

## AN APPLICATION OF RULE ASSESSMENT APPROACH FOR DIAGNOSING MISCONCEPTIONS\*

SIRIDEJ SUSHEVA

### ABSTRACT\*\*

This paper introduced the new diagnostic method for detecting misconceptions. The method was based on the rule assessment approach which stated that each testee had various rules of operation, and that the same responses might be obtained from the same or different rules. Therefore, the rule assessment of the whole set of testee's responses would make possible the tracing back of the rule and the diagnosis of the testee's misconceptions. This diagnostic process consisted of 3 steps : the first one was the developing of cognitive patterns in accordance with the concept of the content wanted to be diagnosed; the second one was the construction of item forms and the test in such a way that the response pattern could be traced back to get the testee's cognitive pattern; the last one was the two-step diagnostic process : exploratory diagnosis and confirmatory diagnosis. From the diagnostic method for misconceptions in mathematics, the negative signed-number addition, of a sample of 940 grade 8 students, it was found that the diagnostic result of the developed method was stable and had criterion-related validity when the teachers' diagnostic

---

\* Paper presented at the International Conference on Teacher Education: Innovative Alternatives for the 21st Century, July 11-14, 1995. Bangkok Thailand.

\*\* Doctoral Dissertation in Educational Measurement and Evaluation, Department of Educational Research, Chulalongkorn University, 1995.

result were used as a criterion. This method provided diagnostic result consistently with that of the Rule Space model's, and was able to avoid the limitation of the Rule Space model; i.e. - to discriminate misconceptions of the testees whose item score response patterns were the same or all zero's. In addition, the computer program written from this diagnostic method functioned as planned.

## INTRODUCTION

Recently several researchers from various disciplines such as cognitive psychology and psychometrics have developed rule assessment methods for diagnosing students' misconceptions. Researchers from each of these disciplines were motivated by an interest in knowledge structure and development in human cognition (i.e. Anderson (1981) and Siegler (1976) from cognitive psychology, and Tatsuoka (1982) from cognitive psychometrics) Tatsuoka and associates (Tatsuoka, 1983a, 1983b, 1985, 1987; Tatsuoka & Tatsuoka, 1985, 1987) have developed a systematic approach to the study of psychometric properties of misconceptions such as the stability of misconceptions committed by a student throughout a test.

Diagnosis of misconceptions can provide useful information in evaluating instruction and instructional materials and can suggest specific prescription for remediating a student. For example, the source of many of the misconceptions constructed by students is the ambiguity of explanation or the lack of precise, accurate instruction (Harnisch & Tatsuoka, 1991). Multiple-choice type diagnostic test is more widely popular in Thailand than other methods. This kind of test can diagnose student's misconception by the choice he/she chooses which is in accordance with the provided misconceptions. In fact there are a lot of misconceptions committed by the students in a specific content, but each item in the test can contain only four or five choices. The students might have other misconceptions which are not in the choices but they must choose one of the provided choices. In addition, the same choices could be obtained from various misconceptions that are not always the specific misconception provided by the test-maker and either the right choices could be obtained from misconceptions. To avoid these limitations of the multiple-choice type diagnostic test, personal indices (Drasgow, 1982; Harnisch, 1981, 1983; Harnisch and Linn, 1981, 1982; Harnisch and Tatsuoka, 1983; Sato, 1975; Wright and Stone, 1979) designing to summarized response patterns are used for detecting anomalous patterns which may result from application of some erroneous rules, but they have only limited power. For instance, these indices cannot diagnose sources of misconceptions or provide prescriptive information for

remediating students. Nevertheless, the use of personal indices is economical and applicable to general areas because such indices can first be used to spot candidates among students who may possibly possess misconceptions and may hence need personal attention from teachers. This procedure seem to be widely used in Japan (Tatsuoka, 1983). To get the more information for the misconceptions diagnosis beyond the personal indices' capability, Tatsuoka and associates (Tatsuoka, 1983a, 1983b, 1985, 1986; Tatsuoka & Tatsuoka, 1983, 1987) have developed a new probabilistic model for handling actual response patterns together with theoretically estimated response curves. This model has been named the Rule Space model. The model begins by mapping all response patterns in a cartesian product space  $\{(\theta, f(x))\}$ , called rule space. Mapping is accomplished using the Standardized Extended Caution Index and the true score ( $\theta$ ) following the Item Response Theory. This approach facilitated the discovery that a large number of actual response patterns match response patterns associated with some erroneous rule of operation by examining the proximity of a student's position in the rule space to the position represented by the erroneous rules. The Rule Space model has the limitation to discriminate misconceptions of the testees whose item score response patterns were the same or all zero's and it does not suitable to use in the classroom level because this model, applying Item Response Theory, has to take a large number of student to estimate the parameters in the model. The developed diagnostic method which is based on rule assessment approach was designed to handle the limitations of the previous approaches and provide the exact prescriptive information for remediating each student who has misconception.

### **RULE ASSESSMENT APPROACH**

The rule assessment approach (Seigler, 1976) focused upon conceptual development. It can be summarized in terms of five core assumptions.

1. Children's conceptual understanding progresses through a regular sequence of qualitatively discrete rules.
2. These rules are ordered in terms of increasing correlation with the correct rule in environments that the children encounter. Children will not adopt a new rule that is less highly correlated with the correct rule than the rule that they are already using.
3. The effectiveness of a learning experience is largely determined by whether the learning experience distinguished the child's existing rule from the correct rule. If the learning experience does not discriminate between the two rules, the child will continue to use the original rule. If the learning experience does discriminate between the two rules, the child will either adopt a more advance rule or will enter into a "rule search" state character-

ized by uncertainty as to which rule is correct.

4. A major reason why children do not immediately adopt the correct rule for all concepts and why they have difficulty in learning it consists in the limited encoding of the correct rules' component dimensions. Understanding of concepts frequently requires integration of several component dimensions. Failure to encode one or more of them can restrict children's ability to learn the concepts.

5. Failure to encode a dimension may be caused by lack of knowledge of the dimension's importance or lack of perceptual salience relative to other dimensions in the situation in which the concept is to be applied. Increasing a child's knowledge of the dimension's importance or increasing the dimension's salience can lead to improve ability to learn, which in turn can lead to improve knowledge. Central to this approach is a methodology for determining which rule an individual child is using. According to this approach, each testee had various rules of operation, and that the same responses might be obtained from the same or different rules. Therefore, the rule assessment of the whole set of testee's responses would make possible the tracing back of the rule and the diagnosis of the testee's misconceptions.

### **THE DEVELOPED DIAGNOSTIC METHOD**

The developed diagnostic method consisted of 3 steps : the development of the rules of operation; the construction of item forms and the test; and the two-step diagnostic process. The first step was the development of cognitive patterns or the rules of operation in accordance with the concept of the content wanted to be diagnosed. A rule was a description of a set of procedures or operations that one used in solving a problem in some well-defined domain such as arithmetic. A right rule was defined as a rule that produced the right answer to every item in a test, but an erroneous rule might fortuitously yeild the right answer for some subset of the items (Tatsuoka & Tatsuoka, 1987). The rules were developed on the basis of logical analysis of the task of solving signed-number addition problems and on analysis of actual responses given by students on a test, followed by clinical interviews. In this study 126 rules of the sign-number addition were developed. Table I describes three examples of the signed-number addition rule.

**TABLE I**  
**The Sets of Response Generated by Three Rules**

Items	Rule 1-2		Rule 2-1		Rule 2-4	
	Responses	Scores	Responses	Scores	Responses	Scores
1) $-9 + 3 = -6$	6	0	-12	0	-12	0
2) $-6 + (-2) = -8$	-8	1	-8	1	-8	1
3) $2 + (-7) = -5$	5	0	-9	0	-9	0
4) $-4 + 8 = 4$	-4	0	12	0	-12	0
5) $-3 + (-9) = -12$	-12	1	-12	1	-12	1
6) $5 + (-3) = 2$	-2	0	8	0	-8	0

Rule 1-2 : If the signs of the two numbers are same then the student adds the two absolute values together and takes the sign of the number with the smaller absolute value in his/her answer.

Rule 2-1 : The two numbers are always added and the sign of the number with the larger absolute value is always taken in the answers.

Rule 2-4 : The two numbers are always added as seen in Rule 2-1 but the negative sign is always taken in the answers.

The second step was the construction of item forms and the test in such a way that the response pattern could be traced back to get the testee's cognitive pattern. The basic strategy that has been used involves the formulation of problem types that yield distinct patterns of correct answers and errors for students using different rules. The six item forms were formulated for the content of the signed-number addition and used to construct the test as shown in figure I

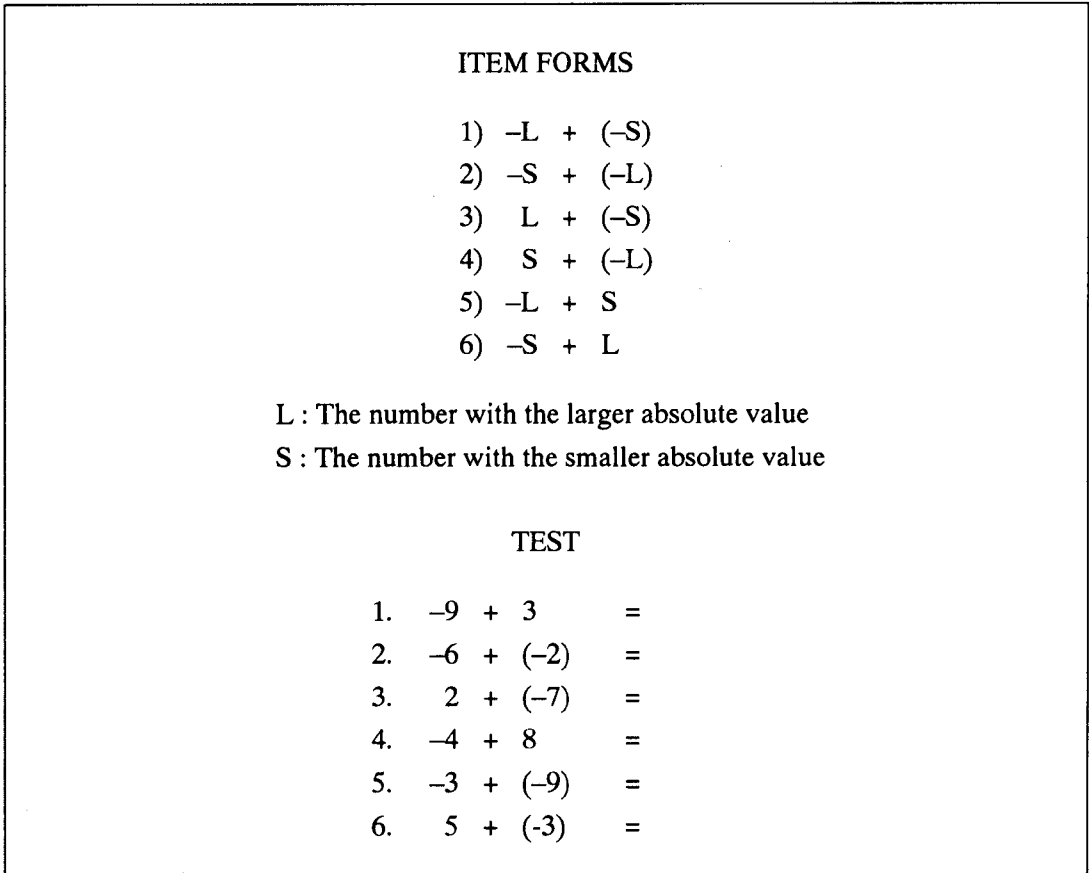


FIGURE I The six item forms for the content of the signed-number addition and the test.

The last step was the two-step diagnostic process : exploratory diagnosis and confirmatory diagnosis.

**Exploratory Diagnosis** : After formulating the response sets of each rule, the student's response set was compared with the response set of the rule one by one. All of 126 rules generated unique response sets. Therefore, the results of the comparisons were classified into three cases.

**CASE I** : Student's response set was the same as the response set of a rule. It will be diagnosed that he/she used that rule.

**EXAMPLE**

	Student's response	Rule produced the same response
1)	$-9 + 3 = \underline{-6}$	#1-5
2)	$-6 + (-2) = \underline{8}$	#1-5
3)	$2 + (-7) = \underline{-5}$	#1-5
4)	$-4 + 8 = \underline{-4}$	#1-5
5)	$-3 + (-9) = \underline{12}$	#1-5
6)	$5 + (-3) = \underline{2}$	#1-5

**CASE II** : Student's response set was not absolutely the same as the response set of any rules, but it was mostly like the response sets of only one rule. It will be diagnosed that he/she used that rule.

**EXAMPLE**

	Student's response	Rule produced the same response
1)	$-9 + 3 = \underline{-6}$	#1-7
2)	$-6 + (-2) = \underline{9}$	
3)	$2 + (-7) = \underline{-5}$	#1-7
4)	$-4 + 8 = \underline{4}$	#1-7
5)	$-3 + (-9) = \underline{12}$	#1-7
6)	$5 + (-3) = \underline{2}$	#1-7

(The response set of Rule 1-7 was -6, 8, -5, 4, 12, 2).

**CASE III** : Student's response set was not absolutely the same as the response set of any rules, but it was mostly like the response sets of more than one rule. This student had to be diagnosed in the confirmatory diagnosis process.

**EXAMPLE**

	Student's response	Rule produced the same response
1)	$-9 + 3 = \underline{-6}$	#1-1, #1-7
2)	$-6 + (-2) = \underline{-8}$	#1-1
3)	$2 + (-7) = \underline{-5}$	#1-1, #1-7
4)	$-4 + 8 = \underline{4}$	#1-1, #1-7
5)	$-3 + (-9) = \underline{12}$	#1-7
6)	$5 + (-3) = \underline{2}$	#1-1, #1-7

(The response set of Rule 1-1 was -6, -8, -5, 4 -12, 2

The response set of Rule 1-7 was -6, 8, -5, 4 12, 2)

**Confirmatory Diagnosis** : The students in case III had to be diagnosed by taking the confirmatory test. It is a true or false type test constructed from the previous test by selecting the items that are not be found all the suspect rules. From case III item I and item V are selected to construct the true or false type test which the provided answer of each item is generated from the rule that cannot be found in the his/her answer in the previous test.

### CONFIRMATORY TEST

$$1) \quad \quad \quad -6 + (-2) = 8$$

$$2) \quad \underline{\quad\quad} -3 + (-9) = -12$$

The provided answer of item 1 is generated from the rule 1-7 and the answer of item 2 is generated from the rule 1-1. The students' response can be generated into two cases.

**CASE I** : True and false or false and true response.

$$1) \quad \underline{/} -6 + (-2) = 8$$

$$2) \quad \underline{x} -3 + (-9) = -12$$

The student with these responses is diagnosed to have rule 1-7.

$$1) \quad \underline{x} -6 + (-2) = 8$$

$$2) \quad \underline{/} -3 + (-9) = -12$$

The student with these responses is diagnosed to have rule 1-1.

**CASE II** : All true or false response.

The student with all true or all false responses is diagnosed to have no consistent rules or use guessing.



## **VALIDITY AND RELIABILITY OF THE DEVELOPED METHOD**

The developed method was used to diagnose misconceptions of a sample of 940 grade 8 students. It was found that the diagnostic result of the developed method was stable and had criterion-related validity when the teachers' diagnostic result were used as a criterion. This method provided diagnostic result consistently with that of Rule Space model's, and was able to avoid the limitation of Rule Space model; i.e. – to discriminate misconceptions of the testees whose item score response patterns were the same or all zero's. In addition, the computer program written from this diagnostic method functioned as planned.

## **CONCLUSION**

The developed diagnostic method, by comparing the students' response sets to the rules' response sets with the confirmatory test for some students, facilitates diagnosis of the misconception that resulted in the erroneous rule. Students' responses are often affected by random errors produced by uncontrollable factors (Harnisch & Tatsuoka, 1991). The developed method enables us to take into account the careless, the erroneous rules, and the guessing. This approach for classifying students' error patterns avoid some of the weakness of personal indices, such as the problems of setting cut-off scores for determining atypicality of response patterns, dependence on a baseline ordering of items, and the lack of prescriptive information for remediating. In addition, it decreases the limitation of the Rule Space model; i.e. – to discriminate misconceptions of the testees whose item score response patterns were the same or all zero's.

## **ACKNOWLEDGEMENTS**

The author would like to gratefully acknowledge and thank thesis advisors, Asst. Prof. Nonglak Wiratchai, Ph.D. and Prof. Suwimon Wongwanich, Ph.D. Special thanks is extended to Dr. Harnisch and Mr. Choi from the University of Illinois at Urbana-Champaign for their assistance with the data analysis.

## REFERENCES

- Anderson, J.R. 1983. *The architecture of cognition*. Cambridge, MA : Harvard University Press.
- Birenbaum, M., et al. 1992. *Toward a stable diagnostic representation of students' errors in algebra* (Research Rep.). Princeton NJ : Educational Testing Service.
- Birenbaum, M. & Tatsuoka, K.K. 1987. Effects of on-line test feedback on the seriousness of subsequent errors. *Journal of Educational Measurement* 2 : 145-155.
- \_\_\_\_\_. 1992. Effect of response format on diagnostic assessment of scholastic achievement. *Applied Psychological Measurement* 16 : 353-363.
- Bloom, B. S. 1971 *Handbook on formative and summative of student learning*. New York : McGraw-Hill.
- Brown, F. G. 1978. 2nd ed. *Principle of education and psychological testing* New York : Holt, Rinehart and Winston.
- Drasgow, F. 1982. Choice of test model for appropriateness measure. *Applied Psychological Measurement* 6 : 297-308.
- Fisher, K. M. 1985. A misconceptions in biology : amino acid and transition. *Journal of Research in Science and Teaching* 22 : 53-62.
- Flavell, J. H. 1985. *Cognitive psychology*. New Jersey : Prentice-Hall.
- Gabel, D. L., et al. 1987. Understanding the particular nature of matter. *Journal of Chemical Education* 64 : 695-697.
- Good, C. V. 1973. *Dictionary of education*. New York : McGraw-Hill Book, Co.
- Gronlund, N. E. 1981. 4 th ed. *Measurement and evaluation in teaching* New York : Mcmillan Publishing Co, Inc.
- Halloun, I.A. and Hestenes D. 1985. Common sense concepts about motion. *The American Journal of Physics* 53 : 1056-1065.
- Harnisch D. L., & Linn, L. R. 1981. Analysis of item response pattern: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement* 3 : 39-87.
- Harnisch D.L., & Tatsuoka, K.K. 1983. A comparison of appropriateness indices based on item response theory. In R. Hambleton (Ed.) *Application of item response theory* Vancouver : British Columbia, 104-112.
- \_\_\_\_\_. 1991. *An application of IRT for diagnosing misconceptions* a paper presented in the proceedings of international educational statistics and measurement symposium. Taiwan : Educational Measurement Center, 12-20.

- Helm, H. 1980. Misconceptions in physics among south africa students. *Physics Educaton* 15 : 92-105.
- Hills, J. R. 1981. 2nd ed. *Measurement and evaluation in the classroom* Ohio : Bell & Howell Company.
- Kiokaew, S. 1988. Comparing college freshmen's concept of covalent bonding and the college of science and the college of education at Prince of Songkhla univesity, Thailand. Doctoral Dissertation, University of Missouri-Columbia.
- Levine, M. V., Rubin, D. B. 1979. Measuring the appropriateness of multiple chioce test scores. *Journal of Educational Statistics* 4 : 269-290.
- Lindquist, Franklin. 1956. *Educational measurement* Washington : American Council of Education.
- Mehrens, W. A. and Lehman, I. J. 1984. *Measurement and evaluation in education and psychology* New York : Holt, Rinehart and Winston.
- \_\_\_\_\_. 1980. 3rd ed. *Standardized tests in ecucation*. New York : Holt, Rinehart and Winston.
- Noll, V. H. and Scannell, D. P. 1972. 3rd ed. *Inrtoduction to educational measurement* Boston : Houghton Mifflin.
- Page, G. T. & Thomson, J. B. with Marshall, A. R. 1977. *International dictionary of education*. Great Britain : The Anchor Press Ltd.
- Peterson, R. F. and Treagust, D. F. 1989. Dvelopment and application of a diagnostic instrument to evaluate grade 11 and 12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science and Teaching* 26 : 301-314.
- Piaget, J. 1964. Cognitive development in the Child. *Journal of Research in Science and Teaching* 2 : 176-186.
- Pines, A. L. and West L. H. T. 1983. A framework for conceptual change special reference to misconceptions. In Novak, J. D. (ed.), *Proceedings of the international seminar misconceptions in science and mathematics* New York : Cornell University, 47-55.
- Ron, H. 1983. Enhancement and assessment of the reliability of instruments of the measurement of conceptual framework. In Novak, J. D. (ed.), *Proceedings of the international seminar misconceptions in science and mathematics* New York : Cornell University, 153-161.
- Sato, T. 1975. *The construction and interpretation of S-P tables* Tokyo : eiji Tosho.
- \_\_\_\_\_. 1982. *Application of S-P curve theory*. Tokyo : Meiji Tosho.

- Schmidt, H. J. 1987. Secondary school students' learning difficulties in stoichiometry. In Novak, J. D. (ed.), *Misconceptions and educational strategies in science and mathematics* New York : Cornell University, 396-404.
- Seigler, R. S. 1976. Three aspects of cognitive development. *Cognitive psychology* 8 : 481-502.
- \_\_\_\_\_. 1978. The origins of scientific reasoning. In R. S. Seigler (Ed.), *Children's thinking : What develops?* Hillsdale, N.J. : Erlbaum.
- Simson, W. W. and Marek A. E. 1988. Understanding and misconceptions of biology concepts holds by students attending small high school. *Journal of Research in Science and Teaching* 25 : 361-374.
- Singha, H. S. 1974. *Modern educational testing*. New Delhi : Sterling Pub.
- Sheehan, K. & Tatsuoka, K. K. 1992. *A diagnostic classification model for problem solving skill* (Research Rep.). Princeton NJ : Educational Testing Service.
- Sternburg, J.R. 1982. *Handbook of human intelligence*. London : Cambridge University Press.
- Tatsuoka, K.K. 1983a. Rule space : An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement* 20 : 345-354.
- \_\_\_\_\_. 1987. Validation of cognitive sensitivity for item response curves. *Journal of Educational Measurement* 24 : 233-245.
- \_\_\_\_\_. 1983b. Changes in error type over learning stage. *Journal of Educational Psychology* 76 : 120-129.
- \_\_\_\_\_. 1984. Caution indices based on item response theory. *Psychometrika* 9 : 95-110.
- \_\_\_\_\_. 1985. A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics* 12 : 55-73.
- \_\_\_\_\_. 1986. Diagnosing cognitive error : Statistical pattern classification and recognition approach. *Behaviormetrika* 19 : 193-206.
- \_\_\_\_\_. 1991. *Item construction and psychometric model appropriate for constructed responses* (Research Rep. No. RR-91-49-ONR) Princeton NJ : Educational Testing Service, 73-86.
- Tatsuoka, K. K. & Birenbaum, M. 1992. *Proficiency scaling based on attribute characteristic curves* (Research Rep.) Princeton NJ : Educational Testing Service.
- Tatsuoka, K. K. & Linn, L. R. 1983. Indices for detecting unusual patterns : Links between two general approaches and potential application. *Applied Psychological Measurement* 7 : 81-96.
- Tatsuoka, K. K., & Tatsuoka, M. M. 1982. Detection of aberrant response patterns. *Journal of Educational Statistics* 7 : 215-231.

- Tatsuoka, K.K., & Tatsuoka, M.M. 1983. Spotting erroneous rules of operation by individual consistency index. *Journal of Educational Measurement*. 20 : 221-230.
- Tatsuoka, K. K., & Tatsuoka, M. M. 1987. Bug distribution and pattern classification. *Psychometrika* 52 : 193-206.
- Treagust, D. 1986. Evaluating students' misconceptions by means of diagnostic multiple choice items. *Research in Science Education* 16 : 199-207.
- Wright, B. D., & Stone, M. H. 1979. *Best design, Rasch measurement*. Chicago : MESA Press.