Employing Online S-P Diagnostic Table for Qualitative Comments on Test Results

Chien-hwa Wang[1] and Cheng-ping Chen[2]
[1]Graphic Arts and Communications, National Taiwan Normal University, Taipei, Taiwan
[2]Information and Learning Technology, National University of Tainan, Tainan, Taiwan
pw5896@ms39.hinet.net
chenjp0820@yahoo.com.tw

Abstract: The major concerns of adaptive testing studies have concentrated on effectiveness and efficiency of the system built for the research experiments. It has been criticised that such general information has fallen short of providing qualitative descriptions regarding learning performance. Takahiro Sato of Japan proposed an analytical diagram called Student-Problem Chart (S-P Chart) in the 1970. The S-P Chart is able to establish a learning diagnostic table which comments student learning performance in a verbal form. The advancement of computer technology has made the S-P analytical process more applicable for school teachers. This study examined how online comments provided by the S-P diagnostic table could affect the students' learning attitude. One hundred sixth grade students were selected to be the subjects of the study. An online embedded test was given to the subjects and an S-P diagnostic table was drawn by the computer to display instant comments on each student's learning performance. A Questionnaire survey and in-depth interviews were performed after the experiment. Results indicated that students liked the online qualitative comments. This is because students were able to instantly understand why they performed well/poor in the test, which is much beyond the numerical scores can explain. The results also showed that the online S-P diagnostic table made students more circumspect on answering the test questions in order to reduce careless mistakes. Students would also be more likely to review what missed on the test. However, the S-P comment table seemed to have no effect on improving their learning performance. An online iterative drilling platform was consequently built to incorporate with the S-P diagnostic process to assist poorly performed students. It may effectively work with the S-P diagnostic process to provide constructive remediation for the students who exhibited a poor performance on the S-P chart.

Keywords: adaptive test, the student-problem chart, learning attitude, iterative drilling

1. Introduction

In the past 20 years, research on computer-supported adaptive assessment of learning achievement has been conducted in various aspects. A line of research has emphasized on-task adaptation of individual differences to the testing environment and has frequently used Item Response Theory (IRT). IRT-based adaptive testing has the advantage of being able to monitor the on-task performance of individual examinees, and the test item that characterizes most matched examinee latent traits can be selected (Xu & Douglas, 2006; Al-A'ali, 2007). Statistical methods such as confidence intervals and the Bayesian reasoning procedure (Frick et al., 1987) were employed to minimize the probability of making computer-generalized error decisions. In an IRT-based adaptive test, test will terminate immediately once a pre-set mastery level has achieved, or a non-mastery level has reached. A major requirement for IRT-based adaptive testing is that numerous samples are necessary for validating each test item. Such a requirement has been criticized to be inappropriate for school teachers in constructing embedded quizzes in their daily classroom activities (Wang and Chen, 2007). Many of the early efforts of adaptive testing were conducted associated with IRT until a simplified method adopting the Sequential Ratio Probability Test (SPRT) was introduced to accommodate the real school environment (Wang & Chuang, 2002). SPRT-based adaptive testing is relatively simple to operate and performs less complex mathematical algorithms. However, the major concerns of these adaptive testing studies have concentrated on the effectiveness and efficiency of the tests per se and do not account for student and teacher perspectives. After performing a typical adaptive test, both students and teachers can only be notified on their personal results (e.g., pass/fail or grades A, B, and C). These simple grading symbols are unable to explain the quality of learning. Criticisms have indicated that such general information falls short of providing qualitative descriptions on learning performance (Wainer, H. & Liely G. L., 1987; Yu, M. N., 2003)..

Reference this paper as: Wang, C. and Chen, C., "Employing Online S-P Diagnostic Table for Qualitative Comments on Test Results" *The Electronic Journal of e-Learning* Volume 11 Issue 3 2013, (pp263-271), available online at www.ejel.org

2. The student-problem chart

Takahiro Sato of Japan proposed an analytical diagram called the Student-Problem Chart (S-P Chart) in 1970, an analytical tool that graphically presents the relationship between test item difficulty pattern and student response pattern. Student learning performance and test item appropriateness are analysed by four indices: disparity index, homogeneity index, item caution index, and student index. Wu (1998) indicated that these indices help teachers diagnose student learning conditions, instructive achievement, and problem quality. Teachers are also able to use the analysed S-P chart data to draw a performance profile curve for each student and construct a learning diagnostic table to qualitatively describe learning performance by calculating the indices of the S-P Chart. The following is a general description of how the process works.

The theoretical basis for S-P-Chart analysis is IRT. However, the S-P-Chart use of IRT significantly differs from that of adaptive testing for test efficiency. The S-P Chart draws two IRT-shape curves: the Student Curve (S-curve) and the Problem Curve (P-curve), according to data recorded in the S-P Chart. The S-curve shows how students agree with problems, and the P-curve shows how problems agree with the students. In an ideal situation, these two curves should coincide, but in a practical situation these two curves diversify (Sato, 1975, 1980, 1985). A sample S-P Chart is shown in Table 1.

Table 1: A sample S-P chart

Sample S-P Chart									
N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
4	1	1	1	1	1	1	1	1	8
3	0	1	1	1	1	1	1	1	7
11	1	1	1	1	1	1	1	0	7
13	1	1	1	1	1	1	0	1	7
15	1	1	1	1	1	0	1	1	7
2	1	1	1	0	1	1	1	0	6
8	1	1	0	1	1	1	0	1	6
10	1	0	1	1	1	1	1	0	6
12	1	1	1	1	0	1	0	1	6
14	0	1	1	1	1	1	0	1	6
1	1	1	1	1	1	0	0	0	5
6	1	1	1	1	0	0	1	0	5
7	1	1	1	0	0	1	0	0	4
9	1	1	0	1	1	0	0	0	4
5	1	0	0	0	1	1	0	0	3
Total	13	13	12	12	12	11	7	7	87

On the S-P Chart, each "1" stands for a correct answer, and each "0" indicates an incorrect answer. The columns of the table sum the correct student answers for each test item, and the rows of the table sum the amount of correct answers for each student. Data on this table were sorted, so that the more questions students answered correctly, the higher the row was placed; for test items, the more questions students answered correctly, the farther left the column was assigned. After the sorting process, a ladder-shaped S-curve was drawn along the right side of the cell boxes according to the correct answers of each student (a solid line on the chart), and a P-curve was also drawn along the bottom of the cell boxes according to the total number of correct answers each item accumulated (a dotted line on the chart).

2.1 Disparity index

After drawing the S- and P-curves, the next step involved calculating the disparity index (D*). According to Yu (2002), D* can be calculated as follows:

$$D^* = \frac{C}{4Nn\bar{p}(1-\bar{p})D_B(M)} \qquad (1)$$

where N is the total number of students, n is the number of test items, p is the average ratio of the correct response, and DB(M) is the constant array that can be accessed by M (M=Gussan(Nn +0.5)). The value range of the disparity index is between 0 and 1. Sato also indicated that a larger disparity index signifies a larger variation between the S-curve and the P-curve. When the disparity index is 0, the two curves are matched perfectly and overlapped.

Caution index

The two types of caution index are caution index for items and caution index for students. The caution index for students is related more to our study. The formula for calculating the caution index for students is as follows (Chen, Lai, and Liu, 2005; Yu, 2002):

$$CS_I = \frac{\sum_{j=1}^{yi.}(1-y_{ij})(y.j) - \sum_{j=yi.+1}^{n}(yij)(y.j)}{\sum_{j=1}^{yi.} y.j - (yi.)(\mu')} \qquad (2)$$

Y is a two-dimensional array that stores both test-taker scores and the percentage of correct responses on test items in descending order (shown in Fig. 1). yij represents the ith person's response on item j. yi. represents the number of correct responses of student i on all test items, y.j represents the number of correct responses for all test-takers on test item j, .μ represents the average score of the students, and μ' represents the average number of correct responses on all test items.

2.2 Learning diagnostic table

The diagnostic result of student learning performance is divided into six categories based on the student caution index and the student correct response percentage. Satio (1985) gave specific qualitative explanations to each of the six categories in terms of student's performance. According to Chen, Lai, and Liu, (2005), and Yih and Lin (2010), Satio's original (in Japanese) is interpreted in Figure 1. Category A indicates that the student has good learning performance and high testing stability. Category B shows middle stability of the student. Category C indicates poor student learning capability, low mastery of the content. Category A' shows good test performance, but occasional wrong responses caused by carelessness. Category B' indicates mediocre student mastery of the content and occasional wrong responses caused by carelessness. Category C' shows low student stability, and very low mastery of the content.

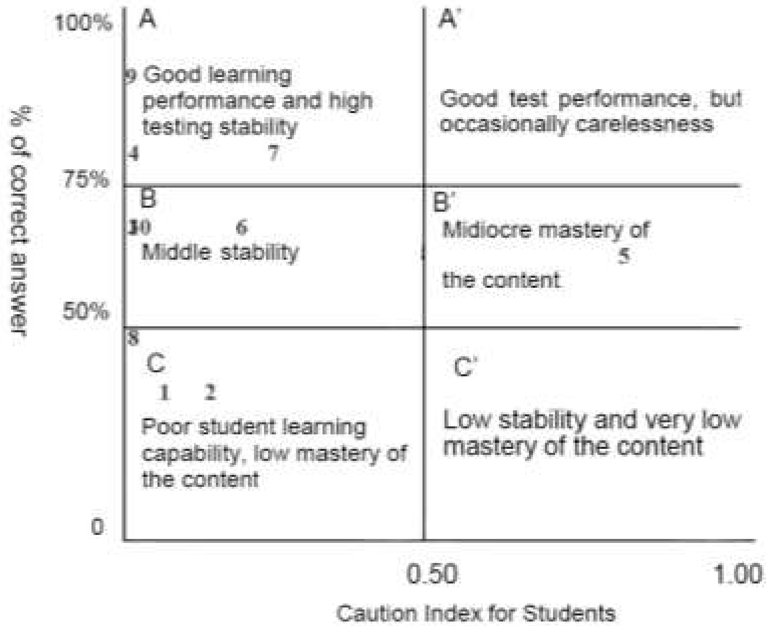

Figure 1: The S-P diagnostic comment table

The learning diagnostic table comments on student learning performance in verbal (non-numerical) form. Using such verbal comments, students can self-assess their performance more qualitatively, and teachers can also better counsel their students.

3. Online assessment using the S-P feedback table

Since late 1980's, the S-P chart has been introduced to the classroom evaluation. Harnisch (1987) designed a multi-purpose, multi-user evaluation system to improve the quality of the reporting of information gathered in testing and evaluation practices. Large-scale experiment was done with 5,945 students and 303 teachers, and the use of overall classroom, school, and district skill proficiency summaries was also discussed in detail. Chacko (1998) discussed the use of S-P chart in identifying learning problems and consequent grouping of students for corrective feedback purposes. He concluded that in order to take the right decisions about instruction, the items should be valid as well as reliable.

The fast advancement of computer technology has made the S-P analytical process more applicable for school teachers. However, local research on the computer-supported S-P Chart has focused on algorithm building and system efficiency appraisals. This line of research started with Chen (1989), who argued that computers should replace hand calculations of S-P indices. He also estimated that the entire analysis process for a 50 x 50 S-P chart should be completed in less than 30 min, including all data entry works. Extended efforts on computerization of the S-P Chart continue. Hu, Chen, and Yan (2005) first introduced a computerized learning assessment system employing the S-P chart, and Zhan (2006) recorded student time spent for answering the whole set of test items as an additional variable for S-P chart analysis.

More recent works continued to be done by local researchers. Lin and Yih (2008) introduced an integrated method that combined the concept structure analysis and the S-P chart. The concept structure analysis could analyse individualized concepts structure and S-P chart could classify students and items into proper types. The results indicated that students with different learning types own varied concept structures. Chang, Hsu, Lee, and Huang (2009) developed a tutoring system that used the S-P chart to classify the learning types for each student, and a rule-space model then was employed to disclose the non-mastered concept. They claimed that their tutoring system could provide expedient suggestions that may help students review the misconceptions instantly. Wang, Sheu, Liang, Tzeng, and Masatake (2012) included the equation of Rasch model into the S-P chart to create a Gray Student-Problem (GSP) chart, the results of their experiment indicated that the graphical GSP chart was able to provide information for teachers to develop remedial instructions.

These local studies were valuable, but did not provide any information on student and teacher attitudes regarding qualitative comments from the S-P diagnostic table.

4. Methodology

Instead of replicating these studies, this study presents answers to questions of how online comments provided by the S-P diagnostic table could affect student learning attitudes and teacher reactions. The specific research questions are as follows:

- Do the S-P online diagnostic comments affect learning attitude afterwards?
- Does subject matter affect student attitude toward the S-P online diagnostic comments?
- What are the teacher appreciations of the S-P online diagnostic comments?

To collect necessary data to answer these research questions, an experiment employing the S-P online diagnostic comments was conducted. An S-P diagnostic online computer program was constructed for the study. The program has four modules: a test item pool, a testing system, a testing result database, and an online diagnostic comments display system.

A total of 100 sixth grade students from six elementary schools in northern Taiwan were selected to be the subjects of the study. Reading and Math were the subjects for the experiment. Test items were developed by the individual school to accommodate the learning agenda of local students. The distribution of students and subject matter are shown in the following table. For answering the research question 3, teachers involved in the experiment were interviewed. The interviews centred upon teaching effectiveness, testing efficiency, learning effectiveness, and appraisals of the online system. Answers were collected via emails and phones.

The experimental process is as follows: Following regular classroom instruction of selected learning units, an online embedded test was given to the participants and an S-P diagnostic table was drawn by the computer to display instant comments on the learning performance of each student. Teachers then explain the meaning of each category in which students were allocated.

Table 2: Experimental content subjects and distribution of students

	Subject	No. of Students
School A	Reading	10
School B	Reading	10
School C	Reading	10
School D	Math	10
School E	Reading	30
School F	Math	30

A questionnaire survey and in-depth interviews were performed to collect necessary information following the experimental testing process. The questionnaire is for acquiring information from students and is divided into two parts: learning attitude appraisal with 15 questions and attitude toward the online S-P system with 14 questions. A pilot study was performed, and internal validities of the questionnaire were examined. Cronbach's α is .705 for Part 1 and .762 for Part 2. The interviews were conducted to acquire teacher opinions. Four teachers experienced in the full experimental process were interviewed.

5. Results and discussions

Questionnaire data were collected immediately after the experimental process and teacher interviews were done within one week after the experiment. Descriptive statistics were done to answer the research question 1 and 2. Coded interviewing records were used to describe teacher's opinions. Ninety-five valid questionnaires were collected and four teachers attended the experiment were interviewed.

5.1 Results regarding student's learning attitude

The average score for this category is 3.01. Seven out of fifteen questions obtained scores over 3.00, and scores for all fifteen questions were over the norm (2.50). We conclude that after employing the S-P diagnostic process, students gained a positive learning attitude. By Further examining each individual question in this category, we found that "I will pay more attention in answering questions to avoid careless mistake" obtained the highest score in this category. The second highest score is associated with the item "With the S-P diagnostic comments, I'll try seriously to answer the question even I don't know how to make the solution initially". Oppositely, the application of the S-P diagnostic comments seemed to have less effect on reinforcing student's intention to review after the class, and also less affected student's regular learning habits. A detailed list of the highest and lowest scored questions is shown in the following table.

Table 3: Higher and lower scored questions for student's learning attitude

	Ranking of the higher scores	Ave. score	Std. Dev.
With the S-P diagnostic comments			
1	I will pay more attention in answering questions to avoid careless mistake	3.44	.614
2	I will try seriously to answer the question even I don't know how to make the solution initially	3.32	.606
3	I will be more serious in reviewing what I missed in the test	3.20	.752
4	I understand more about what I don't understand	3.15	.771
5	I will be more serious in double checking the answers if it still time left	3.11	.893
6	I will concentrate more on listening teacher's lectures	3.05	.720
7	I will be more prepared for the next test	3.05	.763
	Ranking of the lower scores	Ave. score	Std. Dev.
With the S-P diagnostic comments,			
1	I will have more intention to review after the class	2.66	.858
2	I will change my regular learning habits	2.77	.928
3	I will pay more attention to prepare further learning in advance	2.85	.838
4	I will change my study methods	2.86	.918
5	I will work harder	2.91	.839

Further analysing those high and low scored questions, we found interesting patterns within the data. That is, students displayed a higher concern on testing results (questions ranked number 1, 2, 3, and 5 for the higher scores); nevertheless, it also revealed that learning motivation, learning habits, learning preparation, and study methods were less concerned (questions ranked number 1, 2, 3, and 4 for the lower scores). In summary, the online S-P diagnostic comments encouraged students to review what were missed after the test, and also reminded students to make more efforts and to be more careful during the test. Nevertheless, the online S-P diagnostic comments did not help students much for preparing the test. Moreover, such patterns could mean that the S-P comment table only helps students improve their testing attitude and testing skills, but has no effect on improving their learning attitude and learning skills. Further remedial measures are necessary to be taken, especially for those poorly performed students.

5.2 Results regarding student's linking for the S-P diagnostic comments

The average score of this category is 2.85. This score is also higher than the norm. However, such an average score is lower than that of student's learning attitude. It is worth mentioning that in Taiwan, for more than 50 year, the quantitative methods, such as 100-point-scaling and the ordinal ranking, have dominated over any other method in reporting the test results. Therefore, unfamiliar with the functions of the online S-P diagnostic table and meaning of the qualitative comments may have been the major causes for the lower average score of this category.

By further analysing individual questions, we found that most favourite reason for linking the S-P diagnostic comments is that "it make students pay more attention on reading the test result", followed by "it makes students to know what category of my learning effort to which they belong".. Oppositely, the items: "it represents my true learning behaviour", "it will produce more pressure for my further learning", and "it will allow me to keep my learning records" seem to be less attracted by the students.

Reasons for the lower scores of these questions are to be clarified. A detailed list of the highest and lowest scored questions is shown in the following table.

In addition, we originally expect that there will be a subject matter effect on students' liking for the S-P diagnostic comments. We hypnotized that students would prefer reading over math. However, a non-significant result was reported by the t-test (t=.34, p=.74). This means students' liking for the S-P diagnostic comments is the same for reading and math.

Table 4: Higher and lower scored questions for student's linking of the S-P comments

	Ranking of the higher scores	Ave. score	Std. Dev.
	I like the S-P diagnostic comments because		
1	it will make me pay more attention on reading the test result	3.31	.813
2	it will make me know what category of my learning effort to which I belong	3.21	.864
3	it will make me have a clearer picture of my learning condition	3.18	.811
4	it will help me identify my learning weakness	3.14	.808
5	it will allow me use the S-P table to compare with other classmates' performance	3.08	.709
6	it will be easier to show my learning performance than numeric scores	3.07	.841
	Ranking of the lower scores	Ave. score	Std. Dev.
	I like the S-P diagnostic comments because		
1	it represents my true learning behaviour	2.01	.831
2	it will produce more pressure for my further learning	2.22	.889
3	it will allow me to keep my learning records	2.29	.874
4	it will avoid presenting the ordinal ranking	2.46	.897

Interesting patterns could also be found with the scores of questions regarding student's fondness of the S-P comments. We observed that students in favour of the S-P comments mainly because the major concern of the students was the test result. This could be unveiled by examining the questions ranked number 1, 2, and 5 for the higher scores. Furthermore, we also observed that students concerned about their learning performance (question ranked number 3 and 4 for the higher scores), and would like to compare their own performance with other students. However, students seemed to avoid recognising that the S-P comments truly reflected their learning behaviour as well as presenting the ordinal ranking (question number 1 and 4 for the lower scores). Some students may hesitate to recognise that inadequate learning behaviour had resulted in their poor learning performance.

5.3 Results regarding teacher's opinion

Four out of six teachers attended the experiment were interviewed within one week after the experiment. Email was the main platform for the interviewing process. Additional interviews over the phone were performed as necessary. Interviewing results regarding teaching effectiveness, testing efficiency, learning effectiveness, and appraisals of the online system were discussed as follows.

5.3.1 Testing efficiency

Three out of four teachers agree that time needed for correcting student's answers was shortened in compared with the traditional paper-pencil test. All teachers responded that students finished the test within 20 minutes. They agreed that the major reason for the increased testing efficiency was student's anxiety of obtaining immediate feedbacks from the S-P comment table. Another possible reason for the shortened testing time could be the novelty effects as students were experiencing computerized testing for the first time. This means that the S-P system may be more efficient than traditional paper and pencil test in the beginning, but more evidences should be collected to verify a sustained testing efficiency.

5.3.2 Learning effectiveness

Three out of four teachers asserted that learning could be more effective because the S-P diagnostic comment table depicts individual student's learning condition. In Taiwan, student's learning achievement is normally ranked by 100-point scaling without any meaningful explanation. This qualitative type of information provided more descriptive suggestions that numeric scores could not present. For example, the comment: "mediocre student mastery of the content and occasional wrong responses caused by carelessness" provided much more information than that a numeric score or a ranking list could present.

In general, students welcomed the S-P comment table since it provided more meaningful explanation about the test. Moreover, the immediate feedbacks from the S-P table allowed them to quickly recall how the questions were asked and answered.

5.3.3 Teaching effectiveness

One of the major advantages of the online S-P table reported by teachers is that they were able to use the table to consistently monitor each individual student's learning and testing performance. Facilitated by the computer, teachers did not have to correct test answers and calculate test results by hand. All test records and the S-P comment tables were created automatically and the feedbacks could be given immediately after the test. This could save a tremendous amount of time. Furthermore, the information embedded within the table helped teachers to manipulate both the difficulty level and the amount of instructional contents. For example, one teacher remarked that dividing students into six learning dimensions greatly helped him manage the class schedule; another teacher indicated that the S-P table helped her arrange remediation materials. However, one teacher claimed that although the S-P table facilitated teachers to identify individual student's learning performance, the S-P table could not speed up the instructional process

5.3.4 The system appraisal

All teachers agree that the system functioned properly throughout the experiment and the online S-P diagnostic comment system was able to provide the test results and comments immediately and accurately after the test. Without this system, teachers needed to manually correct each student's answer and calculate the total score. Normally, a minimum of one-day delay is required for reporting a hand-calculate test result.

However, all teachers complained that instable Internet connection and the unfamiliarity with the system operation caused problem. Students began to lose their patient if the instable Internet connection continued over ten minutes. "Technology uncertainty" has been an uncontrollable variable for our experiment.

6. Conclusion

In Taiwan, numeric scoring and ranking have been dominant means for reporting student's learning outcomes. A student can only compare his/her score of with classmates. Such quantitative data could be meaningless for planning further remediation. The S-P comment table, which is able to provide more qualitative descriptions of the learning and testing performances, seems to be a feasible alternative for solving the problem. However, complicated computations and sorting process of the S-P algorithm have made most of teachers hesitate to adopt it to their classroom teachings.

This study, however, constructed an online test platform with an S-P diagnostic table generator that is able to report qualitative comments immediately after the test. With this computer-supported system, teachers are no need to make complicated hand calculations. An experiment was conducted to answer the research questions: 1) How online comments provided by the S-P diagnostic table could affect the students' learning attitude? 2) Does subject matter affect student attitude toward the S-P online diagnostic comments? and 3) What are the teacher appreciations of the S-P online diagnostic comments? The results indicate that positive feedbacks from both students and teachers were obtained. Students preferred the online qualitative comments provided by the S-P diagnostic table over traditional numerical scoring. This is because they were able to understand why they performed well/poor on the test. The results also show that the online S-P diagnostic table caused students to become more circumspect in answering test questions to reduce careless mistakes. Students were also more likely to review what was missed on the test.

Moreover, the results of the study also indicated that the S-P comment table only helped students improve their testing attitude and testing skills, but had no effect on improving their learning attitude and learning skills. An online iterative drilling platform was built consequently to incorporate with the S-P diagnostic process to better assist poorly performed students to make progress on their performance. Variable Interval Performance Queuing (VIP) methods (Alessi & Trollip, 2001) were employed for the repetitive drilling process. A preliminary experiment was done to examine the remedial effects of the online drilling system. The results showed that the online practice-and-drill group had a significant gain for the pretest-posttest scores, in compare with no gains for the traditional classroom teaching group (Li and Wang, 2012), indicating that the online iterative drilling platform may effectively work with the S-P diagnostic process to provide constructive remediation for the students who exhibited a poor performance on the S-P chart.

Acknowledgements

Funding of this research work is supported in part by the National Science Council of Taiwan (under the contract numbers NSC 97-2631-S-003 -004 -CC3 and NSC 99-2631-S-003 -014 -CC3).

References

Al-A'ali, M. (2007) “Implementation of an improved adaptive testing theory”, *Educational Technology & Society*, Vol 10, No. 4, pp 80-94.

Alessi, S. M, & Trollip, S. R. (2001) “*Multimedia for learning, methods and development. 3rd ed*. Allyn and Bacon, MA. U.S.A., pp 190-195.

Chacko,I. (1988) “S-p chart and instructional decisions in the classroom”, *International journal of mathematical education in science & technology*. Vol 29, No. 3, pp 445-450.

Chang, W. C., Hsu, T. K., Lee, T. P., & Huang, Y. C. (2009) "The learning analysis system integrate with s-p chart and rule-space model", *2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC)*, Kaohsiung Taiwan 7-9 December 2009, pp 1027-1030. Kaohsiung, National Kaohsiung University of Applied Science.

Chang, W. C., Yang, H. C., Shin, T. K., & Li, M. F. (2008) “Integrating IRT to estimate learning ability with S-P chart in web based learning environment”, *Lecture Notes in Computer Science*, Vol 5328, pp 133-145.

Chen, D. J., Lai, A. F., Liu, I. C., Chen. D. T. (2005) “The design and implementation of a diagnostic test system based on the enhanced S-P model”, Journal of Information Science and Engineering, Vol 21, pp 1007-1030.

Chen, T. S. (1989) “An empirical study on counselling and S-P chart analysis”, *The archive of Guidance & Counselling*. Vol 12, pp 85-141.

Frick, T., Luk, H.K., & Tyan, N. C. (1987) “*A comparison of three adaptive decision making methodologies used in computer based instruction and testing*”, Bloomington, IN: Final Report, Proffitt Foundation. Indiana Univeristy.

Harnisch,D.L. (1987) “*A Multilevel evaluation system using student-problem charts*”, [S.l.]: Distributed by ERIC Clearinghouse.

Hu, M. S., Chen, M. C., & Yang, Z. S. (2005) “The development of a computerized cognition diagnose and assessment system”, *Journal of Air Force Institute of Technology*, Vol 4, 118-129.

Li, S. X., & Wang, C. H. “Employing online practice-and-drill for math remediation in junior high schools” *Taiwan E-learning Forum 2012 (TWELF)* Tainan Taiwan 26-27 October 2012, S-6-5 (CD). Tainan, Chia Nan University of Pharmacy & Science.

Sato, T. (1985) *Introduction to S-P Curve Theory Analysis and Evaluation*, Meiji Tosho, Tokyo, Japan. (in Japanese)

Wang, B. T., Sheu, T. W., Liang, J. C., Tzeng, J. W., & Masatake, N. (2012) “Using grey S-P chart to evaluate english reading performances”, 2*012 International Conference on Education and Management Innovation IPEDR* Kuala Lumpur Malaysia 5-6 May 2012, Vol. 30, pp 140-144. Hong Kong, International Economics Development and Research Center.

Wang, C. H, & Chen, C. P. (2006) “Building an Online Distance Adaptive Testing System for Assessing Unit Learning Achievement”. In: Cheng, J. g. & Han, X, B., (Eds.), *The Proceedings fo the 10th Annual Global Chinese Conference on Computers in Education*, Beijing: Tsinghua University Press, pp 176-184.

Wang, A., & Chuang, C. (2002) “Applicable adaptive testing models for school teachers”, *Educational Media International*, Vol 39, No. 1, pp 55-59.

Wainer, H. & Kiely, G. L. (1987) “Item clusters and computerized adaptive testing: a case for testlets”, *Journal of Educational Measurement,* Vol *24*, No. 3, pp 185-201.

Wu, H. Y. (1998) “Software based S-P Chart Analysis and its applications”, *Proceedings of the National Science Council, Republic of China*, Vol 8, No. 3, pp 108-120.

Xu, X. L., & Douglas, J. (2006) “Computerized adaptive testing under nonparametric IRT models”, *Psychometrika*, Vol 71, No.3, pp 121-137.

Yih, J. M. & Lin, Y. H. (2010) “Concept structure based on response pattern detection of S-P Chart with application in algebra learning”, WSEAS Transactions on Computers, Vol 9, No. 8, pp 847-856.

Yu, M. N. (2002) *Educational Measurement*. Xu-Nan, Taipei.

Yu, M.N. (2003) Educational *Test and Assessment-The Achievement Measurement and the Instructional Assessment (2nd ed.)*. Psychological publishing, Taipei.

Lin, Y. H., Yih, J. M. (2008) "An Integration of Concept Structure Analysis and S-P Chart with Application in Equality Axiom Concepts Diagnosis," *IITA '08 Second International Symposium on Intelligent Information Technology Application*, Vol.2, pp 468-472.

Zhan, Y. W. (2006) “*Response Theory (IRT) and Student-Problem (S-P) Chart — using multimedia English test questions as example*”, Unpublished Master’s thesis, National Chiao Tung University.