DOCUMENT RESUME

ED 337 488 TM 017 319

AUTHOR	Cheung, K. C.; And Others
TITLE	Meaningful Assessment of Problem-Solving Activities in the Classroom: Some Exemplars. Research Monograph No. 2.
INSTITUTION	Institute of Education (Singapore).
REPORT NO	ISBN-9971-953-90-0; ISSN-0219-4326
PUB DATE	Jul 91
NOTE	86p.; Companion to an earlier volume "Meaningful Measurement in the Classroom Using the Rasch Model: Some Exemplars" (1990); see ED 326 544. Two figures may not reproduce well due to small print.
PUB TYPE	Collected Works - General (020) -- Reports - Evaluative/Feasibility (142) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Algebra; Classroom Techniques; *Cognitive Processes; *Educational Assessment; *Elementary School Students; Elementary Secondary Education; Error Patterns; Foreign Countries; Grade 8; Metacognition; *Problem Solving; Secondary School Students; *Word Problems (Mathematics)
IDENTIFIERS	Singapore

ABSTRACT

This collection of four papers deals with problem solving and the measurement of problem solving. "Climbing Up the Competence Ladder: Some Thoughts on Meaningful Assessment of Problem-Solving Tasks in the Classroom" by K. C. Cheung uses the metaphor of a competence ladder to represent the problem-solving continuum with progressive qualitative bands. "A Bane or a Boon?: Meaningful Assessment of Problem-Solving Activities in the Classroom" by K. C. Cheung and L. C. Mooi explores different ways of assessing problem-solving activities and describes a conceptual framework for problem-solving processes. ""On Meaningful Measurement: Stages of Lower Secondary Pupils' Abilities in Solving Algebra Word Problems" by W. F. Loh and K. C. Cheung outlines the critical steps experienced by 130 eighth-graders in Singapore in solving word problems and explains how these activities could be assessed. "On Meaningful Measurement: Metacognition and Hierarchical Modelling of Errors in Algebra Word Problems" by K. C. Cheung and W. F. Loh analyzes the main types of errors made by 130 eighth-graders at different performance levels. A 43-item list of references is included. An appendix contains the algebra test from the fourth study. (SLD)

Centre for Applied Research in Education

RESEARCH MONOGRAPHS

No. 2, 1991

Meaningful Assessment of Problem-Solving Activities in the Classroom: Some Exemplars

by

K. C. CHEUNG
MOOI LEE CHOO
LOH WE FONG

NANYANG TECHNOLOGICAL UNIVERSITY
NATIONAL INSTITUTE OF EDUCATION

ISSN 0219-4326

No. 2, 1991

Meaningful Assessment of Problem-Solving Activities in the Classroom: Some Exemplars

by

K. C. CHEUNG
NGOI LEE CHOO
LOH WE FONG

ISBN 9971-953-90-0

July 1991

FOREWORD

The primary function of the Centre for Applied Research in Education (CARE) is to promote, provoke and publicise educational research activities of the National Institute of Education (NIE), especially those that cut across Schools and facilitate application in schools and NIE itself. Publications would therefore constitute an important means of fulfilling such a function.

While we would continue to encourage and assist NIE staff to publish journal articles and research papers, both locally and overseas, there is pressing need for the publication of more substantial reports of some research studies or collections of closely related studies. The inclusion of more specific details is considered important for the meaningful application, replication or extension of the research. Hence, CARE has undertaken to publish the series of Research Monographs, which would be disseminated to a wider circle of researchers, as well as research-oriented policy-makers and practitioners.

The first two Research Monographs were chosen in order to illustrate not only the kinds of research that we regard as important but also the contrasting styles of reporting which are likely to interest different types of readers. The first one by Quah May Ling, Low Guat Tin and Yeap Lay Leng on "Adolescent Loneliness: Parents and schools can help" is likely to have a wider clientele in view of its more topical subject matter and the authors have rightly attempted to write in a less technical style, without sacrificing the essential elements of the study. The second one by K.C. Cheung, Mooi Lee Choo and Loh We Fong on "Meaningful assessment of problem-solving activities in the classroom: Some exemplars" is a compilation of several studies of a more technical nature. Although the authors have also tried to minimise the use of technical terms, there is a limit to how far they could proceed without losing the essence of their valuable contribution to knowledge.

As these are only our beginning efforts, we would like to hear from readers, especially in terms of their reactions to the first two issues and suggestions for possible future improvement.

Sim, Wong Kooi
Head, Centre for Applied Research
in Education
National Institute of Education

July, 1991

PREFACE

This collection of four papers is essentially a companion volume to an earlier publication entitled *"Meaningful Measurement In the Classroom Using the Rasch Model: Some Exemplars"*, published by ERU last year. This continuity in thinking and research efforts is important if we are to take full advantage of a possible paradigm shift in the way we look at and measure socio-psychological processes given the availability of more sophisticated statistical techniques. The term "meaningful" as applied to the measurement process is interpreted by K C Cheung against a constructivist perspective.

The process of human problem solving is a fascinating field, and as an area of research it has gone through a period of rapid expansion. In this context, I think this collection of four papers is a valuable contribution to the thinking on problem solving and the measurement of it.

In Chapter 1, K C Cheung using the metaphor of a *competence ladder* to represent the problem-solving continuum with progressive qualitative bands marking the steps (of a ladder) explains the theme of this volume, namely meaningful assessment of problem-solving activities. In Chapter 2, K C Cheung and Mooi Lee Choo explore different ways of assessing problem-solving activities in the classroom and at the same time describe a conceptual framework for problem-solving processes. Chapter 3 (by Loh We Fong and K C Cheung) outlines the critical steps lower secondary pupils experienced in solving algebra word problems, and explains how these problem-solving activities could be assessed meaningfully, i.e. basing qualitative interpretation on quantitative measurement. Finally, in Chapter 4, K C Cheung and Loh We Fong analyse both qualitatively and quantitatively the main types of errors made by pupils at different performance levels of the problem-solving proficiency continuum.

While this collection of papers focuses on different aspects of human problem-solving, it is also about psychometrics. I think much of the methodology in this volume was conceived through insights gained from more recent psychometric approaches, particularly those of latent trait analysis.

I would like to thank K C Cheung and two of our MEd colleagues, Mooi Lee Choo and Loh We Fong, for sharing with us their research findings and reflections on the measurement of problem solving processes.

Ho Wah Kam
Ag Dean
School of Education

21 April 1991

CONTENTS

CHAPTER 1

Climbing up the Competence Ladder: Some Thoughts on Meaningful Assessment of Problem-Solving Tasks in the Classroom

K C CHEUNG

Introduction

This is the second monograph in a series, following the publication of the first one on "Meaningful Measurements in the Classroom Using the Rasch Model: Some Exemplars" (Cheung et al., 1990). It endeavours to resolve issues on the assessment of problem-solving tasks. It demonstrates how psychometricians can assess pupils' progress on problem-solving tasks meaningfully. Hopefully, continued efforts in this direction will inform us of a viable pedagogy on problem-solving activities that are so highly endorsed in the school curriculum nowadays.

What is meaningful assessment of problem-solving tasks?

In the first monograph, Cheung (1990) remarked that "it is this view of measurement and testing within a constructivist philosophy that renders the measurement process a meaningful one" (p.2), and "it is the progression of the lower forms of knowing to higher forms of knowing that should be considered to be modelled on a continuum for quantitative measurement with qualitative interpretations" (p.4). This unidimensional continuum, which represents progressive development of forms of knowing that are firmly rooted in the educational objectives of the curriculum, should best be regarded as "resembling a pigtail fashioned as a bamboo stem" (p.4). Three examples from four authors were given: (1) the "part-whole" concept of fraction; (2) levels of Chinese language abilities progressing from "knowledge of characters", "grammatical sensitivity", to "reading comprehension"; and (3) levels of state and trait computer programming anxieties on "confidence", "errors", and "significant others".

It should be emphasised that latent trait theories do not in principle render the measurement process a meaningful one. It serves essentially as a means to structure pupil responses so that objective, quantitative measurement with qualitative interpretations that are specific to a target population becomes a reality. The continuum, once established, is meaningful because the progressive forms of knowing can be understood in terms of constructivist perspectives. It is against this backdrop that the present monograph is cast on the meaningful assessment of problem-solving activities.

Assessment of problem-solving activities is not easy because problem solving is a very complex form of human behaviour. Pupils need to recognise the starting points, understand the finishing points, search for alternative solution paths, control and execute the planned steps of the problem tasks. This process is not unique to individual pupils unless the problems are routine, exercise-type questions. Research on problem solving shows that experts and novices differ in the deployment of perceptual cues, recognition of a "deep structure" of the problem, equipment of a domain-specific knowledge base, understanding of conceptual knowledge, proficiency of procedural skills, deployment of heuristic strategies, awareness of the problem-solving process and its monitoring.

In order to understand this process and its pedagogy, a theory of problem-solving is needed. This theory needs to account for how problem-perception influences problem-solving approaches, the problem-solving stages/cycle, how information is processed within our cognising mind, and alternative conceptions within the domain-specific knowledge base of the problem tasks. Meaningful assessment of problem-solving tasks is an attempt to provide for each pupil a quantitative measure of problem-solving ability with desirable measurement properties such as linearity and objectivity, together with qualitative interpretations that are firmly rooted in an adequate theory of problem-solving. Some pertinent conceptual and methodological issues identified can be resolved. The ensuing chapters are attempts to illustrate these concepts of meaningful measurement using one example on pupils' solving of algebraic word problems.

What problems are we facing?

We can start with simple problems facing teachers everyday. How can we set a classroom test, for example, on the solving of algebra word problems? How can we mark these questions? How can we arrive at a quantitative measure of ability in solving algebra word problems? Can we know from this ability measure the progress of problem solving and the most probable types of errors made by the pupils? How can we assess pupils' work if there are a variety of solution paths? Can pupils be more conscious of their problem-solving processes? These questions, if answered, are valuable in informing us of a viable pedagogy on solving algebra word problems, in the light of an adequate theory of problem-solving mentioned earlier.

How to solve these problems?

Problem 1: Explicating the "deep structures" of problem tasks

Teachers need a blue-print in setting an open-ended test. This blue-print, which conveys the "deep structures" of the questions in the test, forms the basis of assessment. The concept of "deep structure" is a tricky one because even researchers on problem solving can have different opinions on an adequate definition for test construction purposes. <u>In this monograph, questions are said to possess the same "deep structure" if they can be solved by the pupils in particular ways as intended by the teachers, once the pertinent perceptual cues that are built into the questions are recognised. Also, associated with each "deep structure" is a domain-specific knowledge base targeted at the ability level of the pupils under examination</u>. Thus, the same problem may be solved in different ways by pupils of different grades/abilities because the "deep structures" of the same problem are different for the pupils.

This feature of subjectivity on an explication of the "deep structures" of problems concurs with the test grids/plans of multiple choice questions. While the multiple choice test is regarded as objective because of its pre-defined scoring procedures, the art of test construction in terms of contents and emphases is highly subjective. Realising this analogy, what is required in the assessment of problem solving is a method of presenting the "deep structures" of the questions as intended by the teachers as faithfully as possible and setting the questions accordingly. The "deep structures" should indicate how the questions may be solved by the pupils in particular ways, deploying the domain-specific knowledge base as intended and implemented by the curriculum. The use of problem-solving networks, presented in Chapter 2, is a significant attempt to achieve this objective. Problem-solving networks allow us to assess the construct validity of the problems by showing that each of these problems possesses the designed underlying "deep structure". When all the questions in the test are constructed according to a common "deep structure", the possibility of aligning the problems onto a unidimensional problem-solving proficiency continuum in the form of a pigtail fashioned as a bamboo stem is substantially enhanced.

Problem 2: Grading pupil responses for quantitative measurement

If quantitative measurement is desired, some forms of ordering of pupil responses to questions are necessary. There are at least two ways, each of which is applied to different purposes of measurement. The first one resembles the impressionistic marking of compositions. Based on some performance criteria, ordered grades are awarded for each question of the test. Across the questions, the range of grades may be different. Neither would the same literal grades which are awarded to two different questions correspond necessarily to the same ability level. What is necessary is that each of the ordered, graded responses should correspond to an unambiguous, ordinal set of competencies judged by the teachers as best as they can. This scoring scheme is very versatile because only the quality of outcomes is ordered, without bothering about the solution paths taken by the problem solvers. This scoring scheme is best suited for the non-routine problems. The next step after scoring the items is to check the unidimensionality of the questions spanning the test and to transform the ordered, graded responses into measures on an interval scale.

The second method of scoring, which applies to open-ended questions of the routine exercise-type, requires special arrangements for test construction. The purpose of measurement is not only to provide a quantitative measure of problem-solving ability, but also to examine how the problem-solving steps have been taken by pupils of different levels of problem-solving ability. The problem-solving networks, representing the "deep structures" of questions in the test, are needed to decide on the number of key steps in completing the problems. These key steps, deploying conceptual and procedural knowledge, may follow those of a problem-solving cycle such as on problem-understanding, problem-representation, problem-execution, and problem-control. Through an examination of pupil responses, it is also possible to know the key steps that have been taken by the pupils.

The rate-determining steps of a question can be anywhere along the solution path. Should the most difficult step be problem-understanding or problem-representation, pupils should find it relatively straightforward to proceed to completion once this step is secured. Otherwise, pupils may be blocked and are denied the opportunity to complete the ensuing steps. If this is the case, the ability levels of the problem-solving steps following the rate-determining steps may not be estimated accurately. Thus, after a set of core questions of the same "deep structure" have been constructed, it is important to include additional questions targeting at the components of the core questions. In this way, conceptual and procedural knowledge, which are needed to solve the core questions, can also be assessed accurately. The grades awarded to each of the questions in the final test are then the number of steps completed. Although the steps differ in level of difficulty, the problem-solving sequence of routine questions ensures that the more steps a pupil completes, the higher his/her problem-solving ability.

Problem 3: Aligning ordered graded responses onto a unidimensional continuum

Partial Credit modelling, an extension of the Rasch model, is ideal for analysing the two types of ordered, graded responses just described. Chapter 3 shows how this measurement model can be used to analyse pupil responses on the Algebra test in Appendix 1. The outcome is an alignment of all the steps of the questions in the Algebra test onto a unidimensional problem-solving proficiency continuum. Again, model assumptions and requirements have to be met before one can make use of the calibration results. Banding of this continuum into progressive forms of problem-solving competencies, and conceptual and procedural forms of knowing is also possible. This continuum can be likened to "a pigtail fashioned as a bamboo stem", showing how pupils progress towards mastery of solving of algebra word problems . Ideas of meaningful measurement discussed in the first monograph apply in full to this constructed continuum.

Problem 4: Dimensional and hierarchical structuring of errors and misconceptions

Pupils of different abilities commit different types of errors. Very often, errors committed by the high-ability pupils may not be observed in the low-ability pupils. This is because the low-ability pupils may not even understand or represent the problem at all. Errors can be loosely regarded as "inabilities" of the pupils, although the vast literature on "misconceptions" and "alternative conceptions" points to other perspectives. If meaningful measurement is to be pursued, the issue of how these "inabilities" are structured can inform us of the conceptual and procedural barriers towards mastery of problem solving.

Errors made by the pupils cannot normally be regarded as unidimensional. Nevertheless, some of these may be hierarchical in nature. Latent trait theories cannot be applied to mouel the structure of errors. Instead, key types of errors of the questions in the test can be tabulated against the band levels of the problem-solving proficiency continuum, showing the types of errors made by pupils of different levels of problem-solving ability. This contingency table may be explored for any key structural dimensions accounting for the observed frequency patterns. Dual Scaling, presented in Chapter 4 of this monograph, is a multidimensional modelling tool to unveil the hidden structure of a contingency table. The structural dimensions may then be examined to see how the ordered bands of problem-solving proficiency are aligned with the types of errors. If this is carefully carried out, the first structural dimension will account for most of the information in the table and indicate the qualitative shift of the types of errors, including question omissions, made by pupils along a problem-solving ability continuum.

The competence ladder

The discussions lead to a very important concept on the meaningful assessment of problem-solving activities : The Competence Ladder. <u>On the left arm of the ladder is a "pigtail fashioned as a bamboo stem", which represents the progression of conceptual and procedural knowing along the problem-solving ability continuum. The progressive, qualitative bands of this "pigtail" mark the positions of the rungs of the ladder. The rungs symbolically link the left</u>

arm of the proficiency continuum to the right arm of the key structural dimension(s) of errors. From this competence ladder, we know not only how pupils of different abilities perform, but also what types of errors they make. Informed by an adequate theory of problem-solving, a pedagogy of problem-solving may be proposed. Knowledge of metacognitive decision-making processes and affective behaviours of pupils during problem solving help pupils monitor their problem-solving processes. It is this whole package of ideas, not the modelling methodologies, that constitutes the theme of this monograph: meaningful assessment of problem-solving activities.

CHAPTER 2

A Bane or a Boon? : Meaningful Assessment of Problem-Solving Activities in the Classroom

K C CHEUNG & MOOI Lee Choo

Introduction

This chapter explores ways of assessing problem-solving activities in the classroom. It does not deny the usefulness of the "think aloud" methodologies and "protocol analysis" in understanding the processes of problem solving, particularly in solving non-routine problems in novel situations. What this chapter attempts to do is to describe a viable theoretical framework on problem-solving processes, based on a theory of perception, a theory of information processing, a version of problem-solving cycle, and a consideration of domain-specific knowledge base. Network analysis, which is used to systematically organise pupils' responses to problem-solving activities, is introduced as an alternative to flowcharts. It is also used as an aid to understanding the "deep structures" of the problem tasks.

The legitimacy of this kind of qualitative analysis, in terms of the scientific status of knowledge that is constructed, is discussed by drawing an analogy with the Repertory Grid Techniques in the assessment of Kelly's personal constructs. As such, it is argued that network analysis, set within the proposed theoretical framework on problem solving, is compatible with the current interest in a constructivist pedagogy. It is hoped that the ensuing discussion would serve to invoke further discussions and exchanges. It is anticipated that more issues need to be resolved before teachers can assess problem-solving activities in the classrooms meaningfully. Consequently, a boon to psychometricians may be a bane to the classroom teachers.

What is problem-solving?

Problem-solving activities can be analysed as consisting of stages: problem-perception or problem-categorisation, problem-representation or problem-reformulation, problem-control or problem-execution, and reasonably, problem-evaluation. These often sequential problem-solving stages correspond to a theory of perception logic, a theory of information processing, and a version of the problem-solving cycle. These ideas contribute to a viable theoretical framework towards an understanding of problem-solving processes.

1. A theory of perception logic

A neurological perspective, based on the idea of connectionism which is associated with nerve networks in the explanation of the logic of perception rather than the logic of thinking, is now emerging. The perceptual system is likened to a self-organising, pattern-making, and pattern-using system in that perceptual cues bring in past experiences. Upon reaching certain stimulation thresholds of the neural system, the perceptual cues help to reconstruct images that have been visualised before. It is believed that the "circular" movements of patterns and sequences of these constructed images in our brains constitute human thoughts (De Bono, 1990). Rehearsal of perceptual skills, often domain-embedded, is considered to result in increased connectedness and permanency in the long-term memory stores of some parts of the patterned neural networks forming a domain-specific knowledge base. As such, practice and rehearsal enhance the pattern-using power of the perceptual system. Furthermore, boredom and perseverance can have directional effects on perception, in that some people are more ready to attend to perceptual cues and engage in sustained attention than others. Evidences from neurological researches are encouraging in supporting this connectionist theory and logic of perception.

2. A theory of information processing

Guided by a constructivist pedagogy, Osborne and Wittrock's (1985) generative learning model views the brain as an information-processing system for meaning construction. Under the premise that children are motivated to learn and they hold responsibility for their own learning, selective attention to and sustained interest in an experience, as guided by the relevant memory stores and cognitive processes, result in selective perceptions and input of sensory information. Linkages are generated to relate these to the relevant aspects of information in the long-term memory and the generated links are viewed from the constructivist perspective as critical for the meanings that are constructed. Only those recognised as viable are subsumed into the long-term memory. Later information retrieval is made easier since knowledge is structured and ideas interrelated. Recent studies on metacognition illustrate the importance of pupils being conscious of their learning processes. Hopefully, they can monitor and control the cognitive and emotional systems in the construction of knowledge.

3. A version of the problem-solving cycle

A theory of problem solving hinges upon what constitutes a problem which can be delineated on a continuum ranging from routine exercises to non-routine problems. A non-routine problem for one person may only be a straightforward exercise for another, depending on the prior experiences of a person, and the possession and utilisation of a relevant domain-specific knowledge base. In non-routine problem solving, particularly those problems in novel situations, some modifications, transformations, reformulations, or reconstructions of given concepts, relations, and procedures may be required before viable solution paths are in sight. Furthermore, successful problem solvers are often equipped with a number of heuristic strategies, of which some are domain-specific, for the perception, representation, execution, control, and evaluation of the problem-solving processes.

Earlier Gestalt theorists viewed problem solving as an interplay between the successive restructuring of the task environment and the recentering of the emphases/goals of a problem such that gaps can be closed and inconsistencies settled. Insights are considered to play a part for some of the surprise solutions. Later information processing theorists such as Newell and Simon (1972) focussed on how the critical features of the task environment structure the mental representations of the problem (problem space) in which heuristic strategies may be deployed for the search of viable solution paths (search space). A task environment with no recognisable critical features indicative of a "deep structure" (schema) and its associated knowledge base may result in an ambiguous problem representation that may be based on the surface features only. This unclear representation may encourage a trial and error approach rather than the deployment of the more systematic heuristic strategies. For experienced problem solvers, the perception and organisation of a manageable number of "chunks" of information is also an important factor in solving complex problems.

With these in mind, a version of the problem-solving cycle can be understood as composed of the following five key processes which may not proceed in the listed sequence : (1) from problem-perception to "problem-understanding"; (2) a recognition and explication of relevant knowledge schema in "problem-representation"; (3) the use of formalised, goal-oriented rules or heuristic strategies in "problem-execution"; (4) the exercise of meta-level decision-making processes in monitoring the solution paths in "problem-control"; and (5) the formative and summative evaluation of the problem-solving processes in "problem-evaluation". These five processes should best be analysed by network analysis.

What is network analysis?

Network analysis, derived from systemic linguistics, seeks to clarify and make explicit the researcher's subjective scheme of descriptive categories in encoding and interpreting qualitative data. This is done in order to avoid over-simplified taxonomies, or extensive although selective quotation of the data, for a fuller explication of meanings. There are four useful network notations (Bliss, Monk and Ogborn, 1983). Subcategories (bar) are mutually exclusive terms of a classification system such that these terms derive their meanings by the contrasts they make amongst themselves and by the increase in levels of delicacy as a result of further distinctive categorisation. Progressing towards higher levels of delicacy, terminals are reached which are terms without subcategories. On the contrary, the meanings of a term can be derived simultaneously from its distinct and independent facets and these should be co-selected or bracketed (bra) for the purpose of its multi-faceted representation. This co-selection is analogous to the different dimensions of a contingency table.

Repeated selections, whether applied to a bar or a bracket, render the possibility of a number of recursions (rec) into a part of the network for any addition of meanings, whether in combinations or in perspectives. Restricted entry to some parts of a network given some entry conditions is done through the use of restrictive entry conditions (con). These conditions are bracketed by a reverse bracket, which points to a specific term in the network, thus denying other combinations of entry conditions. A paradigm, represented by a trace of a path through the network, is a legitimate combination of terms in a network so that the meanings derived are congruent with the intentions of the researcher and instances of such can be found in the data.

What is a problem-solving network?

Research on problem solving generally reveals that experts and novices begin their problem-representation with specifiably different problem types/categories that they are able to perceive or recognise although experts and novices are cued differently by the surface features of a problem. Novices, as well as experts when they encounter familiar problems, search for familiar features of some typical problems they have encountered so that similar problem solving procedures can be applied. However, when faced with a novel situation, experts can initially recognise configurations of surface features and as a result of their experience suggest principles/schema and make use of the associated knowledge base to start categorising problems. On the other hand, novices are found to base their categorisation essentially on the surface or literal features of a problem task. Research also shows that completion of the problem-representation or problem-reformulation, which results in an articulation of a "deep structure" of the problem, depends on the problem solver's domain-specific knowledge associated with the problem categories. The same "deep structure" can correspond to a set of superficially disparate problems with diverse surface/literal features. Experts are at an advantage in making use of the knowledge base associated with the "deep structure" on embellishing or informing how a problem should be represented or reformulated.

As a result of these different ways of articulation, there may be a number of solution paths to be spelled out in full in terms of problem execution and control. These routes define the problem and search spaces so that the problem solvers' actions can be mapped by a network, which may be documented by the protocols. The network depicts how information is processed and strategies are sequenced and deployed by the problem solvers. It is worth mentioning that the problem space of each problem solver during the problem-solving process is by no means static. Metacognitive and attitudinal aspects of problem solving will exercise overall control over the ever-evolving search space as a result of the reconstruction of the problem space for the more viable heuristic strategies. From this perspective, a problem-solving network is more flexible to represent the problem-solving processes of pupils in a classroom than a flowchart which may not be equally applicable to all problem solvers especially on non-routine problems.

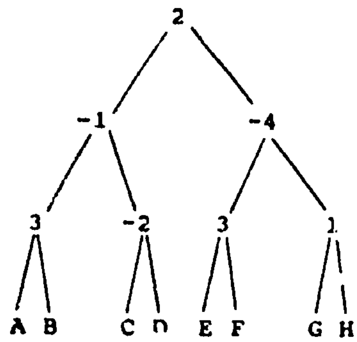
The task approach, task complexity and the accompanying plans and actions in a specific task context for each stage in problem solving can be viewed as of lower "ranks" (ie. units of response data being further broken down into smaller parts) in a network analysis than the problem-solving stages which are, for the sake of network analysis, viewed as of the same "rank". This "rank" scale determines what chunks of qualitative response data are associated in a temporal/causal sequence, at what levels and to what degree of delicacy, so as to represent the problem and search spaces of problem-solving activities adequately.

An example of a problem-solving network

The following presents a problem-solving network of an algorithmic thinking exercise of routine problems in a course on data structures (binary tree) at the Junior College (grade 12) level in Singapore. The two example problems listed below possess the same "deep structure" in the relevant content area and the solution of which is algorithm-based. However, the surface features are different because the problems are framed in different contexts. Pupils are expected to deploy the algorithms/concepts that have been taught in the lessons. Since the problems are routine exercises, not much variations are revealed in the pupils' responses on problem-categorisation, problem-representation, problem-execution and problem-control. As such, the problem and search spaces are uniform across pupils and problems, and these spaces correspond to those of successful problem solvers. Should the problems be non-routine problems in novel situations, a variety of problem representations and use of heuristics proliferate. The task of representing these alternatives and detours in a single network becomes gigantic although not impossible.

1. Problem One:

The results of a knockout football competition for schools can be arranged as follows:



The value at each terminal node in the tree is the name of a school (A,B,...,H). The value at each other node is the goal difference when the "home" team (left-hand) played with the "away" team (right-hand). A negative value means the "away" team won. Describe an algorithm to find the winning team.

2. Problem Two:

In Morse code, each letter of the alphabet is assigned a unique combination of dots and dashes. For example, the letters A, B, C, and D are coded as follows:

A : .-

B : -...

C : -.-.

D : -..

This coding system can be represented by a binary tree such that, except for the root node, each node of the binary tree corresponds to a letter of the alphabet. The Morse code of dots and dashes can then be represented respectively by traversing, starting from the root node, the left and right branches in the successive levels of the binary tree. Assuming the complete binary tree for the Morse code exists, describe an algorithm which uses this tree to print the letter represented by a given Morse code.

Figure 2.1 presents a viable problem-solving network after examining pupils' responses on the two example problems. Associated with each problem-solving process is a set of problem-solving stages, heuristic strategies, procedural skills, and the domain-specific knowledge base. In order to help the reader to understand the network which represents the "deep structure" underlying these two example problems, the following brief notes on binary trees is provided.

"Trees" in computer science grow upside down. The data in a tree are contained in its "nodes", or branching points, and are hierarchically arranged by levels. The node at the highest level of the hierarchy is called the "root" of the tree, whereas the nodes at the lowest extremities are called the "terminal nodes" or "leaves". A "binary tree" is one in which each node may have at most two subtrees: the right and left subtrees. These subtrees are binary trees in their own right. The usual way of representing a tree in computer science involves the use of "pointers", pointing to the addresses of arrays where the roots of the subtrees are stored. For a binary tree, each node consists of the data value, which may be a number or a literal, plus two pointers. One or both of the pointers may have "null values" if they have no subtrees to point to.

A tree may be "traversed" in several ways. Traversing a tree is accessing its stored content in a systematic way. How the tree is traversed is an essential part of the development of an algorithm in solving the programming problem and this cannot be performed without considering the data which the algorithm manipulates. For brevity, tree traversal may proceed by "level" and/or "order" such that a systematic sequence to reach the goal is achieved by combinations of "top-down" and "left-right" searches. In the data structure literature, methods of tree traversals such as "preorder", "inorder" and "postorder" constitute part of the domain-specific knowledge base of pupils faced with problems on binary tree structures.

It should be noted that the two example problems differ in the surface features underlying the binary tree data structure. In problem one, a positive data value indicates that the "home" team won and hence taking the left branch of the tree would lead to the winning team. Applying this scheme to the given binary tree, the mode of tree traversal is:

Starting at the root,

If the data value is positive, take the left branch, else if the data value is negative, take the right branch.

At the last level of the tree, the node reached will give the name of the winning team.

For problem two, a dot indicates taking the left branch while a dash means taking the right branch of the binary tree which represents the system of Morse codes. By following the given sequence of dots and dashes, the code character can be arrived at. The mode of tree traversal is therefore:

Starting at the root,

If the next symbol is a dot, take the left branch, else

if the next symbol is a dash, take the right branch.

At the end of the sequence of symbols, the node reached will give the corresponding code character.

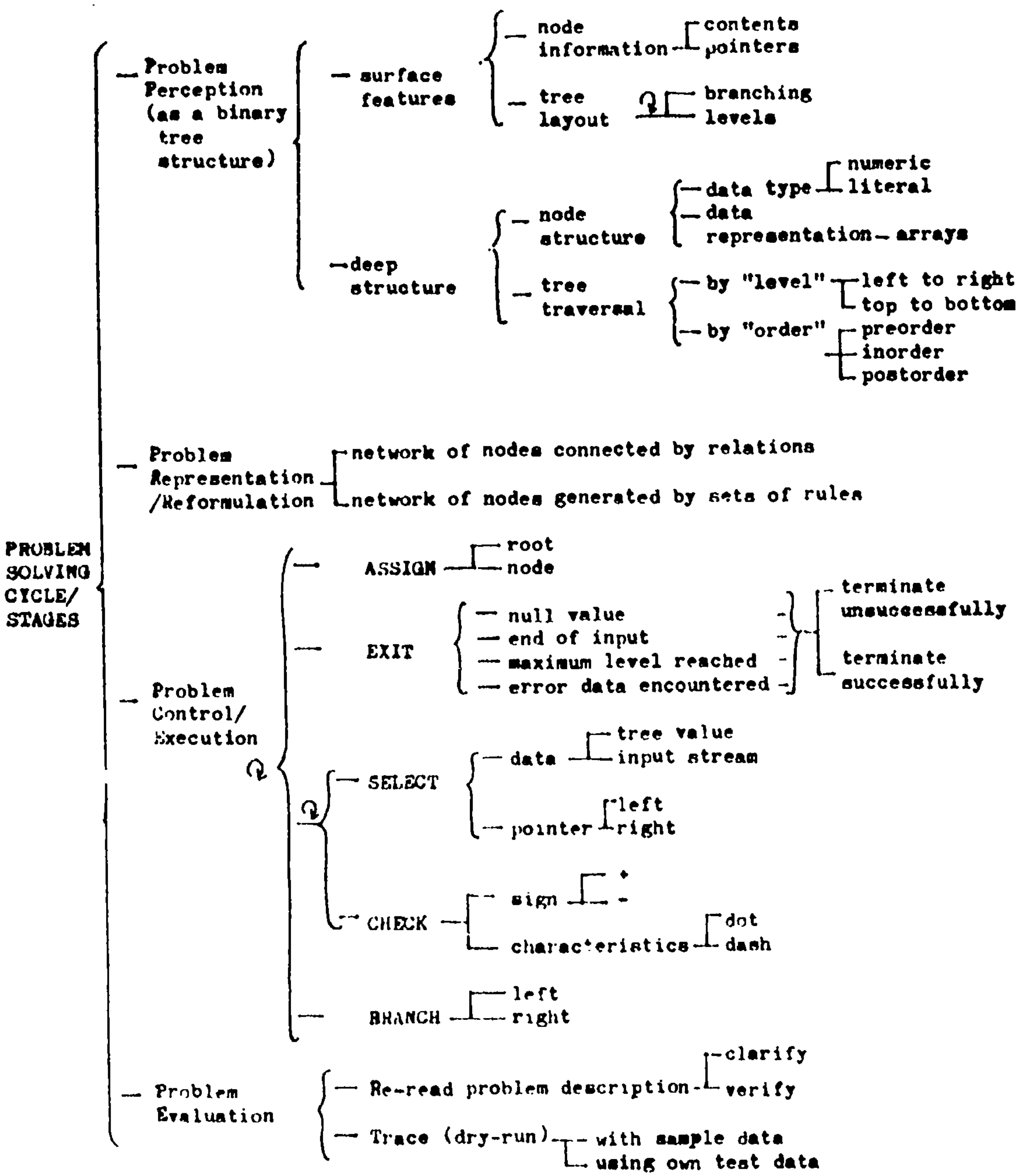


Figure 2.1 Network of problem-solver's actions

Is network analysis scientific?

This section attempts to draw an analogy with networks (cf. network analysis) and Kelly's personal constructs (cf. Repertory Grid Analysis) so as to assess the scientific status of problem-solving networks.

Kelly (1955) commented that man makes sense of the world by constructing and fitting representational models of reality. These mental models will only be amended if they are found to be wanting, for instance, in order to account for anomalies. As a result of daily experiences, man builds up his system of personal constructs which is essentially a system of subjective evaluation of his encounters against his anticipation of events. The repertory grid technique, invented by Kelly, enables the elicitation and exploration of part of this system by simultaneously noting likenesses and contrasting differences of events. The basic philosophy is that it is not possible to affirm logically something without at the same time denying something else.

The technique involves presenting a person with a series of different stimuli so as to trigger a different construct in each case. This triggering process, unlike those of the stimulus-response theory in behaviourism, refers to how one responds to what he perceives to be a stimulus. Hence, the probability of two persons perceiving events in identical ways is not large and this contributes to a person's individuality. For example, by associating London with Sydney and contrasting them with Tokyo the construct of Type of Spoken Language can be obtained, with English

Speaking and Non-English Speaking at the two poles of this reference axis. In this way, a grid of a universe of stimuli/events and a system of personal constructs is constructed. Ratings and rankings of stimuli, and clustering and ordering of constructs are obtained by outcome. Thus, personal/shared meanings of events are determined by where we are in this/these referenced grid structure/s, where the process of knowing and anticipating is channelled.

Since each construct is a personal, resemblance-contrast reference-axis contrasting two poles such that two events can be likened and then simultaneously contrasted with a third, this is reminiscent of the bar of categories of a term in a network in that decisions on categorisations have to be subjective but data-dependent. While the poles and categories are absolute entities once they are decided, the choice of interpretation of an event according to this set of constructs or classification system still depends on how one anticipates a greater possibility amongst the alternatives for a more adequate elaboration of meanings, whether for constricted certainty or broadened understanding.

The system of constructs of a basic concept of an event is analogous to the simultaneous aspects of a term within the same bracket in a network. However, these simultaneous aspects may differ in levels of abstraction and one aspect may be analysed as being more superordinate than the other. Since a construct allows us to locate personal meanings and anticipate events, constructs are necessarily cross-referenced by the users. Thus, the ordered nature of constructs in a personal system is parallel to the levels of delicacy of a network whereas the cross-referenced, structured nature, to some extent, may indicate the branching, recursion, and restricted combinations of some basic terms of a network.

Like the constructs which are useful for the anticipation of a finite range of events, the paradigms of a network are usually finite and can be instanced in the qualitative data. Furthermore, a personal construct system is ever-evolving in coping with anomalies and daily experiences. Thus, a succession of networks is necessary to indicate changes in conception which is a fundamental process of learning. This is because each of these networks resembles a static snapshot of the personal construct system on how we may perceive and anticipate events. The fallibility of the system of personal constructs in the light of anomalies, and the openness and adjustability of the system to include new constructs or to modify old ones, whether they are superordinate or subordinate, are close to the idea of insufficiency of detail/scope/notation of a network structure in coping with new circumstances.

One may also view knowledge in pieces, an idea from di Sessa (1985) that these pieces are loosely-coupled in a flexible, self-contained manner so that the personal system of constructs can be used more creatively. Branching within a network may indicate this kind of knowledge structure as well. Kelly maintained that where one personal construct system is similar to the other, the corresponding psychological processes involved are also similar.

Likewise, each paradigm of a network should correspond to a fixed set of meanings and processes so that the frequency of occurrences of this paradigm can be counted in the network data in an unambiguous manner. In this case, the network can represent shared meanings/perspectives of each instance of the network data.

In summary, network analysis, like the repertory grid technique and its associated Personal Construct Theory, is a way of organising experiences and knowledge. Since human knowledge is both personally constructed and socially shared, the network of qualitative data representing such knowledge is necessarily linked to a process of knowing. In this way, Kelly's ideas of personal constructs and the powerful paradigm of constructivism can furnish a viable philosophical base on how networks should be constructed and utilised. Criteria of validity of a network such as the viability of a network in coping with old and new data, can be examined accordingly. Without a philosophy of knowing, network analysis remains a useful tool in coding and representing experience only.

CHAPTER 3

On Meaningful Measurement: Stages of Lower Secondary Pupils' Abilities in Solving Algebra Word Problems

LOH We Fong & K C CHEUNG

Some contemporary views of a theory of problem solving

Lower secondary pupils experience considerable difficulty in solving mathematics word problems. Despite the attempt by curriculum planners to include more problem-solving activities in the mathematics curriculum and the large volume of research on "expert-novice" problem-solving, problems related to how to teach for both conceptual understanding and procedural proficiency at the lower secondary level appear to have persisted. Perhaps two major problems are that meaningful assessment of problem-solving activities is not an easy task for classroom teachers, and that a viable pedagogy of problem-solving in mathematics is by no means clear. This chapter seeks to resolve the first problem of meaningful assessment, whereas the second issue of pedagogy will be explored in Chapter 4 of this monograph.

Chapters 1 and 2 of this monograph outlined a perspective of problem-solving, taking into account a theory of perception logic, a constructivist theory of information processing, a version of a problem-solving cycle, and a consideration of an associated domain-specific knowledge base of the problem tasks. It is clear that without such a theoretical framework, recommendations on a viable pedagogy on problem solving may not be meaningfully proposed. Furthermore, an advocation on the use of problem-solving networks in constructing problem-solving tasks and analysing response behaviours is a first and significant step towards expressing what processes should have been undertaken, and describing what and how the outcomes are obtained when pupils are engaged in problem-solving activities. Network analysis, unlike research-based techniques such as "protocol" and "episode" analyses used in "think aloud" researches, is recommended for its versatility in summarising pupils' solution steps and results, and clarity for classroom teachers in explicating the designed "deep structure" of problem-solving tasks during test construction.

Some known misconceptions in solving algebra word problems

Recent studies have shown that success in solving mathematical word problems is due to abilities in constructing appropriate representations of problem situations, reformulating problem conditions, developing solution plans using a relevant domain-specific knowledge base, and deploying procedures and heuristics to generate viable solution paths (Davis, 1985; De Corte and Verschaffel, 1985; Venezky and Bregar, 1988). Amongst these abilities which teachers and pupils find troublesome to teach and learn are problem-understanding and problem-representation, which are critical stages in a number of proposed models of mathematical problem-solving (Schoen and Oehmke, 1980; Davis and Silver, 1983; Lesh, Landau and Hamilton, 1983; Riley, Greeno and Heller, 1983;

Bransford, et al., 1986; Venezky and Bregar, 1988; Kouba, 1989; Schroeder and Lester, 1989).

Put in a simple way, problem understanding helps the construction of an initial representation of a problem, and the setting up of explicit goals which will direct the course of problem solving (Janvier, 1987; Lesh, et al, 1983; Silver, 1987). In this regard, Resnick (et al., 1981) proposed that there are at least three possible ways to represent a problem: (1) informal or linguistic, (2) physical or visual, and (3) algebraic. Combinations of these ways may be applied in a single problem-solving task, affecting the heuristic strategies and metacognitive decision-making that are deployed during the problem-solving process (Lesh et al., 1983). Apart from those difficulties that arise from the problem-solving process, misconceptions in the content and process skills of algebra can hamper the successful solving of algebra word problems as well. Some of these misconceptions identified in a number of research studies are:

1. Different interpretations of letters and variables - Children generally do not understand the idea of a letter as a variable. They tend to interpret letters as standing for specific numbers, and different letters must necessarily represent different numbers. They do not understand that a letter can represent zero or can take on any value, positive or negative. Some are confused over the distinction between a letter as representing the number of a given object, and as a shorthand notation representing the actual object (Kuchemann, 1981; Booth, 1984).

2. Improper use of brackets - Children ignore the proper use of brackets by assuming that the problem context determines the order of operations, or that operations are simply performed starting from left to right within an algebraic expression (Booth, 1984).

3. The need to give numerical or single-term answer - Children are reluctant to record an algebraic statement as an answer, thinking that a numerical answer is required (Booth, 1984). They also tend to give numerical values to letters and variables at their earliest opportunity (Collis, 1975; Kuchemann, 1981; Booth, 1984; Ellerton, 1985; Dickson, 1989).

4. Different interpretations of terminology such as "solve", "simplify", "factorise" and "evaluate" - Children have difficulty in deploying the formal methods that have been taught to them even in the case of simple arithmetic. An example is that the procedures used by pupils in solving arithmetic problems are difficult to be symbolised (Booth, 1984). Terminology and the rules of algebra that are obvious to teachers may be a source of puzzlement and confusion to the children (Thwaites, 1982; Ellerton, 1985).

5. Understanding of the equality symbol - Equations of the form "a=bx+c" are misinterpreted as trying to give a numerical answer to the "sum" on the right-hand side of the equality symbol. Similarly, equations of the form "ax+b=cy+d" are seen as two sums

to be evaluated. The equality symbol is not conceived as an operation which should be applied identically to both sides of the equation. Rather, it is viewed as an operator symbol requiring the respondent to do something on the "sums" that are given separately on both sides of the equation (Behr, Erlwanger and Nichols, 1976; Kieran, 1981; Dickson, 1989).

6. The tendency to invent "malrules" - This occurs when some prototype rules are created by generalising taught systematic rules in order to extrapolate new rules which are considered by experts as incorrect (Radatz, 1979; Matz, 1982; Sleeman, 1982; Thwaites, 1982; Van Lehn, 1983; Resnick, et al., 1987). An example on the use of "malrules" is to extrapolate "a+(b*c)=(a+b)*(a+c)" from the taught systematic rule "a*(b+c)=(a*b)+(a*c)" by creating a prototype rule obeying the distributive law of algebra.

Purpose of study

Children attend differently to contextual cues, gain differential access to appropriate domain-specific knowledge base, deploy a diversity of heuristic and metacognitive strategies, search for viable alternative solution paths, commit various types of conceptual and procedural errors, and are found lost in detours and trapped in blind alleys in their solving of algebra word problems. The purpose of this chapter is to examine how lower secondary pupils in Singapore engage in solving algebra word problems, and to explore how problem-solving activities can be meaningfully assessed according to the criteria described in Chapter 1 of this monograph.

Network analysis of two routine exercise-type algebra word problems (items 10 and 11 in Appendix 1) reveals the commonly designed "deep structure" of the problem-solving tasks. This allows other simpler questions (items 1 to 9) to be constructed, based only on some key problem-solving steps such as problem-understanding and problem-representation and some key conceptual and procedural skills listed as components in this network. Analyses of the problem-solving steps taken by the pupils and their misconceptions contribute to a scoring scheme on item performance levels, which reflect critical response thresholds along the solution paths for successful problem solving. Through the use of Partial Credit Modelling, the calibrated item response thresholds are checked to ascertain whether they behave as designed within qualitative, progressive bands such that the levels of both conceptual and procedural understanding they represent are firmly rooted in the problem-solving research literature.

Design of the sample and algebra test

The sample for this study consists of 130 secondary two (grade 8) pupils chosen from four classes of a typical government-aided secondary school in Singapore. The sample comprises one class of high-ability pupils, two classes of average-ability pupils in the "express" stream, and one class of below-average ability pupils from the "normal" stream. In Singapore, the "express" stream pupils normally take four years whereas the "normal" stream pupils take at least one additional year to complete their secondary school education.

Algebra is first introduced to pupils in Singapore when they are in primary six. At the end of secondary one, pupils are expected to be able to simplify algebraic expressions of one unknown which may involve brackets, evaluate algebraic expressions by substitution, use symbols and letters to represent numbers and express physical situations, solve simple algebraic equations and algebra word problems (Ministry of Education, Singapore, 1981). Appendix 1 contains 11 questions, with items 7 and 9 in two parts, pertaining to this syllabus. Since the purpose of this test is to establish a progressive, qualitative continuum indicating how competencies may be deployed and barriers may be surmounted, those conceptual knowledge and procedural skills leading to the successful solving of items 10 and 11 are built into items 1 to 9 of the test. Otherwise, should problem understanding and problem representation be the most difficult steps, the difficulties of the ensuing steps on problem execution and control may not be estimated accurately. This is because pupils are not given the opportunities to demonstrate their competencies when they fail to secure a good start.

The common designed "deep structure" of items 10 and 11 is shown in Figure 3.1. Based on this problem-solving network and alternative conceptions obtained from analyses of pupils' responses, the performance score levels of each of the test items

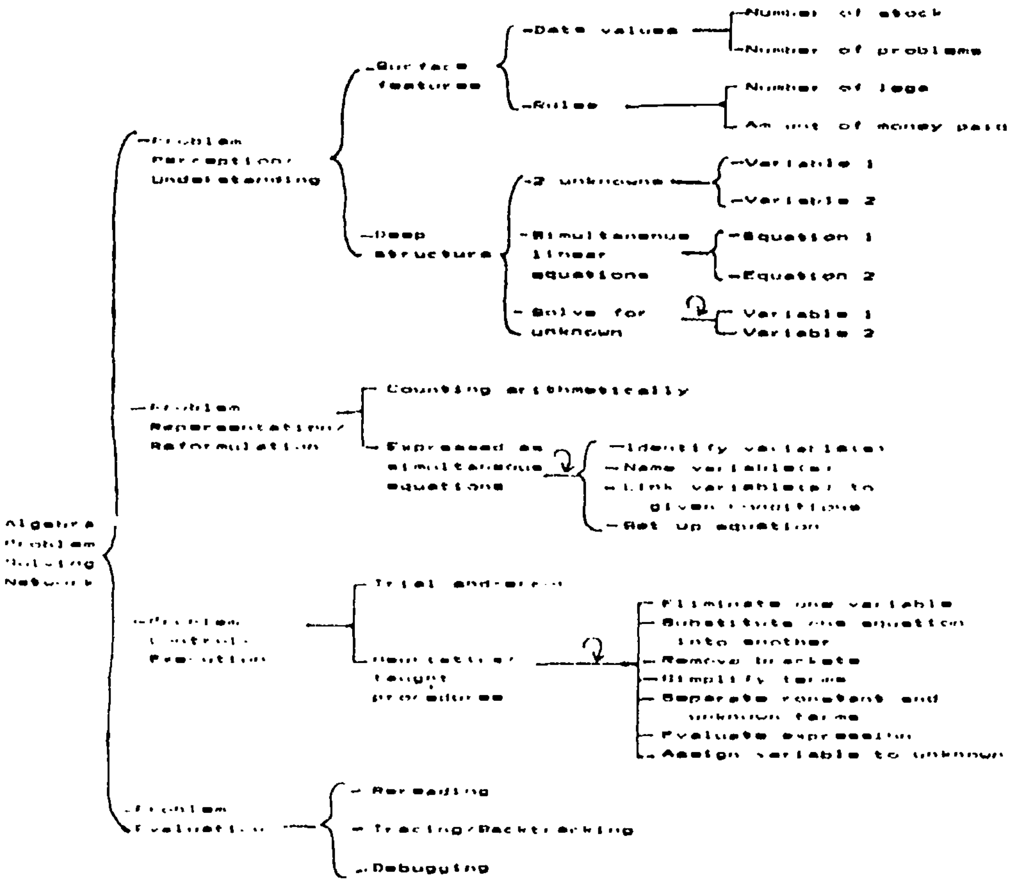


Figure 4.1 Common 'Deep Structure' of Items 10 and 11.

are determined. These levels are shown in Table 3.2, where their calibration results are also presented and discussed later. It should be noted that the test being of routine exercise-type enables precise definitions of item performance score levels although it is observed that some pupils use trial and error methods when they fail to represent questions 10 and 11 for an algebraic solution. This issue of quantitative modelling of alternative solution routes, an educational more than a statistical problem for researchers, remains unresolved if a combined analysis is to be sought.

The use of partial credit modelling in the construction of the algebra problem-solving proficiency continuum

Partial Credit Modelling (Masters, 1982) was used to analyse the pupils' ordered, graded responses of items according to a scoring scheme so that partial credits for partial success are assignable to items that can be mapped on a unidimensional continuum. Partial Credit Modelling permits the maximum item scores, which reflect the highest performance levels attainable of pupils, to vary across items. In addition, it allows the structure of competencies or procedural steps that these performance score levels represent to vary across items. The proposed scoring scheme mentioned earlier serves precisely this purpose of grading pupils' responses into an ordinal scale of performance levels. Table 3.1 presents a frequency distribution of performance levels of test items based on pupils' responses that have been graded according to the proposed scoring

Table 3.1 Frequency Distribution of Performance Levels Across Items

Item	Item Performance Level (Item Score) 1 (0)	2 (1)	3 (2)	4 (3)	5 (4)
1	10	120			
2	29	40	61		
3	2	13	115		
4	17	13	100		
5	19	14	30	67	
6	22	108			
7(i)	24	4	102		
7(ii)	33	6	91		
8	55	8	67		
9(i)	73	11	9	37	
9(ii)	18	6	106		
10	80	7	35	4	4
11	45	6	46	11	22

scheme.

The outcomes of the Partial Credit Modelling, if the model assumptions and requirements are met, are sets of pupil ability and item step difficulty parameter estimates on a common linear logit scale. Table 3.2 shows the item step difficulty estimates and their associated fit statistics. Figure 3.2 presents the calibrated ability distribution of the pupils in the sample. The various item steps of the algebra test items demonstrate excellent statistical fit to the Partial Credit Model and they span a long range of at least six logits (item step difficulties estimates range from -3.27 to 2.95) with the average difficulty of the item steps being arbitrarily set at zero logit. Judging from the ability distribution of the pupils, it can be concluded that the test has been appropriately targeted at the pupils in the calibrated sample because the modal region of the roughly normal distribution is approximately at 0.8 logit. All of these statistics demonstrate the internal validity of the algebra test.

The item step difficulty estimates are not necessarily ordinal because they are response thresholds indicating the difficulty of proceeding from one given performance level to the next when a pupil is already at a given level. Notwithstanding this, a pupil's ability logit estimate generally increases with increasing performance score levels. It is noteworthy that some intermediate performance score levels are less likely to obtain because once the cognitive or procedural barriers have been surmounted the ensuing steps simply follow. It is this "step" property which renders

Table 3.2 Item Estimates and Their Fit Statistics

Description of item step	Step Difficulty Estimate	Precision (Standard error)	Fit Statistics v	q	t
1.Simpl fy terms	-2.11	.34	1.00	.27	.10
2.Remove bracket	.17	.23			
Simplify terms	.13	.19	.93	.17	-.74
3.Substitute values of a,b,c	-2.11	.76			
Evaluate expression	-1.78	.29	1.21	.26	.82
4.Separate constant & unknown	.2?	.29			
Solve fr · y	-1.54	.23	.76	.16	-1.66
5.Multiply out brackets	.17	.29			
Add 4x+E to 2x	-.50	.24			
Solve for x	-.20	.20	.96	.12	-.30
6.Translate twice of x to 2x	-1.22	.25	.93	.15	-.45
7i)Know relationship between variables from word sentence	2.38	.26			
Write algebraic expression	-3.27	.25	.73	.15	-2.04
7ii)Know relationship between variables from word sentence	2.12	.23			
Write algebraic expression	-2.49	.22	.91	.11	-.81
8.Intepret 5m and 2p	2.30	.21			
Interpret 5m + 2p	-1.67	.20	1.07	.09	.77
9i)Know relationship between variables from word sentence	1.90	.21			
Write overtime pay as 2h	.90	.23			
Equate w=500 + 2h	-.51	.24	1.08	.11	.69
9ii)Calculate overtime pay	1.22	.29			
Calculate total wage	-2.52	.26	.92	.17	-.46
10.Write first equation	2.82	.21			
Write second equation	-1.04	.22			
Solve equations partially	2.95	.39			
Solve equations completely	1.26	.55	1.18	.13	1.34
11.Write first equation	2.30	.22			
Write second equation	-1.80	.22			
Solve equations partially	1.98	.24			
Solve equations completely	.30	.27	1.01	.11	.14

Scale Performance Level	Raw Score	Ability (Logits)	Frequency
		5.0	
		4.8	
		4.6	
		4.4	
		4.2	
		4.0	
4		3.8	
		3.6	
		3.4	
		3.2	
		3.0	
		2.8	
		2.6	
		2.4	
		2.2	XX
		2.0	XX
		1.8	
	26	1.6	XXXXXXX
	25	1.4	XXXXX
3		1.2	XXXXXXXXXXXXXX
		1.0	XXXXXXXXXX
		.8	XXXXXXXXXXXXXλXXXX
		.6	XXXXXXXXXXX
		.4	XXXXXXXXXXXX
	16	.2	XXXXXXXXXXXXXXXX
2		.0	XXXXXXXXXX
		-.2	XXXXXXXXXXXX
		-.4	XX
		.6	XXXX
		-.8	
	6	-1.0	XX
		-1.2	XX
1		-1.4	
		-1.6	
		-1.8	
		-2.0	
		-2.2	X
		-2.4	
		-2.6	
		-2.8	
		-3.0	

Figure 3.2 Ability Distribution of Pupils

Partial Credit Modelling to be different from those graded response models that are based on "category boundary" approach, of which it must always be easier to reach a low performance level than the higher ones. By considering the odds of passing from a given performance level to the next and equating these odds to a set of logistic probability curves, the Partial Credit model is actually an extension of the Rasch and Rating Scale model in the number of steps taken and handling of inordinal item step difficulties. Consequently, this measurement procedure satisfies all the requirements such as specific objectivity and linearity of a measurement model (Cheung, 1990) and meaningful measurement such as those exemplified by Koh and Cheung (1990) is also achievable.

The item step difficulty estimates of all the test items help to define the progressive, qualitative aspects of the constructed proficiency continuum on the solving of algebra word problems. Special care should be taken to interpret these estimates because they indicate the response thresholds of completing those steps that are associated with the various performance score levels, and not the actual ability level of the known competencies that may be applied within the problem-solving sequences. Thus, the problem-solving sequences of the test items as explicitly reflected in the scoring scheme, and the patterns of response thresholds particularly those of the rate-determining steps, are important considerations in establishing the external validity of the problem-solving proficiency continuum.

This content analysis is summarised in Figure 3.3 as an item map, indicating the probability of more than 0.5 of scoring the performance levels 1 and above rather than 0, 2 and above rather than 0 and 1, and so on across all items that have been arranged in an order along the proficiency continuum. This continuum is found to match with the design properties of the test. The actual probabilities are not shown in this item map. Instead, separate "zone" maps of each item can be drawn, indicating the probability of scoring at the various performance levels. Figure 3.4 shows an example zone map of item 8, of which the first step is the rate-determining step, so that the probability of scoring one is negligible. In passing, one may want to know that the item map can be obtained from the set of zone maps by drawing a line through "probability = 0.5" on each zone map to obtain the most probable performance levels along the ability logit continuum.

Quantitative measurement with qualitative interpretations using the proficiency continuum

Items 1, 3 and 6 help to define the lowest portion (scale performance level = 1, up to -0.9 logits) of the continuum. They involve simplifying "2a+5b+a" to "3a+5b" for item 1, substituting the values of a, b, and c into "3a-b+5c" and evaluating this expression correctly for item 3, and translating word sentences into algebraic expression "2x" to indicate "twice of x" for item 6. The response thresholds of these steps in logits are -2.11, -2.11 and -1.78, and -1.22 respectively for the three items.

Items 4, 7(i) and (ii), and 9(ii) contribute to a definition of

Scale Performance Level	Raw Score	Progressive Forms Of Knowing	Ability (Logits)	10	9i	11	8	2	5	7ii	7i	9ii	4	6	3	1
		Know relationships	3.0	4	3	4	2	2	3	2	2	2	2	1	2	1
		between variables	2.8	4	3	4	2	2	3	2	2	2	2	1	2	1
4		from word problem.	2.6	4	3	4	2	2	3	2	2	2	2	1	2	1
		Write equation to	2.4	4	2	4	2	2	3	2	2	2	2	1	3	1
		represent problem.	2.2	3	3	4	2	2	3	2	2	2	2	1	2	1
		Solve simulta-	2.0	3	2	4	2	2	3	2	2	2	2	1	2	1
		neous equations.	1.8	2	3	4	2	2	3	2	2	2	2	1	2	1
			1.6	2	3	4	2	2	3	2	2	2	2	1	2	1
	25	Remove bracket &	1.4	2	3	3	2	3	3	2	2	2	2	1	2	1
		simplify terms.	1.2	2	3	3	2	2	3	2	2	2	2	1	2	1
		Solve equation	1.0	2	2	2	2	2	3	2	2	2	2	1	2	1
3		with brackets &	.8	1	2	2	2	2	3	2	2	2	2	1	2	1
		negative solution.	.6	0	1	2	2	2	3	2	2	2	2	1	2	1
		Interpret algebraic	.4	0	0	2	2	1	3	2	2	2	2	1	2	1
		expression.														
	16	Seperate constant	.2	0	0	2	0	1	2	2	2	2	2	1	2	1
		from unknowns to	.0	0	0	0	0	1	2	2	2	2	2	1	2	1
2		solve equation.	-.2	0	0	0	0	1	2	1	2	2	2	1	2	1
		Understand alge-	-.4	0	0	0	0	1	1	0	2	2	2	1	2	1
		braic relationship	-.6	0	0	0	0	0	1	0	0	1	1	1	2	1
		between known &	-.8	0	0	0	0	0	0	0	0	0	1	1	2	1
		unknown terms.														
	6	Simplify terms.	-1.0	0	0	0	0	0	0	0	0	0	0	1	2	1
		Evaluate expres-	-1.2	0	0	0	0	0	0	0	0	0	0	1	2	1
		sion.	-1.4	0	0	0	0	0	0	0	0	0	0	0	1	1
		Translate simple	-1.6	0	0	0	0	0	0	0	0	0	0	0	1	1
1		sentence to	-1.8	0	0	0	0	0	0	0	0	0	0	0	1	1
		algebraic expres-	-2.0	0	0	0	0	0	0	0	0	0	0	0	1	1
		si .	-2.2	0	0	0	0	0	0	0	0	0	0	0	1	0
			-2.4	0	0	0	0	0	0	0	0	0	0	0	1	0
			-2.6	0	0	0	0	0	0	0	0	0	0	0	0	[illegible]
			-2.8	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 3.3 Item Map of the Algebra Test

```
 2.5 1122222222222222222222222222222222222222222222222222222222222
 2.4 12..........................................................2
 2.3 012.........................................................2
 2.2 012.........................................................2
 2.1 012.........................................................2
 2.0 01*2........................................................2
 1.9 0X12........................................................2
 1.8 0X1*2.......................................................2
 1.7 0XX1*2......................................................2
 1.6 0XXX1*2.....................................................2
 1.5 0XXX1**2....................................................2
 1.4 0XXXX1**2...................................................2
 1.3 0XXXXXX1*2..................................................2
 1.2 0XXXXXXX1**2................................................2
 1.1 0XXXXXXXX1**2...............................................2
 1.0 0XXXXXXXXXX1**2.............................................2
  .9 0XXXXXXXXXXXX1**2...........................................2
  .8 0XXXXXXXXXXXXXX1***2........................................2
  .7 0XXXXXXXXXXXXXXXX1***2......................................2
  .6 0XXXXXXXXXXXXXXXXXXX1***2...................................2
  .5 0XXXXXXXXXXXXXXXXXXXXX1***2.................................2
  .4 0XXXXXXXXXXXXXXXXXXXXXXXX1***2..............................2
  .3 0XXXXXXXXXXXXXXXXXXXXXXXXXXX1***2...........................2
  .2 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1***2........................2
  .1 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1**2......................2
  .0 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1***2...................2
 -.1 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1**2.................2
 -.2 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1***2..............2
 -.3 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1***2............2
 -.4 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1***2..........2
 -.5 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1**2.........2
 -.6 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1**2.......2
 -.7 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1**2......2
 -.8 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1*2.....2
 -.9 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1*2....2
-1.0 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1*2...2
-1.1 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1*2...2
-1.2 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1*2..2
-1.3 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX12..2
-1.4 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1*2.2
-1.5 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX12.2
-1.6 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX12.2
-1.7 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX12.2
-1.8 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX122
-1.9 0XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX122
-2.0 0000000000000000000000000000000000000000000000000000000000000
      1     0     .1     .2     .3     .4     .5     .6     .7     .8    .9
                                   PROBABILITY
```

Figure 3.4 Zone map for Item 8

the next lower portion (scale performance level = 2, -0.9 to 0.3 logits) of the continuum. Amongst these three items in four parts, the first steps are all critical (response thresholds are 0.22, 2.38, 2.12 and 1.22 respectively) and the ensuing steps simply follow (-1.54, -3.27, -2.49 and -2.52 respectively). These critical steps include separating the constant and unknown terms of "4y-8=2y-4" into "2y=4" for item 4, understanding the two algebraic relationships of the known and unknown from word sentences for items 7(i) and 7(ii), and forming the algebraic equation "w=500+2*10" for item 9(ii). The ensuing steps respectively involve solving for "2y=4", coding correctly expressions "20-x" and "x-y", and calculating the values of "500+2*10". With hindsight, one can see the flexibility and power of Partial Credit Modelling in treating what should constitute a problem-solving step. For example, one can easily regard the ensuing steps as part of the package of critical steps. Thus, these ensuing steps are not explicitly estimated without any major consequential effects on the overall qualitative structure of the proficiency continuum.

Items 2, 5 and 8 contribute to an understanding of the higher portion (scale performance level = 3, 0.3 to 1.5 logits) of the continuum. The response thresholds of the two steps of item 2 are very close (-0.17 and 0.11) and the completion of which indicates a person is progressing from scale performance level 2 to 3. These two steps involve grouping the like terms together by removing the brackets of "(8y+6z)-(4y+3z)" and simplifying this expression to form "4y+3z". Similarly, the response thresholds of the three steps of item 5 are very close, with the first step a little bit relatively more difficult (0.17, -0.50, -0.20). The completion of these steps prepares one for entering into scale performance level 3. The steps include multiplying "4(x+2)" to form "4x+8", then adding "4x+8" and "2x" to form "6x+8" and finally solving "6x+8=4" to give "x=-2/3". Item 8 is a two-step item with a difficult first step and easy second step (2.12 and -2.49). The first step involves an interpretation of "5m" and "2p" in a mathematical rather than informal qualitative way, and the second step an interpretation of the algebraic expression "5m+2p" as the cost of 5 mangoes and 2 papayas.

Completion of items 9(i), 10 and 11 shows the clearest indication that a person has attained scale performance level 4 (greater than 1.5 logits). The response thresholds of the three steps of item 9(i) are 1.90, 0.90, and -0.51 respectively, indicating that the first step is the most difficult step. This item requires the pupils to understand the relationship between total wage, basic wage, and overtime pay, and then express overtime pay as part of the total wage, followed by equating "w=500+2h". The item response behaviours of items 10 and 11 are similar, with each of the steps of item 10 more difficult than those of item 11 because the latter is a standard example in the textbook. The response thresholds are 2.82, -1.04, 2.95 and 1.26 for item 10, and 2.30, -1.32, 1.98 and 0.30 for item 11. As revealed, the first step of both items, which requires the pupils to translate the word sentences into algebraic equations, are the rate-determining steps.

The second step of both items, which are relatively straightforward once the first steps are completed, seeks to supply a second equation for the simultaneous equations to complete the problem representation. The last two steps involve substituting one equation into another and solving for the two unknowns accordingly. It should be noted that representing the problems in an algebraic way and solving a system of simultaneous equations in two unknowns are difficult tasks for the pupils.

A point noteworthy of documentation is the attempt in this study to incorporate those successful problem solvers who obtained their answers using trial and error (approximately 25% of the sample) into the Partial Credit Modelling procedures. This is done by grading these responses at performance level 2. While this practice warrants further examination in Chapter 4 and is unlikely to be resolved on statistical grounds only, the internal validity of this proficiency continuum appears credible because of the conformity of the response data with the stringent measurement model. It is anticipated that pupils' responses obtained from different solution paths may need to be modelled separately. They may be combined only if the item response behaviours are statistically conformable to the Partial Credit model and the corresponding performance levels of the alternative solution paths as laid down in the scoring schemes can be educationally justified.

Conclusion

This paper demonstrates the possibility of modelling problem solving tasks of the routine exercise-type so that quantitative measurement with qualitative interpretation is possible. This is a first but significant step towards meaningful assessment of problem-solving tasks laid down in Chapter 1 of this monograph. At least two other issues need to be resolved. First, the various solution paths may not be equated for the assignment of performance score levels because of the different structures of competencies involved and uses of heuristic strategies during the problem-solving process. It is also because of the varying degrees of opportunity to learn of the pupils. Second, there are metacognitive and attitudinal aspects of problem solving informing how errors, false starts, detours, and blind alleys have come about. These information, if integrated with the problem-solving proficiency continuum, makes meaningful measurement of problem-solving activities more complete. Some of these are examined in Chapter 4. The findings on the difficulties of solving algebra word problems in this chapter not only substantiate those in the literature, but also statistically demonstrate that problem-understanding and problem-representation are critical rate-determining steps in problem solving, even if the problems are of the routine-exercise type.

CHAPTER 4

On Meaningful Measurement: Metacognition and Hierarchical Modelling of Errors in Algebra Word Problems

K C CHEUNG & LOH We Fong

Some contemporary views of metacognition on mathematical problem solving

When a person is actively engaged in an attempt to solve problems, particularly non-routine problems, there are meta-level decision-making processes which are most likely to occur between transitional episodes of the problem-solving process, indicating the occurrences of reflective, regulatory, evaluative and control behaviours (Schoenfeld, 1983). Biggs (1987) summarised six kinds of metacognitive knowledge: (1) content-knowledge of what one is talking about; (2) self-knowledge of what one already knows or does not know; (3) fore-knowledge of purposes and goals; (4) situational awareness of the ways things are; (5) progress-knowledge of where one is going or changing course; and (6) strategic-knowledge of alternative ways to work best in circumstances.

Kilpatrick (1967) pioneered the use of protocol coding schemes to examine cognitive processes. Since then, the use of verbal reports such as those obtained from the "think-aloud" procedures provides legitimate data for analysing problem-solving behaviours (Ericsson and Simon, 1980). Patterns of behaviours and streams of events of these ongoing cognitive processes that come to conscious attention, as opposed to those obtained through introspection and retrospection, are analysed into sequences and episodes for explanation through processes of metacognition and affection. From an information theorist's standpoint, it is essentially those psychological data in the short-term memory, and the problem-solving processes carried out in the central processor that are the focus of the study.

Polya's famous book "How to solve it ?" provides a general four-phase model consisting of heuristics "Examine-Plan-Do-Check" of the problem-solving process on mathematical problems in a variety of contexts. Recently, Schoenfeld (1987) developed four within-classroom techniques for the teaching of heuristic strategies with a metacognitive focus: (1) using videotapes to create awareness of the thinking and metacognitive processes, (2) teacher as a role model for metacognitive behaviour, (3) teacher as a moderator rather than an arbitrator in whole class problem solving, and (4) teacher as a facilitator in small group problem solving. Despite all these efforts, a viable full-fledged pedagogy of problem solving is not available mainly because a theory of problem solving is not fully developed yet. Chapter 2 of this monograph is an attempt to provide some ideas for a more adequate theory. It is hoped that the findings of researches and teaching experiments can inform us of a better pedagogy for use in the classroom.

Purpose of study

This study, a continuation of Chapter 3 of this monograph, seeks to analyse both qualitatively and quantitatively the key types of errors made by the pupils at the different performance levels of the problem-solving proficiency continuum. For the purpose of this study, five items along this proficiency continuum are selected. They represent a list of concepts, procedural skills, and key types of errors on solving algebra word problems. The key types of errors and the bands of performance levels of solving algebra word problems are then modelled using dual scaling, a multidimensional scaling procedure, to ascertain the dimensional and hierarchical nature of errors. Metacognitive behaviours are examined to explicate the problem and search spaces of the problem-solving processes. Guidelines for a pedagogy of solving algebra word problem are also presented.

Error analysis of selected algebra word problems

Error analyses of five test questions (3, 4, 5, 8 and 10) along the algebra problem-solving continuum are reported here although a full-scale analysis of the whole test can be done as well. Most of the errors on these five items are made by pupils with scale performance levels 2 and 3 (87% of the sample). However, there are some key types of errors that are made more often by pupils at one level than the others. At the same time, some types of errors are committed by pupils at the four levels of performance. For the purpose of this study, key types of errors are those errors viewed by the present authors as educationally significant and are committed by a large number of pupils in the sample. Table 4.1 shows how the key types of errors are distributed across the four performance levels of the problem-solving continuum.

Item 3 (scale performance level 1):

What is the value of 3a-b+5c if a=2, b=3 and c=5?

Nearly 90% of the pupils answer this question correctly (60%, 91%, 90%, 91% respectively at the four levels of performance). The key error is that pupils at levels 2 and 3 (8% and 10% of pupils at the two levels respectively) evaluate the numerical values wrongly after substituting the values of a, b and c into the expression correctly. The response thresholds of the two steps of substitution and evaluation are -2.11 and -1.78 logits respectively and this key error occurs in the slightly more difficult second step of evaluation.

Item 4 (scale performance level between 1 and 2):

Given that 4y-8=2y-4, find the value of y?

Nearly 80% of the pupils answer this question correctly (0%, 64%, 93%, 100% respectively at the four levels of performance). The key error is committed at the second step where pupils at performance levels 2 and 3 (16% and 6% of pupils at the two levels respectively) are required to simplify the constant and unknown terms. This is done after the first step of grouping the terms to either sides of the equality symbol in order to solve the equation. The response thresholds of the first step of rearranging the

Table 4.1 Cross-tabulation of Key Types of Errors Against Scale Performance Levels

Scale Performance Level	Item 3		Item 4		Item 5			Item 8			Item 10			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
4 (n=11)	10	1	11	0	10	0	1	10	1	0	5	4	0	1
3 (n=70)	63	7	65	4	48	7	6	50	6	1	3	19	8	23
2 (n=44)	40	3	28	7	11	4	5	13	10	12	0	8	5	11
1 (n=5)	3	0	0	0	0	1	0	0	2	3	0	1	0	2
Total	116	11	104	11	69	12	12	73	19	16	8	32	13	37

A: No error
B: Error in evaluation
C: No error
D: Error in simplifying terms
E: No error
F: Simplify instead of solve
G: Ignore negative sign in solution
H: No error
I: Treat variable as object
J: Omitted
K: No error using algebraic method
L: No error using trial and error method
M: Use wrong equation
N: Use superficial cues

terms, and the second step of simplifying and solving the equation are 0.22 and -1.78 logits respectively.

Compared with the second step of item 3, the second step of item 4 invited more pupils at level 2 to commit errors although the step difficulties are comparable (both at -1.78 logits). This is because those who can complete the first step of item 4 will tend to find the ensuing step relatively easy (the first step is at a relatively higher threshold than the second step). In the sample, none of the pupils at level 4 makes this key type of error after successfully grouping the constant and unknown terms. The same occurs for pupils at performance level 1 because they cannot even complete the first step which for them is relatively difficult. This error pattern of the high and low ability pupils is consistent with the relatively higher step difficulty of the first step of grouping terms than the second step of simplification and evaluation.

Item 5 (scale performance level between 2 and 3):

Solve the equation $4(x+2)+2x=4$.

Nearly 52% of the pupils solve this equation correctly (0%, 25%, 69%, 91% respectively at the four levels of performance). Amongst the operations of removing brackets, grouping for constant and unknown terms, and solving for the unknown, two key errors stem out that are unique to the given equation. The first is that while they have no difficulty in understanding "find the value of ..." they cannot understand the technical word "solve". As a result, pupils stop at the first two steps of removing brackets and simplifying.

20%, 9%, and 10% of pupils at performance levels 1, 2 and 3 make this error. The second error stems from the uneasiness of "x" taking the value of a negative number. Pupils simply ignore the negative sign in the solution. 11% and 9% of pupils at levels 2 and 3 commit this error.

These two types of errors are characteristics of pupils of level 3 and have nearly accounted for all the errors made by pupils at this level. Also, it is observed that these two types of errors are seldom made by pupils of performance level 4 and 1. Clearly, pupils at performance level 1 cannot even remove the brackets to solve for the unknown. The response thresholds of the three steps of removing brackets, grouping and simplifying, and solving are 0.17, -0.50, and -0.20 logits respectively. The two key errors occurred at the second and third steps respectively and these are of higher step difficulties than those key errors of items 3 and 4.

Item 8 (scale performance level between 3 and 4):
Mangoes cost m cents each and papayas cost p cents each. If I buy 5 mangoes and 2 papayas, what does 5m+2p stand for?

Nearly 57% of the pupils have no problem of understanding the algebraic expression correctly (0%, 30%, 71% and 91% respectively at the four levels of performance). The key error is that pupils misinterpreted "5m" as 5 mangoes and "2p" as 2 papayas. Thus, "5m+2p" is interpreted qualitatively as the number of mangoes and papayas bought. 40%, 23%, 9%, and 9% of pupils of performance levels 1 to 4 make this error.

Furthermore, a related type of error is that 60% and 27% of pupils at performance levels 1 and 2 do not respond to this question, showing that they have difficulty in comprehending the given expression. Once the first step of understanding what "5m" and "2p" mean the second step of stating that the expression stands for the cost of fruits is relatively straightforward. The response thresholds of these two steps are 2.30 and -1.67 logits respectively. The high response threshold of the first step of understanding clearly leads to an expectation that there is a higher proportion of pupils at performance levels 1 and 2 to treat variables "m" and "p" as objects or to omit the question entirely.

Item 10 (scale performance level 4):

A mother promised to pay her son 20 cents for every math problem that he got right and fined him 5 cents for each wrong answer. After 10 problems, the mother had to pay the boy $1.50. How many problems did the boy answer correctly?

Only 31% of the pupils can solve this algebra word problem adequately (20%, 18%, 31%, and 82% respectively at the four levels of performance). Two methods of problem solving are evident, namely, the algebraic metnod and the trial and error method. The latter kind is four times more popular for those successful problem solvers (6% and 25% comprising the 31% of successful problem solvers). There are comparable chances for successful problem solvers at performance level 4 to deploy the algebraic method (56%) than the trial and error method (44%). However, successful problem solvers at levels 2 and 3 cannot use algebraic method but need to use trial and error for problem solution (0% and 14% compared to 100% and 86% at levels 2 and 3 for the two methods).

Consequently, pupils at intermediate performance levels can still use heuristics to search for a viable solution whereas those high-ability pupils can make use of the "deep structure" of the question and its associated knowledge base to attempt a solution in an algebraic way. This assertion is further supported by the greater percentages of pupils of performance levels 1, 2 and 3 using only part of the question and superficial cues for their problem solving (40%, 25%, 33% and 9% respectively at the four levels of performance).

Another type of error is made by pupils of performance levels 2 and 3 only (11% and 33% of pupils at the two levels respectively). It occurs because although pupils understand the problem correctly and have the know-how to solve for a solution they make a mistake in representing the problem using the wrong algebraic equation. Pupils at performance level 1 do not commit this error because they cannot understand the problem at all whereas pupils at performance level 4 find the problem-solving procedures straightforward once they have understood the problem. These explain the inordinal pattern of the response thresholds of the four steps of problem-understanding, problem-representation, and the two steps of solving of the system of two simultaneous equations of unknown "x" and "y" (2.82, -1.04, 2.95 and 1.26 logits respectively). The third step is noticeably more difficult because it involves solving equations of two unknowns.

The findings of the error analyses on the relationships between key types of error and problem-solving ability are summarised below:

1. For pupils at all the four levels, particularly the first three levels, they rely on superficial cues when they do not understand a problem. In particular, pupils at levels 2 and 3 will proceed to attempt a solution when some superficial cues can get them going.

2. For pupils at level 4, they can use algebraic methods to solve word problems, whereas for pupils at levels 3 and 4 they can make use of other heuristics such as trial and error if they have understood the problem but fail to proceed using the algebraic method. Pupils at levels 1 and 2 treat variables as concrete objects so that they usually have difficulties in problem-understanding and problem-representation in an algebraic way. Pupils at these two levels very often cannot even proceed to the problem-execution phase of using the various algebraic skills and they use heuristics sparingly.

3. For pupils at levels 2 and 3, they show a confusion between terminologies such as "solve" and "simplify" but do not have difficulties when a simple phrase such as "find the value of" has been used instead. They tend to make errors in simplifying and evaluating algebraic expressions and cannot admit variables which take on negative values. The solving of simultaneous equations

involving two unknowns is also well beyond their abilities.

Relationships between key types of errors and problem-solving ability

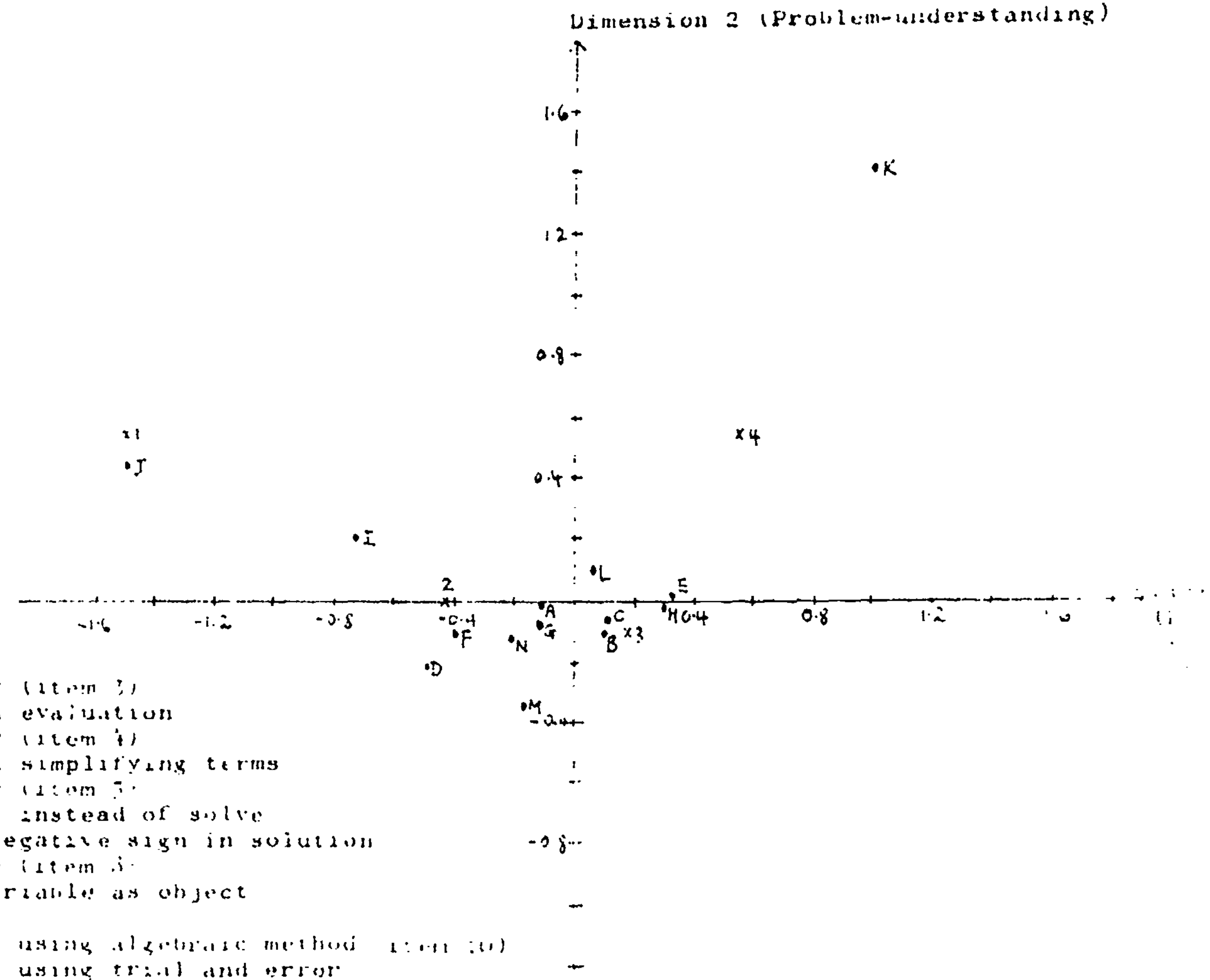
This section attempts to model statistically the key types of errors of the five selected items along some structural dimensions that are parallel to that of the problem-solving proficiency continuum. Latent trait modelling such as the Partial Credit Modelling cannot be used because of the unrealistic assumption that these key types of errors can be aligned onto a unidimensional trait. Instead, analytical tools that do not rely on this unidimensionality assumption are needed. Furthermore, it should be capable of modelling frequency counts of key types of errors and successful completion of items as those in Table 4.1. Dual scaling (Nishisato, 1980) is a versatile method for the exploration of the hidden structure of categorical data such as two-way contingency tables and quantification of row and column categories along parsimonious structural dimensions. Statistically, it decomposes a table into row and column structural dimensions, with optimal weights assigned to row and column categories along these dimensions, such that both the between-row and between-column discriminations are simultaneously maximised.

These discrimination indices are expressed by the "squared correlation ratio" following Guttman's principle of internal consistency (Guttman , 1941) in minimising within-subject (row or column) discrepancies and maximising between-subject (row or column) differences in values of optimal weights. The method resembles a multidimensional decomposition of data with the most "informative" structural dimension extracted first, then the second most "informative" dimension, and so on, until the information in the data are exhaustively extracted. From the point of view of dual scaling, when the two dimensions of categories spanning the contingency table are independent of each other, there exists no structural dimension in the data because the information level of this table is zero. Consequently, associated with each structural dimension is a statistic "Delta Partial", indicating the percentage of information explained by that dimension. The optimal weights are further weighted to reflect this relative importance of structural dimensions for comparing amongst structural dimensions. These weighted optimal weights of a structural dimension can be plotted against another revealing the underlying structure of the data table.

Table 4.1 provides all the necessary input for dual scaling. The two dimensions of this contingency table are firstly the four scale performance levels, and secondly the key types of errors together with the categories of counts on the successful completion of the five selected items along the proficiency continuum. Three structural dimensions result from the dual analysis, accounting for 72%, 21% and 7% of the information in the data. The weighted optimal weights of the categories of the two dimensions of Table 4.1 can be plotted along the first and second structural dimensions, showing both the hierarchical and dimensional properties of the key types of error, stages or performance levels, and ability levels in the successful completion of the test items. This plot is shown in Figure 4.1.

The first structural dimension clearly is the proficiency continuum upon which key types of errors are structured. Examinations of the coordinates J, I, D, F, N, M, G, B relative to coordinates 1, 2, 3, 4 recapitulate exactly the findings on how pupils of different performance levels may have the opportunities to commit the various key types of errors. It demonstrates that by making reference to the first structural dimension, those "inabilities" as defined by the key types of errors can be hierarchically structured in the same way as the progression of knowing/procedural skills in solving algebraic word problems. This is revealed by coordinates A, C, E, H, K that they are sequenced along the scale performance continuum by outcome. It should be noted that only if a pupil is well past performance level 2 will s/he successfully complete questions 3, 4, 5 and 8. This is in accordance with the probability of pupils of the four performance levels passing the respective items. Also, in order to successfully solve the algebra word problem (item 10), a pupil need to be in performance level 4.

Furthermore, a pupil needs to be able to understand and represent a problem, as suggested by the coordinates K, 4, J and 1 interpreted along the second minor structural dimension. The position of L shows that it is mostly those level 3 pupils, who



A:no error (item 3)
B:error in evaluation
C:no error (item 4)
D:error in simplifying terms
E:no error (item 5)
F:simplify instead of solve
G:ignore negative sign in solution
H:no error (item 5)
I:treat variable as object
J:omitted
K:no error using algebraic method (item 10)
L:no error using trial and error
M:use wrong equation
N:use superficial cues

Figure 4.1 Hierarchical Structure of Errors in relation to Levels of Algebra Problem-solving Ability.

63

possess most of the basic algebraic skills, when faced with problems in representing the problem situations algebraically will deploy the trial and error methods. The first structural dimension thus constitutes the right arm of the competence ladder which is to be paired with the left arm of the problem-solving proficiency continuum described in the last chapter. The four performance levels act symbolically as the rungs linking the progression of the problem-solving steps on one arm, to the hierarchical structure of key types of error on the other. This ladder shows a genuine realisation of quantitative assessment of problem-solving tasks with meaningful qualitative interpretations on both the structure of problem-solving steps and key types of errors that are firmly rooted in the literature described earlier.

A metacognitive exploration of the problem and search spaces of problem solvers

Space does not allow a detailed reporting here of how metacognitive decision making and attitudinal behaviours affect pupils engaged in problem solving. Instead, four excerpts from two pupils (P1 and P2) solving two algebra word problems (A and B) in front of the researcher (T) are discussed. These excerpts illustrate the contrasting problem and search spaces of the problem solvers. Problem A is a standard textbook question asking for the number of cows and chickens given the total number of the stock and the total number of legs. Both pupils do not have problems in understanding the question, knowing that cows have 4 legs whereas chickens have

2 legs.

Pupil 1 simply does not have the slightest incentive to attempt an answer when he cannot recognise a to proceed. Pupil 2 plunges into solving the problem quickly by considering only the total number of legs. His method is faulty, leading him to an incorrect answer. He is not aware that his answer on the total number of animals is different from that given in the question. This shows that at this stage he is still using partial information in order to proceed with a solution. However, he recognises that he has not reached the goal yet because he does not know the number of cows. He does not know what has gone wrong and is looking for an alternative method. He remembers that if he can write two equations to represent the problem he may solve the problem in an algebraic way. Clearly, pupil 2 is equipped with a domain-specific knowledge base for him to categorise some familiar problems and to solve the simultaneous equations of two unknowns. Despite his success, he fails to evaluate why his initial attempt has not worked.

Excerpt A1:

P1: A woman has altogether 50 cows and chickens. These animals have a total of 172 legs. How many cows are there?

T : Do you understand what the question is asking you?

P1: [nods her head] Yes.

T : How will you answer this question?

[pause]

What is the question asking you?

P1: How many cows are there?

T : How will you solve it?

P1: [pâuse a while]

I only know cows have 4 legs and chickens have 2 legs.

T : What else do you know?

P1: I don't know how to do it.

T : You said you understand the question ...

P1: Yes, I know what it's asking, hmm ..., but I don't know how to find the answer.

T : You have no idea at all?

P1: No.

Excerpt A2:

P2: A woman has altogether 50 cows and chickens. These animals have a total of 172 legs. How many cows are there?.

T : Do you understand what the question is asking you?

P2: It just says that a woman has altogether 50 cows and chickens. These animals have a total of 172 legs.

A cow has 4 legs, chicken 2 legs, that is 6 legs.

I think 172 divide by 6 ..., that's 27 animals.

But I'm not sure how many cows and how many chickens yet.

T : Why do you divide 172 by 6?

P2: The total number of legs of the animals.

T : You obtain 27 as the answer ...

P2: Yes, but I still do not know how many cows and chickens are there.

May be I'll represent x for the cows and y for the chickens. So x plus y equals 50.

Then 4 plus 2 equals ... [pause]

T : What are you trying to do now?

P2: I'm trying to write another equation.

T : So you're trying to use another method now.

Why don't you want to use the first method?

P2: May be that is the same as this one.

May be I can get the answer here also.

T : Okay, how would you proceed?

P2: x is cow ..., so 4x plus 2y equals 172.

I'll substitute this as one [pointing to the first equation : x+y=50] and that as two [referring to the equation : 4x+2y=172].

[proceeded with solving a set of simultaneous equations].

I get x as 36 ..., so there are 36 cows.

Problem B is also thoroughly understood by both pupils, although the pupils in the sample may not have encountered similar questions before. This time, pupil 1 attempts an answer by making use of all pieces of given information in the question. Since he lacks algebraic skills, he can only proceed in an arithmetical way. He starts by assuming that the son gets the 10 problems right, and ultimately through some more sensible steps, he obtains a contradictory result that the son gets the 10 problems wrong. He is aware of this discouraging result and then gives himself up after he fails to see another way to proceed.

Pupil 2 cannot obtain an answer this time because problem B is less familiar to him. He attempts an answer in an arithmetical way first and can go no further. The trouble is that the question is so well-designed that pupils will realise whether they are wrong or not. Realising that he is wrong, pupil 2 attempts to solve the problem in an algebraic way but in vain. He is confused but perseveres because he knows that if he can write two equations he can obtain the answer He allows himself a final attempt by rereading the question and studies some special cases, hoping by using trial and error he may see a structure in the problem to obtain an answer. All these attempts are fruitless and he has to admit that he cannot answer.

Excerpt B1:

P1: A mother promised to pay her son 20 cents for every math problem that he got right and fined him 5 cents for each wrong answer. After 10 problems, the mother had to pay the boy one-fifty.How many problems did the boy answer correctly?

T : Do you understand the question?

P1: Yes.

[teacher checks whether P understands the problem]

T : Do you know how to do it?

P1: 20 cents ..., okay, 20 cents times 10 is 2 dollars.

Not all the 10 problems he got right.

5 cents ...,

2 dollars minus one-fifty is 50 cents ...

That means 50 cents is his wrong answer.

So ..., how many 5 cents will there be if you want 50 cents ...,

[counting with her fingers] ...., 10

I think it's wrong already.

T : Why do you think it's wrong?

P1: Because he cannot answer all 10 problems wrongly.

T : Can you think of any other way to solve it?

P1: No, I cannot do it.

T : Have you seen such a problem before?

P1: No.

Excerpt B2:

P2: A mother promised to pay her son 20 cents for every math problem that he got right and fined him 5 cents for each wrong answer. After 10 problems, the mother had to pay the boy one-fifty. How many problems did the boy answer correctly?

T : Do you understand the question?

P2: Yes.

T : Do you know how to solve it?

P2: I'll try.

If he got right one problem, he'll get 20 cents.

so 1 problem 20 cents ..., there are 10 probl ms.

10 times 20 is 2 dollars ...,

If he get 10 problems right, he'll get 2 dollars.

If he get 10 sums wrong, 5 cents times 10 ..., 50 cents ...,

he has to pay 50 cents.

[pause]

T : What are you thinking now?

P2: I'm trying to get an equation ..., hmmm ...,

T : What sort of equation are you trying to get?

P2: Something plus something equals one-fifty.

[pause]

T : What are you thinking now?

P2: I'm confused.

T : But you understand the question?

P2: Yes, but I cannot set up an equation ..., because of the 20 cents and the 5 cents.

Let me see ..., if he get 5 sums right ..., hmm ...,

that will be one dollar.

Then the 50 cents ..., I'll add two more sums to make it one-forty [scratch her head and giggles]

I think I did it wrongly.

T : Why do you think so?

P2: I think I should get an equation ...,

[rereads the question].

5 sums ..., one dollar, two more sums one-forty ...,

but this is not algebra.

T : You'll prefer to use the algebra method?

P2: I think it's easier, but I can't set the equation.

T : Do yoũ want to proceed with this method?

P2: I don't think it's correct.

T : Do you want to try further?

P2: No. I don't think I can get the answer.

Towards a pedagogy of problem solving on algebra

This section presents some findings that contribute to a pedagogy of solving algebra word problems. Some useful questions for helping pupils monitor their problem solving are also given. Three lines of research findings are evident:

1. In order to solve an algebra word problem successfully, pupils need to be conversant with a number of exemplary problems so that problem-categorisation and problem-representation can be made easier. Otherwise, even though pupils do not have any difficulties in understanding the given problem, they cannot represent it in a way for an algebraic solution. Those who are keen may use a trial and error method or approach the problem in an arithmetical way. It should be noted that problem-categorisation and problem-representation cannot be completed if the surface features have not been sorted out by the pupils. The reason is that pupils may not recognise the underlying "deep structures" that are meant for them. These two steps of categorisation and representation are vital for solving word problems and prove most difficult even for

the high ability pupils.

2. Pupils plunge into their solving of the problems often too quickly, failing to recognise that there are important cues that have to be used. As a result, pupils often cannot represent a problem properly. When reaching an impasse, they also lack heuristic strategies to explore the problem in a systematic way. It should be noted that associated with each problem representation there is a domain-specific knowledge base. In solving algebra problems, pupils are thus keen to find equations linking the givens to the unknowns. Solving a variety of problems which have the same "deep structure" in numerous contexts may help pupils in the setting up of these equations. Preferably, the problems should be designed in such a way that pupils will immediately recognise whether their solution paths will lead to a detour or blind alley.

3. There is a hierarchy of concepts and procedural skills showing how pupils of different competencies progress towards mastery. In a parallel manner, pupils of different levels of abilities commit different types of errors. Consequently, instructions may be applied according to pupils' performance levels. The fashionable constructivist teaching approaches in the educational literature will help pupils construct their concepts. As far as procedural skills are concerned, rehearsal within a personally meaningful context may drive mastery towards automation. Engaging pupils in problem solving tasks reveal the diversity of perceptions and

approach , so much needed to be understood by classroom teachers if meaningful learning is the outcome of schooling.

As far as metacognition is concerned, the following questions may be asked covertly or overtly when pupils are engaged in problem solving. These questions aid one's awareness of the progress towards a goal state, and at the same time control and monitor the solution path in the problem-solving process. These questions are relevant to the various stages of problem solving, namely, problem-understanding, problem-representation, problem- execution, problem-control, and problem-evaluation.

- What the problem asks me to do ?
- Do I need to break a problem into subproblems ?
- What goal-directed content knowledge and heuristic strategies are accessible to me ?
- Is there a related problem or pattern seen before ?
- Do I need to look at some simple cases for some tentative explorations ?
- Do I need to make a guess and check ?
- Is the problem representation appropriate and complete ?
- When to apply particular heuristic strategies ?
- When to pursue alternative solution routes ?
- When to review progress, within or between episodes, on how things are and where they lead to ?
- What to do when an impasse has been reached ?
- Am I surprised, irritated, frustrated, anxious, and confused ?
- Can I tolerate ambiguity of results and premature closure ?

- Is there another way to look at the problem ?
- Do I need to check solution by retracing steps ?

In sum, the present chapter completes the discussion on the meaningful assessment of problem-solving tasks in the classroom. The problem-solving network, the proficiency continuum, the hierarchical structural dimensions of errors, the problem and search spaces of problem solvers, and most important of all, the conceptual framework which is based on a theory of perception logic and constructivist information processing all contribute to the proposed art and technology of meaningful asssessment of problem-solving tasks in the classroom.

AUTHORS

Dr K. C. CHEUNG joined the Educational Research Unit, Institute of Education, Singapore in 1989 after obtaining his Ph D from the Centre of Educational Studies, King's College, London. His thesis on science achievement in Hong Kong won him the Bruce H Choppin Memorial Award. He was lecturing in the University of Hong Kong before going on to London on a Commonwealth Academic Staff Scholarship for Science Education. His research interests cover constructivism in learning and teaching, science education and modelling of educational data. He is now serving at the Centre for Applied Research in Education (CARE), as well as at the Division of Psychological Studies, School of Education of the Nanyang Technological University in Singapore.

Miss MOOI Lee Choo obtained her Master's degree in Education in 1991. She is currently teaching Computing (Advanced Level) in a junior college in Singapore. Her main interests are educational computing, modelling, testing and measurement.

Madam LOK We Fong is currently writing her dissertation on "Secondary School Students' Understanding of Algebraic Expressions and Equations" for the Master's degree in Education. Her interests include mathematical problem-solving and modelling of errors.

References

Behr, M., Erlwanger, S. & Nichols, E. (1976). How children view equality sentences. PMDC Technical Report No.3, Florida State University. ERIC Doc. ED 144802.

Biggs, J. (1987). Reflective thinking and school learning - An introduction to the theory and practice of metacognition. SET Research Information for Teachers, Item 10, No. 2.

Bliss, J., Monk, M. & Ogborn, J. (1983). Qualitative Data Analysis for Educational Research. London and Canberra: Croom Helm.

Booth, L. R. (1984). Algebra: Children's Strategies and Errors. NFER-NELSON.

Bransford, J., Sherwood, R., Vye, N. & Rieser, J. (1986). Teaching thinking and problem solving. American Psychologist. 41, 1078-1089.

Cheung, K. C., Koh, W. K., Soh K. C. & Mooi, L. C. (1990). Meaningful Measurement in the Classroom Using the Rasch Model: Some Exemplars. Singapore: Institute of Education. ERIC Doc. TM015767.

Cheung, K. C. (1990). Bamboo stems and pig-tails: Some thoughts on meaningful measurement and testing in the classroom. In K. C. Cheung, et al. Meaningful Measurement in the Classroom Using the Rasch Model: Some Exemplars. Singapore: Institute of Education.

Collis, K. F. (1975). A Study of Concrete and Formal Operations in School Mathematics: A Piagetian Viewpoint. ACER.

Davis, R. B. & Silver, E. A. (1982). Mathematical behaviour of children. In Mitzel, H. E. (Ed.). Encyclopedia of Educational Research. AERA, 1157-1163.

Davis, R. B. (1985). The role of representations in problem solving: Case Studies. Journal of Mathematical Behavior, 4, 85-97.

de Bono, E. (1990). I am Right - You are Wrong. New York: Viking Penguin.

de Corte, E. & Verschaffel, L. (1985). Beginning first graders' initial representation of arithmetic word problems. Journal of Mathematical Behavior, 4, 3-21.

Dickson, L. (1989). Equations. In Hart, K. (Ed.). Children's Mathematical Framework. NFER-NELSON.

di Sessa, A. (1985). Knowledge in Pieces. Address to the 15th Annual Symposium of the J. Piaget Society. Philadelphia, June 1985.

Ellerton, N. (1985). A child's perspective of algebra. SET Item 4 (A). ACER.

Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. Psychological Review, 87(3), 215-251.

Guttman L. (1941). Quantification of a class of attributes: A theory and method for scale construction. In J. P. McIver & E. G. Carmines (Eds.). Unidimensional Scaling. London: Sage Publications, 1981.

Janvier, C. (1987). Representation and understanding: The notion of function as an example. In C. Janvier (Ed.). Problems of Representation in the Teaching and Learning of Mathematics. Hillsdale, NJ: Erlbaum.

Kelly, G. A. (1955). The Psychology of Personal Constructs (Volume 1 & 2). New York: W. W. Norton.

Kieran, C. (1981). Concepts associated with the equality symbol. Educational Studies in Mathematics , 12, 317-326.

Kilpatrick, J. (1967). "Analysing the Solution of Word Problems in Mathematics - An Exploratory Study". Unpublished doctoral dissertation, Stanford University.

Koh, W. K. & Cheung, K. C. (1990). On meaningful measurement: Primary school pupils' understanding of part-whole concept of fractions. In K. C. Cheung et al. Meaningful Measurement in the Classroom Using the Rasch Model: Some Exemplars. Singapore: Institute of Education.

Kouba, V. L. (1989). Children's solution strategies for equivalent set multiplication and division word problems. Journal for Research in Mathematics Education, 20(2), 147-158.

Kuchemann, D. E. (1980). Algebra. In Hart,K. (Ed.) Children's Understanding of Mathematics . London: Murray.

Lesh, R., Landau, M. & Hamilton, E.(1983). Conceptual models and applied mathematical problem-solving. In R. Lesh and M. Landau (Eds.). Acquisition of Mathematics Concepts and Processes. New York: Academic Press.

Masters, G.N. (1982). A rasch model for partial credit scoring. Psychometrika, 47(2), 149-174.

Matz, M. (1982). Towards a process model for high school algebra errors. In D. Sleeman and J. S. Brown (Eds.). Intelligent Tutoring System. New York: Academic Press.

Ministry of Education (1981). Elementary Mathematics Syllabus For The New Education System. Singapore: Ministry of Education.

Newell, A. & Simon, H.A. (1972). Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.

Osborne, R. & Wittrock, M. (1985). The generative learning model and its implications for science education. Studies in Science Education, 12, 59-87.

Radatz, H. (1979). Error analysis in mathematics education. Journal For Research in Mathematics Education, 10, 163-172.

Resnick, L. B. & Ford, W. W. (1981). The Psychology Of Mathematics For Instruction. NJ: Lawrence Erlbaum Associates.

Resnick, L. B., Cauzinille-Marmeche, E. & Mathieu, J. (1987). Understanding algebra. In J. A. Sloboda and D. Rogers (Eds.). Cognitive Processes in Mathematics. New York: Oxford University Press.

Riley, M. S., Greeno, J. G. & Heller, J. I. (1983). Development of word problem solving ability. In. H. P. Ginsburg (Ed.). Development of Mathematical Thinking. New York: Academic Press.

Schoen, H. L. & Oehmke, T. (1980). A new approach to the measurement of problem-solving skills. In S. Krulik and R. E. Reys (Eds.). Problem Solving in School Mathematics. NCTM 1980 Yearbook.

Schoenfeld, A. H. (1983). Episodes and executive decisions in mathematics problem solving. In R. Lesh and M. Landau (Eds.). Acquisition of Mathematical Concepts and Processes. New York: Academic Press.

Schoenfeld, A. H. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.). Cognitive Science and Mathematics Education. Hillsdale, NJ: Erlbaum.

Schroeder T. L. & Lester, F. K. (1989). Developing understanding in mathematics via problem solving. In New Directions For Elementary School Mathematics. NCTM.

Silver, E. A. (1987). Foundations of cognitive theory and research for problem-solving instruction. In A. H. Schoenfeld (Ed.). Cognitive Science and Mathematics Education. Hillsdale, NJ: Erlbaum.

Sleeman, D. (1982). Assessing aspects of competence in basic algebra. In D. Sleeman and J. S. Brown (Eds.). Intelligent Tutoring Systems. New York: Academic Press.

Thwaites, G. N. (1982). Why do children find algebra difficult? Mathematics in School, 11(4), 16-17.

Van Lehn, K. (1983). On the representation of procedures in repair theory. In H. P. Ginsburg (Ed.). The Development of Mathematical Thinking. New York: Academic Press.

Venezky, R. L. & Bregar, W. S. (1988). Different levels of ability in solving mathematical word problems. Journal of Mathematical Behavior, 7, 111-134.

Appendix 1

ALGEBRA TEST Time:60 minutes

1. Simplify 2a+5b+a.

2. Simplify (8y+6z) - (4y+3z).

3. What is the value of 3a-b+5c if a=2, b=3 and c=5?

4. Given that 4y-8 = 2y-4, find the value of y.

5. Solve the equation 4(x+2) + 2x = 4.

6. One stick is twice as long as the other. If the length of the shorter one is x cm, find the length of the longer one.

7. The price of a desk and a chair is $x. What is the price of the desk if the price of the chair is: i) $20, ii) $y?

8. Mangoes cost m cents each and papayas cost p cents each. If I buy 5 mangoes and 2 papayas, what does 5m + 2p stand for?

9. Mary's basic wage is $500 per month. She is also paid another $2 for each hour of overtime that she works. If h stands for the number of hours of overtime that she works and if w stands for her total wage in $,
   i) write down an equation connecting w and h,
   ii) what would Mary's total wage be if she worked 10 hours of overtime?

10 A mother promised to pay her son 20 cents for every Maths problem that he got right and fined him 5 cents for each wrong answer. After 10 problems, the mother had to pay the boy $1.50. How many problems did the boy answer correctly?

11 A woman has altogether 50 cows and chickens. These animals have a total of 172 legs. How many cows are there?