s11191-004-6179-3
- Zachos, P., Hick, T. L., Doanne, W. E., & Sargent, C. (2000). Setting theoretical and empirical foundations for assessing scientific inquiry and discovery in educational programs. *Journal of Research in Science Teaching*, 37(9), 938-962. Retrieved from [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1098-2736](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1098-2736)