*International Journal of*
*Innovative Management, Information & Production* ISME International©2011 ISSN 2185-5439
Volume **2**, Number **1**, June **2011** PP. **87-101**

# COMPARISONS OF RULE USAGE IN PROBABILITY REASONING BASE ON OT AND ISM

YUAN HORNG LIN[1] AND BERLIN WU[2]
[1]Department of Mathematics Education
National Taichung University
Taichung City, Taiwan
lyh@mail.ntcu.edu.tw

[2]Department of Applied Mathematical Sciences
National Chengchi University
Taipei City, Taiwan
berlin@nccu.edu

ABSTRACT. *The purpose of this study is to implement an Internet assessment of probability reasoning that provides graphs of rule usage and similarity coefficients compared to expert reasoning. The probability reasoning assessment is based on the rule-assessment approach provided by R. S. Siegler. The method of constructing the structural graphs of rule usage in probability reasoning combines ordering theory (OT) with interpretive structural modeling (ISM). Additionally, set operations are adopted to calculate the similarity coefficient for graphs of rule usage compared with expert reasoning. Several programming languages, including PHP, MySQL and FLASH, are used to implement the assessment system. An empirical study of pupils shows that rule usage varied with the total score and response patterns. In addition, there is a significant difference in the similarity coefficients based on the variables of age and gender and their interaction. Finally, some recommendations and suggestions for future research are discussed based on the findings and results.*

**Keywords:** Internet Assessment System; Ordering Theory; Probability Reasoning; Rule Assessment

**1. Introduction.** Probability reasoning is one of the logic thinking and many psychological development researchers acknowledge the importance of probabilistic reasoning in cognitive development (Piaget and Inhelder, 1975). Probabilistic reasoning can develop quite early in childhood and children display misconceptions stemming from misinterpretation of the reasoning about uncertain events (Konold, 1989). Siegler indicates that there are three defective rules and one correct rule as to the problem solving rules for marble test, which is a popular probability reasoning test (Siegler, 1981). Although probability reasoning is an important issue, most researchers adopt paper-pencil test and analyze performance of examinee based on total score. Results of total score provide limited information on cognition diagnosis and little is known about the knowledge structures of probability reasoning. In addition, paper-pencil test is quite time-consuming.

Therefore, development of Internet assessment system for probability reasoning with diagnostic information on rule usage should be a prospective research.

Based on the discussions above, Internet assessment system of probability reasoning test which is extended from research of Siegler will be implemented in this study. Moreover, ordering theory (OT) combined with interpretive structural modeling (ISM) is used to calculate the subordinate relationship among rules so that individualized rule usage could be displayed in the form of graphic structure. Besides, similarity coefficient for graphs of rule usage between examinee and expert will also be developed to present and compare characteristics of individualized rule usage (Hartigan, 1967). With the Internet assessment, an empirical study for pupils will be investigated and discussed.

**2**. **Literature Review.** Literature will include probability reasoning, OT and ISM. Related research about development and assessment methodology of probability reasoning is discussed. Rules of problem-solving for probability reasoning marble test will also be discussed. OT and ISM are introduced in view of their algorithm and application.

**2.1. Probability Reasoning and Rule Usage of Marble.** Piaget and Inhelder determine that adolescents in the period of formal operations are capable of manifesting probabilistic reasoning; whereas, individuals in lower cognitive periods are not. As a result, they view probabilistic reasoning to be a characteristic of formal operations. Quite a few researches aim at probability reasoning tasks and they find proficiency of probability reasoning may vary with age and gender (Siegler, 1981; Cosmides and Tooby, 1996).

Fischbein provides the foundation of intuitive thinking as precursors to development of probability reasoning (Fischbein and Gazit, 1984). He claims that there are two kinds of intuitions, which influence the development of probability reasoning. One is primary intuitions that are related to personal experiences and appear prior to instruction. The other is secondary intuitions that appear by way of the instructional process. The results show that intuition can support surprisingly precocious performance in young children and contribute to the biases evident in adult judgments and decisions. One common finding claims that the understanding of ratio and proportion constitutes a precondition to fine probabilistic reasoning (Tarr and Jones, 1997). However, it is also found that many students have misconceptions regarding chance, randomness, and probability (DelMas and Bart, 1989). Konold provides a set of questions, called “Weather Problem”, and indicates that subjects display misconceptions stemming from misinterpretation of the reasoning about uncertain events (Konold, 1989). Some researchers employ instructional activities and demonstrate that refined instruction could promote students’ probability reasoning (Tarr and Jones, 1997; Garfield and Ahlgren, 1988).

One well-known probability reasoning is called marble test (Siegler, 1981). Figure 1 is an example of marble test design and it is represented as (3, 4) vs. (2, 3). It means that there are 3 black marbles and 4 white marbles in set A and 2 black marble and 3 white marbles in set B.

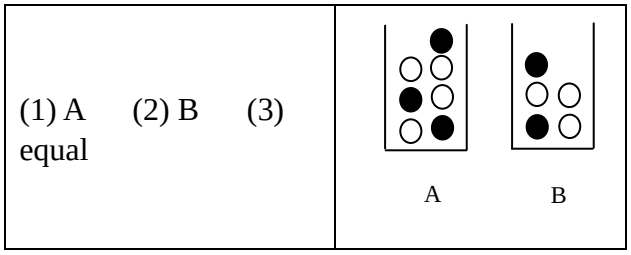


FIGURE 1. Example of the item of marble test

Students are asked to imagine picking one marble randomly from the two sets and requested to decide which set provides greater chance of picking a black marble, set A or set B or "equal." If these two sets have the same chance, subjects must choose "equal.". Siegler (1981) indicates that there are three defective rules and one correct rule when students respond to marble test items. These four rules are depicted in Figure 2, 3, 4 and 5. Rule 4 is the correct rule and the other rules are defective rules. The following conventions are used in these figures: (a) RB = number of black marbles on the right side (set B); (b) RW = number of white marbles on the right side (set B); (c) LB = number of black marbles on the left side (set A); and (d) LW = number of white marbles on the left side (set A).

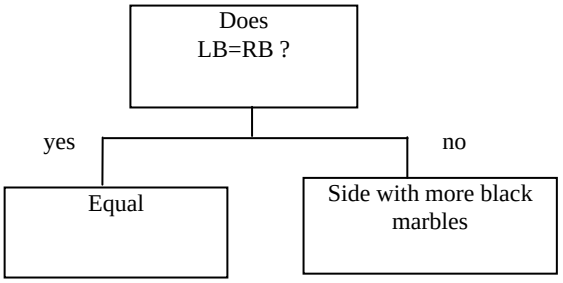


FIGURE 2. Flowchart of rule 1

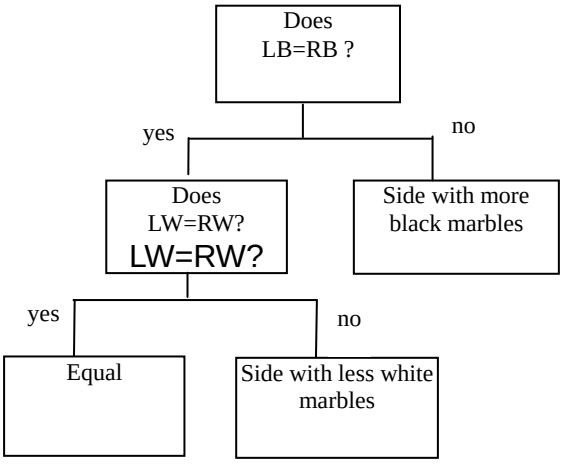


FIGURE 3. Flowchart of rule 2

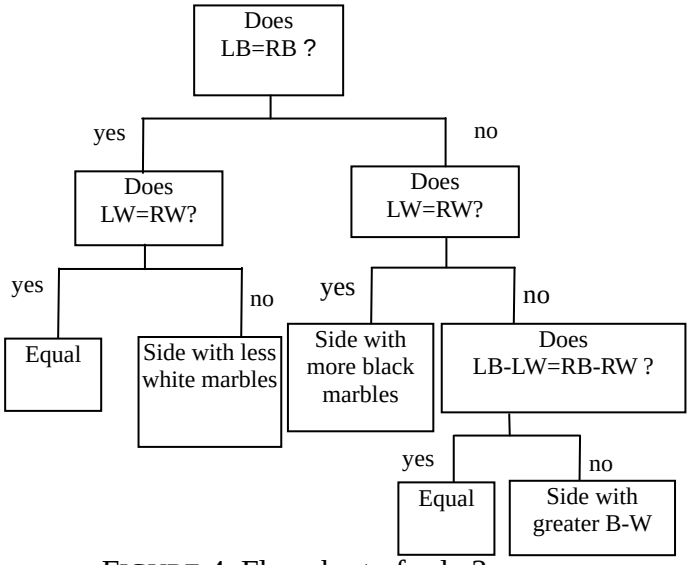


FIGURE 4. Flowchart of rule 3

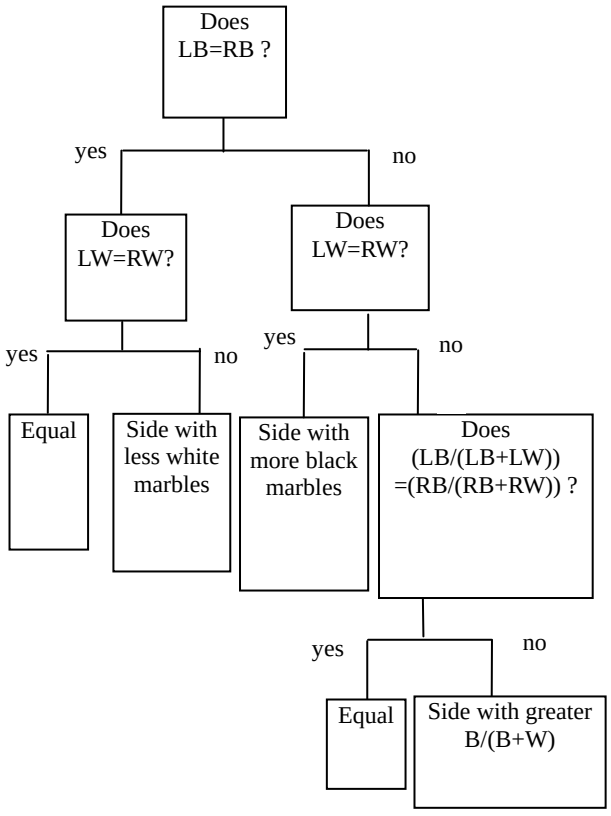


FIGURE 5. Flowchart of rule 4

Latent class analysis and other clustering technique are popular methods to classify students so that students within the same group own homogeneity in rule usage and knowledge structures. Research also find that students do not use unique a rule when responding to marble test (Lin and Hung, 2007). In other words, students may change rule

in the process of answering questions. Therefore, except for the classification on students for rule usage, how to construct the hierarchies and relationship about rule usage of probability reasoning should be a prospective study.

As to rules of problem-solving, some are correct and some are incorrect. There are two kinds of scoring. One is raw score and the other is raw rule score. Raw score is applied in most traditional probability reasoning research. This scoring simply focuses on the correct rule. It is coded as 1 when the response conforms with the correct rule. Otherwise, it is coded as 0. Almost traditional research adopts raw score and it just focuses on scoring of correct rule without any information from incorrect rules. However, raw rule score depends on the information of all rules, including correct and incorrect rules. The response vector is coded by the conformity of each rule and the information of raw rule score covers that of raw score.

**2.2. Ordering Theory.** It is the common viewpoint of psychometrics that items within a test exist subordinate relationship (Bart and Krus, 1973). OT is one branch of psychometric methodology to organize item hierarchies and its purpose is mainly to determinate the ordering relationship and precondition among items. With the analysis of OT, item hierarchies could be displayed.

Suppose there be two dichotomous items item $i$ and item $j$ ($i \neq j$). It is coded 1 when examinees give right answer; otherwise it is coded as 0. The frequencies of examinee within each cell are showed in Table 1.

TABLE 1. Contingency table of response frequency for item

| | | Item $j$ | | Sum |
|---|---|---|---|---|
| | | 1 | 0 | |
| Item $i$ | 1 | $n_{11}$ | $n_{10}$ | $n_{1\bullet}$ |
| | 0 | $n_{01}$ | $n_{00}$ | $n_{0\bullet}$ |
| Sum | | $n_{\bullet 1}$ | $n_{\bullet 0}$ | $n = n_{11} + n_{01} + n_{10} + n_{00}$ |

The response pattern (0, 1) means disconfirmatory pattern because this response pattern doesn't satisfy the condition that item $i$ is a precondition of item $j$ (Bart et al., 1997). Hence, $r_{ij} = n_{01}/n$ is defined as ordering coefficient and it is the percentage of disconfirmatory pattern. It is obvious that $0 \le (n_{01}/n) \le 1$ and smaller value means more possibility that item $i$ is the precondition of item $j$. Tolerance level $\varepsilon$ ($0 < \varepsilon < 1$) is to decide whether the binary precondition relationship between two items exists or not. It is

$$r_{ij}^{*} = \begin{cases} 1 & , \quad r_{ij} < \varepsilon \\ 0 & , \quad r_{ij} \ge \varepsilon \end{cases} \tag{1}$$

If $r_{ij}^{*} = 1$ exists, it means item $i$ is the precondition of item $j$ with linkage from item $i$ to item $j$. On the other hand, there is no linkage from item $i$ to item $j$. Some research suggest that $\varepsilon$ should be no greater than 0.2. Namely, it had better $\varepsilon < 0.2$. With the construction of hierarchies and relationship among items, item hierarchies based on OT will be implemented.

OT is mainly used in the investigation of intelligence development, psychological development of formal thinking and knowledge structures of mathematics ability (Bart and Williams-Morris, 1990; Bart et al., 1994; Bart et al., 1997). In this study, the researcher will adopt OT to calculate the ordering coefficient for subordinate relationship of rule usage.

**2.3. Interpretive Structural Modeling.** ISM, developed by J. N. Warfield, it is based upon discrete mathematics and graph theory and it aims to arrange elements of binary relationship by hierarchical graph (Warfield, 1977). For a system containing $K$ elements, their binary relationship among elements is known and it is denoted by binary relation matrix $A=(a_{ij})_{K\times K}$. $a_{ij}=1$ means $A_i$ is the precondition of $A_j$. Otherwise, means $a_{ij}=0$ presents $A_i$ is not the precondition of $A_j$.

Boolean operation is used to acquire transitive closure and reachability matrix in ISM algorithm Warfield(1982). The transitive closure is $\hat{A}=A\oplus A^2\oplus A^3\oplus\cdots A^P$ and reachability matrix is $R=\hat{A}\oplus I=(A\oplus I)^P$. With transitive closure $\hat{A}$ and reachability matrix $R$, the hierarchical graph with regard to binary relation matrix $A=(a_{ij})_{K\times K}$ is derived. To take the binary relation matrix $A=\left(a_{ij}\right)_{K\times K}$ in Figure 6 for example, the construction of simplified graph by ISM is depicted.

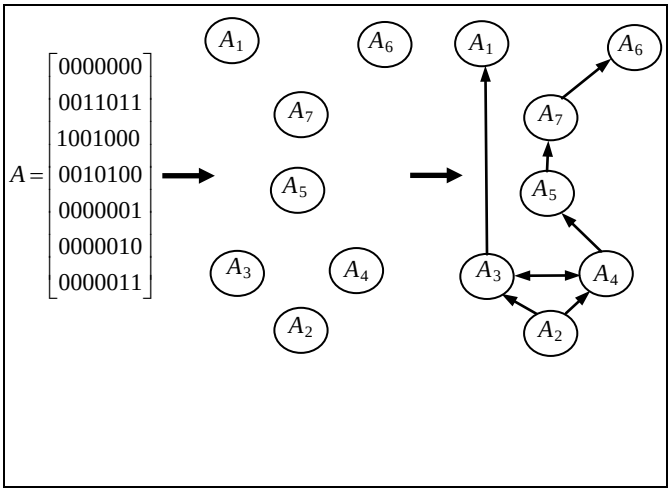


FIGURE 6. Construction of hierarchical graph by ISM

**3. Research Design and Data Resource.** The researcher designs Internet system of probability reasoning assessment. Formula for precondition relationship calculation, probability reasoning items, Internet assessment system design and sample will be discussed.

**3.1. Ordering Calculation for Rule Usage.** Suppose there be $M$ ($m=1,2,\cdots,M$) items and $R$ ($r=1,2,\cdots,R$) rules in a test. There is $N$ ($n=1,2,\cdots,N$) students who take the test. For student $n$, if his response on item $m$ conforms with rule $r$, raw rule score for item $m$ on rule $r$ is denoted by $s_{mr}=1$; otherwise it is $s_{mr}=0$. The raw rule score matrix for student $n$ is $S_n=(s_{mr})_{M\times R}$. Contingency table based on rules of raw rule score matrix $S_n=(s_{mr})_{M\times R}$ is depicted in Table 2.

TABLE 2. Contingency table of frequency on raw rule score pattern for two rules

| | | Rule $r'$ | | Sum |
|---|---|---|---|---|
| | | 1 | 0 | |
| Rule $r$ | 1 | $f_{11}$ | $f_{10}$ | $f_{1\bullet}$ |
| | 0 | $f_{01}$ | $f_{00}$ | $f_{0\bullet}$ |
| Sum | | $f_{\bullet 1}$ | $f_{\bullet 0}$ | $f$ |

OT is used to determine the subordinate relation between rule $r$ and rule $r'$. The ordering coefficient is $f_{rr'} = f_{01}/f$ and its binary subordinate relationship is

$$f_{rr'}^{*} = \begin{cases} 1 & , \quad f_{rr'} < \varepsilon \\ 0 & , \quad f_{rr'} \geq \varepsilon \end{cases} \tag{2}$$

Where $\varepsilon$ is tolerance level and it is $0 < \varepsilon < 1$. Similarly, it is suggested that $\varepsilon < 0.2$. The matrix of binary subordinate relationship among rules for student $n$ is $F_n^{*} = (f_{rr'}^{*})_{R\times R}$. $F_n^{*} = (f_{rr'}^{*})_{R\times R}$ provides the source matrix to construct graphs of rule usage and the algorithm of graph construction is ISM, which arrange rules in the form of hierarchical and relational graph.

**3.2. Similarity Calculation for Graph of Rule Usage.** Suppose there be two students $n$ and $n'$ with their matrix of binary subordinate relationship $F_n^{*}$ and $F_{n'}^{*}$ respectively. The similarity coefficient for graphs of rule usage between these two students is $s_{nn'}$ and it is

$$s_{nn'} = (\frac{1}{R})\sum_{r=1}^{R} \frac{\#\left(G_n(\nu_r) \cap G_{n'}(\nu_r)\right)}{\#\left(G_n(\nu_r) \cup G_{n'}(\nu_r)\right)} \tag{3}$$

$G_n(\nu_r)$ is defined as $G_n(\nu_r) = \{\nu_r \,|\, f_{rr'}^{*} = 1\}$ and it means the set of rules which rule $r$ is the precondition of them for student $n$. $\#(G_n(\nu_r) \cap G_{n'}(\nu_r))$ is the number of rules belonging to the intersection of $G_n(\nu_r)$ and $G_{n'}(\nu_r)$; while $\#(G_n(\nu_r) \cup G_{n'}(\nu_r))$ is the number of rules belonging to the union of $G_n(\nu_r)$ and $G_{n'}(\nu_r)$. The larger the similarity coefficient $s_{nn'}$ is, the more similar in graphs of rule usage these two students will have.

Expert have more advanced and refined knowledge structures Goldsmith et al.(1991). Suppose expert use correct rule through all items. The matrix of binary subordinate relationship among rules for expert is $F_{expert}^{*} = (f_{rr'}^{*})_{R\times R}$. The similarity coefficient for graphs of rule usage between student $n$ and expert is defined as follows. As previously stated, a larger similarity coefficient $s_{n(expert)}$ implies a better rule usage.

$$s_{n(expert)} = (\frac{1}{R})\sum_{r=1}^{R} \frac{\#\left(G_n(\nu_r) \cap G_{expert}(\nu_r)\right)}{\#\left(G_n(\nu_r) \cup G_{expert}(\nu_r)\right)} \tag{4}$$

**3.3. Probability Reasoning Items and Internet Assessment System.** Internet assessment system contains modules of data management, assessment management and environment management. Platform and software are CentOS 4.5, Apache 2.0, PHP and MySQL. Operation procedure of the system is depicted in Figure 7.

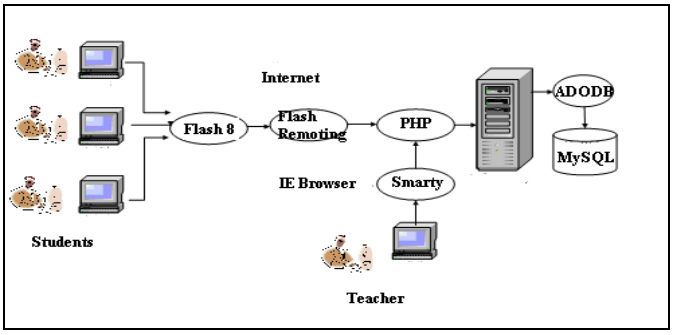


FIGURE 7. Operation procedure of Internet assessment system

The probability reasoning test includes 20 marble problems. As shown in Figure 8, students are asked to imagine picking one marble randomly from the two cups respectively. They need to decide which cup provides greater chance of picking a black marble, cup A or cup B or "equal." If these two cups have the same chance of picking a black marble, they must choose "equal." Composition of item design and response based on all rules are shown in Table 3.

TABLE 3. Items of composition and response based on four rules

| Item No. | Composition | Response | | | |
|---|---|---|---|---|---|
| | | Rule 1 | Rule 2 | Rule 3 | Rule 4 |
| 1 | (1,2) vs. (2,4) | B | B | A | E |
| 2 | (1,2) vs. (3,6) | B | B | A | E |
| 3 | (1,2) vs. (4,8) | B | B | A | E |
| 4 | (1,3) vs. (2,6) | B | B | A | E |
| 5 | (2,1) vs.(3,2) | B | B | E | A |
| 6 | (2,4) vs.(1,2) | A | A | B | E |
| 7 | (2,6) vs.(1,3) | A | A | B | E |
| 8 | (2,8) vs.(1,4) | A | A | B | E |
| 9 | (3,1) vs.(5,3) | B | B | E | A |
| 10 | (3,2) vs.(8,7) | B | B | E | A |
| 11 | (3,6) vs. (1,2) | A | A | B | E |
| 12 | (3,7) vs.(3,2) | E | B | B | B |
| 13 | (4,1) vs.(4,6) | E | A | A | A |
| 14 | (4,1) vs.(9,6) | B | B | E | A |
| 15 | (4,6) vs.(4,1) | E | B | B | B |
| 16 | (5,1) vs. (5,7) | E | A | A | A |
| 17 | (5,3) vs. (3,1) | A | A | E | B |
| 18 | (5,4)vs. (2,1) | A | A | E | B |
| 19 | (7,5) vs. (3,1) | A | A | E | B |
| 20 | (7,5) vs. (4,2) | A | A | E | B |

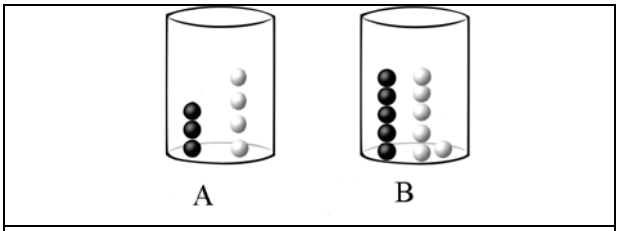


FIGURE 8. Example of item (3, 4) vs. (5, 6)

**3.4. Subjects.** The subjects of this study include 3339 students from fifth to eighth graders. They come from Taichung City and Taichung County of Taiwan. All the subjects take the Internet probability reasoning assessment at their computer classroom. Before they begin to take probability reasoning assessment, their computer teachers spend 10 minutes to explain the purpose and the details.

Limitation of testing time is 30 minutes. Subjects must finish the test by themselves without any discussion. Once subjects finish the test and submit their answer, all their response data will be transmitted to the server system immediately. The researcher obtains raw response, graphs of rule usage and similarity coefficient instantly.

TABLE 4. Subjects with grade and gender

| Grade | Gender | | Total |
|---|---|---|---|
| | Male | Female | |
| 5 | 438 | 395 | 833 |
| 6 | 371 | 339 | 710 |
| 7 | 543 | 535 | 1078 |
| 8 | 373 | 345 | 718 |
| Total | 1725 | 1614 | 3339 |

**4. Results.** Two way analysis of variance (two way ANOVA) based on grade and gender will display whether there are differences respective to correct ration and similarity coefficient. Furthermore, graphs of rule usage according to different total score and response pattern will show the characteristics of rule usage.

**4.1. Two Way ANOVA for Correct Ration of Test.** As mentioned before, probability reasoning may vary with age and gender. Hence, two way ANOVA based on grade and gender for correct ration of probability reasoning is shown in Table 5. There is significant difference on grade. However, there is no significant difference for neither gender nor interaction of them.

TABLE 5. Two way ANOVA on grade and gender for correct ration

| Source | SS | df | MS | F |
|---|---|---|---|---|
| Grade | 42.76 | 3 | 14.25 | 203.57*** |
| Gender | .01 | 1 | .01 | .14 |
| Grade × Gender | .51 | 3 | .17 | 2.42 |
| Error | 245.89 | 3331 | .07 | |
| Total | 289.17 | 3338 | | |

***p<.001

Table 6 indicates post hoc comparison for grade and there exist significant differences between some pairs of comparison. Both seventh and eighth graders perform better than fifth and sixth graders respectively.

TABLE 6. Post hoc comparison on grade for correct ratio

| | Grade | | | |
|---|---|---|---|---|
| | 5 | 6 | 7 | 8 |
| 5 ( $\bar{X}_5$ =.4152 ) | | | | |
| 6 ( $\bar{X}_6$ =.4196) | .0044 | | | |
| 7 ( $\bar{X}_7$ =.6314) | .2162*** | .2118*** | | |
| 8 ( $\bar{X}_8$ =.6608) | .2456*** | .2412*** | .0294 | |

***p<.001

**4.2. Rule Usage of Different Total Score**. Although method in this study could display individualized graph of rule usage, however, it is too lengthy to exhibit each individualized graph. For convenience, subjects are divided into three groups in accordance with total score so that characteristics of rule usage vary with total score are easily compared. Subjects within the highest 27% total score belong to high score group. On the contrary, subjects within the lowest 27% total score belong to low score group. The others belong to middle total score group. Three students are randomly selected from the above three groups respectively. Tolerance level $\varepsilon$ is decided $\varepsilon = 0.1$.

As shown in Figure 9, student A, B, C, won distinct graph of rule usage. Their similarity coefficients compared with expert are 1, .71 and .33 respectively. As to student A, who belong to high score group and has total score 18, he adopts rule 4 in advance. Once he gives up rule 4, he will use rule 2 or rule 3. Student B belong to middle score group and has total score 11. Student B will adopt rule 4 or rule 2 firstly. However, rule 2 is not correct rule and this strategy will result in wrong answer. Student C belongs to low score group and has total score 2. He will adopt rule 1, rule 2 or rule 3 in advance. The above three rules are defective rules and this is the reason why student C belongs to low score group.

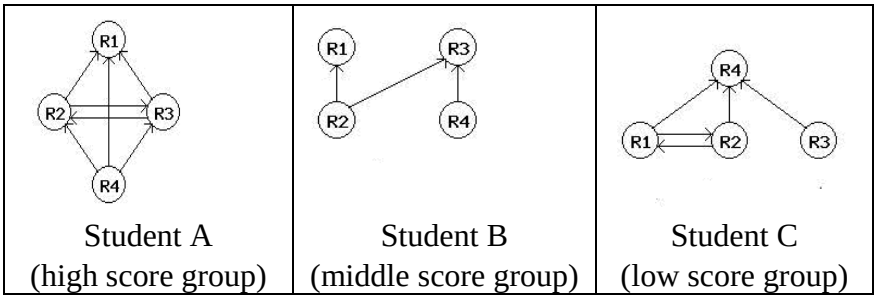


FIGURE 9. Graphs of rule usage for three students of different groups

**4.3. Rule Usage of Different Response Pattern with Same Total Score.** There is limitation of cognition information on total score and response pattern will provide more information for rule usage. Two pairs of students with the same total score and different response pattern will be discussed below.

As shown in Figure 10, the first pair of student D and student D' , whose similarity coefficients compared with expert are .71 and .85, have the same total score 11 with varied response pattern. Their response patterns are as follows.

Response pattern of D is : (11011000010110110110)

Response pattern of D' is : (00001000100111111111)

As predicted above, their graphs of rule usage are quite different.

Similarly, the second pair of student E and student E', whose similarity coefficients compared with expert are .33 and .44, have the same total score 2 with different response pattern. Their response patterns are as follows.

Response pattern of E is: (10000000000000000010)

Response pattern of E' is: (00000000000001000001)

Their graphs of rule usage vary a lot. The above results in this section coincide with the viewpoints and findings in some cognition diagnosis literature.

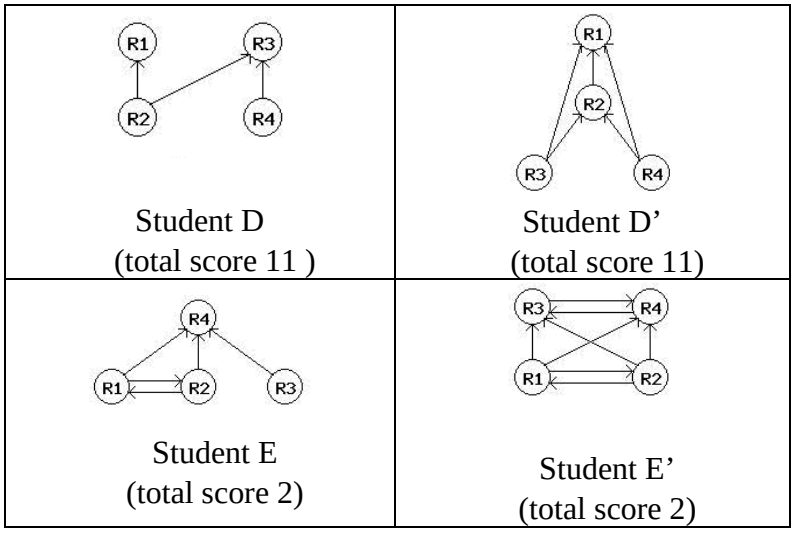


FIGURE 10. Comparisons on graphs of rule usage for two pairs of students

**4.4. Comparisons on Similarity Coefficient.** As already explained, expert will use correct rule (rule 4) when responding to marble problems. Similarity coefficient compared with expert is to measure the differences on graph of rule usage between student $n$ and expert. The graph of rule usage for expert is depicted in Figure 11.

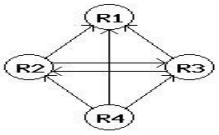


FIGURE 11. Graphs of rule usage for expert

As previously discussed, age and gender are possible two factors to be considered as to variance of rule usage. Therefore, two way ANOVA with independent variables of age and gender and dependent variable of similarity coefficient compared with expert is analyzed. As shown in Table 7, there are significant differences on factors of grade and gender. Besides, there also exists interaction between grade and gender.

TABLE 7. Two way ANOVA on similarity coefficient for grade and gender

| Source | SS | df | MS | F |
|---|---|---|---|---|
| Grade | 16.70 | 3 | 5.57 | 139.25*** |
| Gender | .20 | 1 | .20 | 5.00* |
| Grade × Gender | .41 | 3 | .14 | 3.50* |
| Error | 138.13 | 3331 | .04 | |
| Total | 155.44 | 3338 | | |

*p<.05 ***p<.001

Table 8 and Table 9 display the post hoc comparison for main effect of grade and gender. Table 7 demonstrates that there is significant difference between any pair of comparison except for pair of grade 5 and grade 6. Table 9 displays that significant difference between male and female and similarity coefficient of female is higher than male. Table 10 summarizes the hoc comparison for simple main effect of grade × gender. It reveals that results of post hoc comparison on grade depend on male and female. Moreover, it only occurs to sixth graders that similarity coefficient of female is higher than male.

TABLE 8. Post hoc comparison for main effect of grade

| | Grade | | | |
|---|---|---|---|---|
| | 5 | 6 | 7 | 8 |
| 5 ( $\bar{x}_5$ =.6416 ) | | | | |
| 6 ( $\bar{x}_6$ =.6259) | -.0157 | | | |
| 7 ( $\bar{x}_7$ =.7603) | .1187*** | .1344*** | | |
| 8 ( $\bar{x}_8$ =.7950) | .1534*** | .1691*** | .0347** | |

**p<.01 ***p<.001

TABLE 9. Post hoc comparison for main effect of gender

| | Gender | |
|---|---|---|
| | male | female |
| Male ( $\bar{X}_{male}$ =.7013 ) | | |
| female ( $\bar{X}_{female}$ =.7184) | .0171* | |

*p<.05

TABLE 10. Post hoc comparison for simple main effect of grader×gender

| Source | | SS | df | MS | F test and post hoc comparison |
|---|---|---|---|---|---|
| Gender | male | 9.81 | 3 | 3.27 | F = 78.07*** 8 > 7 > 6 > 5 |
| | female | 7.39 | 3 | 2.46 | F = 60.00*** 7 > 5, 7 > 6, 8 > 5, 8 > 6 |
| Grade | 5 | .01 | 1 | .01 | F = .25 |
| | 6 | .48 | 1 | .48 | F = 11.67** female > male |
| | 7 | .11 | 1 | .11 | F = 2.36 |
| | 8 | .01 | 1 | .01 | F = .03 |

**p<.01 ***p<.001

**5.5. Conclusions.** Probability reasoning Internet assessment system is successfully built based on the algorithm of OT and ISM. This assessment system could display individualized graph of rule usage. The main findings of this study are as follows.

One major result of this study is that the calculation and Internet assessment for graph of rule usage are established. Teachers or researcher could acquire real-time results immediately after students finish their assessment. The graphs of rule usage provide information of hierarchies and relationship among rules. The individualized graph could help teachers understand cognition processing and logic thinking of each student so that they could design adaptive instruction and remedial instruction in accordance with the graphic information.

Secondly, investigation on difference between novice and expert is the state-of-art issue in cognition psychology. Formula of similarity coefficient for graphs of rule usage is developed so that comparisons among graphs of rule usage are feasible.

Thirdly, two independent variables, age and gender, are considered to be independent variables in this study. When the correct ration based on only correct rule is considered, there exits only significant difference on grade. However, when it comes to consider similarity coefficient, the results show there are significant differences on grade, gender and interaction of grade and gender. Generally speaking, there are significant differences among grades and this shows that rule usage development increase rapidly as grade. It also

reveals that female is like expert more than male. Post hoc comparisons for interaction of grade and gender show that divergence on rule usage depends on each other. The foregoing results may indicate that similarity coefficient could furnish more details about cognitive information on rule usage than traditional total scoring.

This study builds on-line assessment with graph of rule usage and similarity coefficient for probability reasoning test which is beyond the limitation of paper-pencil test. Further research could be done on theoretical development of rule assessment methodology. Other cognitive issue on problem-solving rules like proportion reasoning and relation reasoning based on methodology in this study could be a prospective research Siegler (1982) Jansen and Han van der Maas (1997).

## REFERENCES


[1] B. R. J. Jansen and M. L. J. Han van der Maas(1997), Statistical test of the rule assessment methodology, *Developmental Review*, vol. 17, pp. 321-357.

[2] C. Konold(1989), Informal concepts of probability, *Cognition and Instruction*, vol. 6, pp.59-98.

[3] E. Fischbein and A. Gazit(1984), Does the teaching of probability improve probabilistic intuitions?, *Educational Studies in Mathematics*, vol. 15, pp.1-24.

[4] J. N. Warfield(1982), Interpretive structural modeling (ISM), *in Group Planning & Problem Solving Methods in Engineering*, Olsen, S. A. W.E., Walter, W.V., and Lehner, W. (Eds.), *New York: Wiley* , pp. 115-201.

[5] J. A. Hartigan(1967), Representation of similarity matrices by trees, *Journal of the American Statistical Association*, vol. 62, pp.1140-1158.

[6] J. E. Tarr and G. A. Jones(1997), A framework for assessing middle school students' thinking in conditional probability and independence, *Mathematics Education Research Journal*, vol. 9, pp.39-59.

[7] J. Garfield and A. Ahlgren(1988), Difficulties in learning basic concepts in probability and statistics: implications for research, *Journal for Research in Mathematics Education*, vol. 19, pp. 4.

[8] J. N. Warfield(1977), Crossing theory and hierarchy mapping, *IEEE Transactions on System, Man, and Cybernetics*, vol. 7, pp. 505-523.

[9] J. Piaget and B. Inhelder(1975), The Origin of the Idea of Chance in Children, *Routledge and Kegan Paul*.

[10] L. Cosmides and J. Tooby(1996), Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty, *Cognition*, vol. 58, pp.1-73.

[11] R. S. Siegler(1981), Developmental of sequences within and between concepts, *Society for Research in Child Development Monographs*, vol.46, Whole No. 189.

[12] R. DelMas and W. M. Bart(1989), The role of an evaluation exercise in the resolution of misconceptions of probability, *Focus on Learning Problems in Mathematics*, vol. 11, pp. 39-54.

[13] R. S. Siegler(1982), The rule-assessment approach and education,*Contemporary Educational Psychology*, vol.7, pp.272-288.

[14] T. E. Goldsmith, P. J. Johnson and W. H. Acton(1991), Assessing structural knowledge, *Journal of Educational Psychology*, vol. 83, pp.88-96 .

[15] W. M. Bart and D. J. Krus(1973), An ordering-theoretic method to determine hierarchies among items, *Educational and Psychological Measurement*, vol. 33 , pp.291-300.

[16] W. M. Bart and R. Williams-Morris(1990), A refined item diagraph analysis of proportional reasoning test, *Applied Measurement in Education*, vol. 3, pp. 143-165.

[17] W. M. Bart, T. Post, M. Behr and R. Lesh(1994), A diagnostic analysis of a proportional reasoning test item: an introduction to the properties of a semi-dense item, *Focus on Learning Problems in Mathematics*, vol. 16, pp.1-11.

[18] W. M. Bart, W. Rothen, and S. Read(1986), An ordering-analytic approach to the study of group differences in intelligence, *Educational and Psychological Measurement*, vol. 46, pp. 799-812.

[19] Y. H. Lin and W. L. Hung(2007), Robust clustering on rule usage of probability reasoning with raw rule score, *In The 4th International Conference on Fuzzy Systems and Knowledge Discovery*, China, Haikou, pp- 251-255.