Assessing tenth-grade students’ problem solving ability online in the area of Earth sciences

Chun-Yen Chang a,*, James P. Barufaldi b, Ming-Chao Lin a, Yi-Chun Chen a

a Department of Earth Sciences, National Taiwan Normal University, 88, Section 4, Ting-Chou Road, Taipei 11677, Taiwan, ROC
b Center for Science and Mathematics Education, The University of Texas at Austin, USA

Available online 23 March 2006

Abstract

This study examined tenth-grade students’ \( n = 263 \) problem solving ability (PSA) online through assessing students’ domain-specific knowledge (DSK) and reasoning skills (RS) in Earth sciences as well as their attitudes toward (AT) Earth sciences related topics in a secondary school of Taiwan. The students’ PSA was evaluated based on a previous model (Chang, C. Y. (2004, November 26–27). Trends in assessing student earth science problem solving ability: the importance of domain-specific knowledge and reasoning skills in earth sciences. Paper presented at the Seoul Conference for International Earth Science Olympiad (IESO), Seoul, Korea; Chang, C. Y., & Barufaldi, J. P. (submitted). Does problem solving = prior knowledge + reasoning skills in science? An exploratory study. Journal of Experimental Education; Chang, C. Y., & Weng, Y. H. (2002). An exploratory study on students’ problem-solving ability in earth science. International Journal of Science Education, 24(5), 441–452) which empirically established that students’ PSA is a composite of DSK, RS and AT subscales. Major findings are as follows: (a) The correlation coefficient among students’ DSK, RS and AT was relatively small, indicating that these subscales might have successfully represented different constructs of students’ PSA; (b) a significantly positive correlation existed between students’ PSA total scores and each subscale. It is, therefore, suggested that students’ PSA may be potentially assessed online by measuring their essential components in the area of Earth sciences.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Problem solving; Secondary school; Science education; Assessment

* Corresponding author. Tel.: +886 2 2935 4393; fax: +886 2 2933 3315.
E-mail addresses: changcy@ntnu.edu.tw (C.-Y. Chang), jamesb@mail.utexas.edu (J.P. Barufaldi), ardo168@hotmail.com (M.-C. Lin), chyichun@ese.geos.ntnu.edu.tw (Yi-Chun Chen).

0747-5632/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2006.02.014
1. Introduction

Developing and enhancing the problem-solving abilities of students have long been important objectives of science education. Problem-solving ability is generally viewed as the ability to think critically, to reason analytically, and to create productively, which all involve quantitative, communication, manual, and critical-response skills (American Association for the Advancement of Science, 1993). Research has substantiated the important contributions of domain-specific knowledge to students’ domain learning (Alexander & Judy, 1998; Murphy & Alexander, 2002). After a study of related literature on problem solving and extensive discussion, Project 2061 (American Association for the Advancement of Science, 1993, p. 282) concludes that:

- Students’ ability and inclination to solve problems effectively depend on their having certain knowledge, skills, and attitudes.
- Quantitative, communication, manual, and critical-response skills are essential for problem solving...

The most recent standards for grades 1–9 science and life technology curriculum in Taiwan (Chang, 2005; Ministry of Education, 2001) also emphasize the need to develop students’ abilities of independent thinking and problem solving as well as stimulate their creativity and potential, as one of six major goals of primary and middle school education. Recent science education standards in the USA also propose that teaching must involve students in engaging problem solving, planning, decision making, and group discussions (National Research Council, 1996).

It is interesting to find that problem-solving ability, science knowledge, and science process skills have been closely linked in research literature. To be successful in problem solving, it appears that students need to have some background knowledge and to possess certain science process skills. Accordingly, some researchers have suggested the possibility of improving students’ problem-solving ability through teaching science-process skills (Germann, 1991), or vice versa (Geban, Askar, & Ozkan, 1992; Holley, 1996a, 1996b). Others have proposed employing different modes of problem-solving associated instruction to improve students’ creative/critical thinking, problem-solving ability, science-process skills or Earth science achievement (Basaga, Geban, & Tekkaya, 1994; Chang, 2001a, 2001b, 2002, 2003; Chang & Barufaldi, 1999; Chang & Mao, 1999; Germann, 1989; Tobin & Capie, 1982).

Researchers have assessed the relative contribution of knowledge and process to problem solving and processing skills (Champagne & Klopfer, 1981). Twenty-seven eighth-grade students participating in the researchers’ study were asked to solve two types of verbal problems (analogies and set-membership) drawn from concepts in physical geology. The set-membership problem involves students in grasping the relationship that is common among three of the four given terms. The results revealed that processing skills and science knowledge both contribute more to the successful solution of set-membership problems than to the solution of analogies problems. The researchers found that students with higher processing skills on structuring tasks would score higher on set-membership problems in physical geology than those with lower processing skills. The educational implication of this study suggested that good instruction should consist of process and content. In several previous studies in this series (Chang, 2004; Chang & Barufaldi, under
review; Chang & Weng, 2002), it was found that students’ problem-solving ability (PSA) and their domain-specific knowledge (DSK), reasoning skills (RS), and attitudes (AT) seem to be closely related using the traditional testing methods. We empirically established that students’ PSA is basically a composite of DSK, RS and AT subscales as shown in Fig. 1. Research on student PSA has focused mostly on traditional paper-and-pencil or performance type assessments. However, relatively little is known about methods of evaluating student PSA through online web-based testing. Furthermore, limited research is available on the exploration of student PSA, DSK, RS and AT in the areas of Earth sciences for senior high students with an online testing system. Therefore, this study took further steps and attempted to assess students’ problem solving ability online by mapping their domain-specific knowledge of, reasoning skills in, and attitudes toward Earth science related topics in a secondary school in Taiwan.

2. Method

2.1. Sample

The participants in the study were 263 tenth-grade students (131 females and 132 males), with a mean age of 16 years, enrolled in a compulsory Earth science course (seven Earth science classes) at a public senior high school located in the middle region of Taiwan. The Earth science course, taught 2 h per week for one semester, is required of every tenth-grade student in the secondary schools in Taiwan.

2.2. Instruments

Quantitative data were obtained online on students’ domain-specific knowledge and reasoning skills as well as attitudes through the use of the domain-specific knowledge test (DSKT), the reasoning skills test (RST), and the attitudes test (AT) in the area of Earth sciences to determine their interrelationships.

2.3. Domain-specific knowledge test (DSKT)

The domain-specific knowledge test (DSKT) is a 14 question multiple-choice test designed to measure students’ Earth science knowledge central to the debris-flow hazards topic, encompassing students’ knowledge and understanding of scientific facts, concepts, principles, laws, and theories. The test consists of multiple-choice items for the objective and online testing purposes. A panel of experts including three university professors from the Department of Earth Sciences, National Taiwan Normal University and three high

Fig. 1. The PSA as a composite of DSK, RS and AT in the area of Earth sciences.
school teachers verified the content validity of the instrument. These experts checked the
degree of correspondence between the curriculum content and DSKT test items and deter-
determined that the nature of the test items corresponded to the important concepts introduced
in the curriculum. The Kuder-Richardson Formula 20 (KR-20) was used to determine the
reliability coefficient of DSKT. The reliability coefficient ranged from 0.55 to 0.78. DSKT
items emphasize recollection of concepts or recall of ideas in debris-flow hazards and
require students to understand ideas or concepts in related topics. An example of test items
included in the DSKT is shown in Fig. 2 (correct answers denoted by *).

2.4. Reasoning skills test (RST)

The reasoning skills test (RST) is an 11 question multiple-choice test designed to
measure students’ reasoning skills to assess students’ reasoning ability in Earth science
in addition to their domain knowledge. The RST encompasses the following three compo-
nents: data interpretation and explanation (interpreting information presented in tabular
or graphic form), conflicting views debated in science (evaluating two or three viewpoints
on a specific phenomenon), and summarizing of experimental data (comprehending or

Fig. 2. One item example of DSKT.

8. Given the same formations with the same dips in the following cross sections,
which one might BEST represent a dip slope that could possibly cause a landslide
or debris-flow hazard in location ★ ?
analyzing the design of an experiment), which all focus on students’ ability to elicit and analyze explanations and are considered the most important elements in scientific reasoning. The same panel of experts also established the content validity of the RST. They checked the degree of equivalence between the guiding principle of reasoning ability and the RST items and confirmed that the nature of the test items were aligned with the important skills introduced in the guidelines. The reliability coefficient was estimated at 0.50–0.76 (KR-20) for the previous and current study. An example of test items included in the RST is shown in Fig. 3 (correct answers denoted by *).

2.5. Attitudes test (AT)

The AT consists of 15 items and was designed specifically to tap students’ attitudes toward debris-flow hazard topics using three subscales measuring learning interest in the topic (6 items), preference for the test format (5 items) and self-confidence in the topic (4 items). This instrument was constructed and developed by the researchers and was validated by the panel of experts previously cited. Factor analysis and principle component analysis with varimax rotation was then used to clarify the structure of AT scales, which

Fig. 3. One item example of RST.
can be characteristically grouped into three orthogonal factors accounting for 64.56% of total variance explained. Reliability was also established through internal consistency. The Cronbach reliability coefficients of 0.85, 0.74, 0.57, and 0.77 were reported for the AT, learning interest, preference for the test format, and self-confidence subscales respectively. The following three questions are samples cited from the AT instrument: (a) I am very interested in the topic on the test (interest subscale); (b) I like the test format very much (preference for the test format scale); and (c) I am satisfied with my performance (self-confidence subscale); and each AT item has a five-point scale with categories ranging from “agree a lot” (5) to “disagree a lot” (1). The mean scores, total/subscale scores divided by the total/subscale items for each student’s responses on the items, were used as indicators to display their degree of AT. The highest possible mean score on the AT is 5 and the lowest possible mean score is 1, therefore, 3 is treated as the median (neutral) AT. Table 1 summarizes the range of scores, means, standard deviations and reliability coefficients of the study.

When pushing assessments from the end of paper-based testing to the end of online testing, one purpose is to make it easier to integrate formative or even summative evaluation into nowadays’ ubiquitous web-based learning environments. On the one hand, a large quantity of samples can be rapidly assessed and analyzed by online tests in supporting advanced educational studies; on the other hand, online learners’ learning statuses can then be recorded, scored, and captured in computer systems, and subsequent pedagogical feedbacks or remedial instructions could be provided by computers or teachers accordingly. Therefore, the validity of the online test responses should be discussed from two perspectives, (1) the perspective of supporting large scale educational assessments, and (2) the perspective enabling online instructional decisions.

In this study, the first perspective that supports large scale educational assessments was emphasized. Therefore, the online tests were administrated by teachers in secondary schools’ computer rooms. In other words, students were tested in a centralized manner under the supervision of staff persons and the possibility of students’ cheating, such as discussing with other students, accessing external resources, etc, was minimized and deemed equivalent to traditional paper-based testing. Accordingly, the data used in this study is considered to be valid in terms of characterizing students’ problem-solving abilities and the way we determine the validity in this case is also identical to conventional paper-based tests. However, it is noteworthy that once the online test modules were integrated into web-based learning environments in which students’ learning is of a self-paced nature, it is unlikely that the testing process can be always monitored by teachers. The major purpose of obtaining the data is to inform online instructional decisions, either made by computers or online teachers. Under this condition, it is necessary to revisit the way we define and determine validity for the instrumentation. This issue seems to be a new and interesting one that may merit further investigation.

Table 1
Summary of means, standard deviations and reliability coefficients (n = 263) of the DSKT, RST, and AT scores

<table>
<thead>
<tr>
<th>Instruments</th>
<th>Number of items</th>
<th>Score range</th>
<th>Mean</th>
<th>SD</th>
<th>Reliability coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSKT</td>
<td>14</td>
<td>5–14</td>
<td>10.88</td>
<td>1.48</td>
<td>0.55–0.78</td>
</tr>
<tr>
<td>RST</td>
<td>11</td>
<td>0–10</td>
<td>5.14</td>
<td>1.67</td>
<td>0.50–0.76</td>
</tr>
<tr>
<td>AT</td>
<td>15</td>
<td>1–4.73</td>
<td>3.19</td>
<td>0.55</td>
<td>0.78–0.85</td>
</tr>
</tbody>
</table>
2.6. Design and data analysis procedures

A correlation research design (Campbell & Stanley, 1966) was adopted for the study. The participants were tested online with the web-based DSKT, RST and AT for a total of 2 h in December, 2004. Care was taken to ensure that the procedures for administering these tests online were as similar as possible. The relationships among students’ domain-specific knowledge, reasoning skills and attitudes were then determined for the total group of students using a Pearson product-moment correlation method.

To meet contemporary calls for improvement in the interpretation and reporting of quantitative research in education (Rennie, 1998; Thompson, 1996), this study reports practical significance (effect magnitudes) along with statistical significance test. The effect size index $r$ was used since it is more appropriate for measures of association. According to Cohen’s rough characterization (Cohen, 1988), $r = 0.1$ is deemed as a small effect size, $r = 0.3$ a medium effect size, and $r = 0.5$ as the large effect size, in light of the nature and characteristics of behavioral or social sciences (p. 78–83). This kind of data presentation method is important in terms of interpreting research results. Researchers have cautioned that using only the result of statistical significance testing in statistical inference (Cohen, 1988; Daniel, 1998; McLean & Ernest, 1998) is insufficient. This is mainly because the computation of statistical significance is related to the sample size involved in the analysis. Moreover, it is common to observe a statistical significance with a large sample size, even if there was actually little practical effect. Therefore, to protect against statistical significant finding in terms of large sample size for this study, effect magnitudes were also reported along with statistical significance test. The correlations were determined using SPSS 11.5 (Statistical Package for Social Sciences version 11.5 made by SPSS Inc., Chicago, IL).

3. Results and discussion

Table 2 presents a $4 \times 4$ matrix of Pearson product-moment correlation coefficients on students’ DSKT, RST, AT and their composite (PSA total) scores. The major findings revealed that (1) the correlation coefficient among students’ DSK, RS and AT are quite small ranging from only 0.046 to 0.174 ($r^2 = 0.02–0.03$), small effect sizes. This finding indicates that these subscales might have successfully represented different constructs of students’ PSA, as shown in the RS and AT columns of Table 1; (2) a significant positive and high correlation exists between students’ PA (total scores) and their DSK ($r = 0.631$, $p < .01$), RS ($r = 0.684$, $p < .01$) and AT ($r = 0.618$, $p < .01$) scores, with large effect sizes.

Table 2
Summary of Pearson product-moment correlation coefficients among students’ DSK, RS, AT and their composite (PSA total) scores

<table>
<thead>
<tr>
<th></th>
<th>DSK</th>
<th>RS</th>
<th>AT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSK</td>
<td>1</td>
<td>0.174**</td>
<td>0.046</td>
<td>0.631**</td>
</tr>
<tr>
<td>RS</td>
<td>1</td>
<td>0.148*</td>
<td>0.148*</td>
<td>0.684**</td>
</tr>
<tr>
<td>AT</td>
<td>1</td>
<td>0.618**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

* $p < 0.05$.  
** $p < 0.01$.  

Where \( r^2 = 0.40 \) for DSK, \( r^2 = 0.47 \) for RS and \( r^2 = 0.38 \) for AT as shown in the last column of Table 1; and (3) students’ PSA are more significantly correlated with their reasoning skills \( (r = 0.684, \quad r^2 = 0.47, \quad p < .01) \) than with their domain specific knowledge and attitudes \( (r = 0.618–0.631, \quad r^2 = 0.38–0.40, \quad p < .01) \), as shown in the last column of Table 1.

Overall, there existed a significantly positive correlation between problem-solving ability and domain-specific knowledge, reasoning skills and attitudes of tenth-grade students in the area of Earth sciences in a secondary school in Taiwan. Correlations were \( r = 0.62–0.68, \quad r^2 = 0.38–0.47. \)

The results are expected since students’ PSA is a composite of their respective subscales. However, the results further verify the previous model (Chang, 2004; Chang & Weng, 2002), that students’ PSA are highly correlated with their DSK, RS, and AT components. These results are reasonable since solving a problem is less difficult when students know more about the subject, possess more reasoning skills, and hold more positive attitudes toward the subject; conversely, that problem solving ability in a ‘knowledge-skill-attitude vacuum’ leads to minimal understanding in the area of Earth sciences.

The aforementioned results are, to a certain degree, consistent with findings from other studies (Alexander & Judy, 1998; Murphy & Alexander, 2002), which demonstrated that students’ subject-matter knowledge plays an important role in their domain learning. Furthermore, the above results demonstrate that reasoning skill is more correlated with PSA compared to DSK and AT subscales. The finding is also somewhat consistent with results and conclusions from other studies in this area. For example, reasoning ability was found to be significantly related to students’ success at solving stoichiometry problems (Robinson & Niaz, 1991), concept acquisition (Lawson & Worsnop, 1992), and their science grades (Bitner, 1991). These findings, along with conclusions from this study, have a practical implication for the practice of science teaching and learning in the area of Earth sciences. If students are well developed in their domain-specific knowledge, reasoning skills and attitudes in regular science classrooms, they might be capable of improving their problem solving ability.

The current study, along with the researchers’ work, found that 38–47% of the variation in problem-solving ability was attributable to differences in domain-specific knowledge, reasoning skills, and attitudes of students with large effect sizes. These data further confirmed that DSK, RS and AT may serve not only as important components of successful problem solving in Earth sciences, but also might be strong predictors (or explainers) when test development or other related issues are considered. Since solving a problem requires students to fully understand the problem, make appropriate inferences, and finally find solutions to the problem. These findings suggest that Earth science instruction or assessment in the secondary schools should emphasize the merit of enhancing students’ domain-specific knowledge, reasoning skills, and attitudes of students because pupils receiving this type of instruction or assessment are provided with the opportunity to acquire basic knowledge, develop reasoning skills, think independently, and foster positive attitudes. Accordingly, the enhancement and development of knowledge and reasoning skills as well as attitudes might help students to improve their problem-solving abilities.

This study found that the correlation coefficients among students’ DSK, RS and AT are rather small, signifying that these components may characterize different constructs of students’ PSA. As suggested by Stevens (2002, p. 345), covariate variables should be significantly correlated with the dependent variable and that have low correlations among themselves (Stevens, 2002). Only two to three percent of the variation in respective sub-
scales was attributable to differences in domain-specific knowledge, reasoning skills and attitudes of students with small effect sizes. These data indicate that students’ PSA may be assessed online through mapping their essential components such as DSK, RS, and AT in the area of Earth sciences.

The results of the study offer an alternative perspective or solution to reduce the difficulty of assessing students’ higher-order thinking ability, such as problem solving ability, through web-based testing as encountered by many researchers worldwide. Since the DSK, RS, and AT have proven to effectively reflect PSA in the area of Earth sciences as a result of this study, online web-based test developers might consider adopting a similar approach in assessing student higher-order thinking ability as proposed by the study and will somehow decrease computer-software development costs. In addition, the approach will move online assessment toward an alternative direction that considers essential components of PSA and implements appropriate elements of online assessment techniques. Besides, web-based test developers in problem solving ability may wish to reflect on the results of the study when designing tests to assess students’ fundamental knowledge, reasoning skills, and attitudes. Instead of developing complicated ways and complex techniques to assess students’ PSA or other higher-order thinking ability online, we might need to go back to the basics in terms of developing highly reliable online instruments with sound validity to assess students’ knowledge, skills, and attitudes in the subject matter. Furthermore, the model proposed by the study could be applicable in other areas of science.

The results of the study provide new information on the interrelationship of students’ problem solving ability in, domain specific knowledge of, reasoning skills in, and attitude in Earth science related topics through web-based testing methods. It is noted, however, that a positive correlation does not ensure causality. Consequently, causality merits further investigation, which is presently being investigated in Taiwan. In addition, to further corroborate the model presented in the current paper, we are now developing a user modeling framework which may assess and model learners’ problem-solving ability from their self-explained ideas against a specific scenario of problem-solving criteria with the aims of additionally substantiating the findings of the study.

Acknowledgements

This paper is based on work supported by the National Science Council under Contract Nos. NSC 92-2524-S003-010, 93-2524-S003-001, 94-2524-S003-014, and was completed while the first author was a Senior Research Fellow at Center for Science and Mathematics Education, The University of Texas at Austin, USA in 2005. The authors wish to thank the NSC of Taiwan for financing this project. The data presented, the statements made and the views expressed are solely the responsibility of the authors. The authors gratefully acknowledge the assistance of Jing-Ye Jhan and Hao-Chuan Wang.

References


Chun-Yen Chang, Ph.D., is Professor and Chair of Department of Earth Sciences at National Taiwan Normal University, Taipei, Taiwan. His major research interest includes computer-assisted instruction, problem-solving models, science curriculum development, inquiry teaching, and pedagogical understandings and conceptual change instruction in the area of science education.

James P. Barufaldi, Ph.D., is Ruben E. Hinojosa Regents Professor and Director of Center for Science and Mathematics Education, The University of Texas at Austin, USA. His special areas of interests are curriculum design, instructional strategies, implementation, evaluation, and science teacher education. He is currently investigating the process of building successful collaboratives in the science education community and variables, which may contribute to high intensity, sustained collaboration.

Ming-Chao Lin, M.S., is Doctoral student of Department of Earth Sciences at National Taiwan Normal University. His research interest is web-based learning and test development.

Yi-Chun Chen, B.S., is Master student of Department of Earth Sciences at National Taiwan Normal University. Her research interest includes problem solving and test development.