A taxonomy of mathematical problem solving behaviours

Foong Pui Yee

Paper presented at the 13th Biennel Conference on "Mathematical Turning Points: Strategies for the 1990s", jointly organised by the Australian Association of Mathematics Teachers and the Mathematics Education Research Groups of Australasia at the University of Tasmania, Australia, from 5-10 July 1990
A TAXONOMY OF MATHEMATICAL PROBLEM SOLVING BEHAVIOURS
Foong Pui Yee
Institute of Education, Singapore
Monash University, Australia

The taxonomy described in this paper was developed from a research that had its major objective the investigation of the process of mathematical problem solving in terms of definable behaviours. The taxonomy was developed for use as an instrument to classify and encode behaviours in their sequence of observed occurrence during the process of mathematical problem solving. It formed the basis of a behavioural analysis framework formulated in the research to examine the problem solving protocols of individuals for comprehensive information about the problem solving process itself, the individual differences in the behaviours of subjects and the strategies applied by each in dealing with the different tasks used in the study.

Introduction

Much of the recent research on mathematical problem solving attempts to describe and characterise the problem solving process, derived from verbal reports. It has become increasingly important for research to develop valid and reliable instruments which reflect actual behaviours in the problem solving process. Kilpatrick (1967), Lucas et al. (1980), Rowe (1980), Schoenfeld (1983), and many others have devised protocol analysis frameworks to record and analyse behaviours in a sequence of observed occurrence.

This paper describes a framework for analysing the "thinking aloud" data from problem solving performance on certain non-routine mathematical problems. The methods of data collection and protocol analysis were derived from research within the framework of information processing influenced by Newell and Simon's (1972) theory of problem solving. The present study was able to develop a comprehensive taxonomy comprising cognitive, metacognitive and affective behaviours manifested in individuals during problem solving activities.

The taxonomy identified 28 behaviour categories which could be coded reliably. These behaviours could be considered as low inference measures that were objective, observable and most importantly could be defined in operational terms. Inter-coder agreement was used to establish reliability and to refine concepts which had to be reconciled with the actual data under the conditions set up by the methodology in the present study.
**Design**

The Sample

The nine volunteer subjects in this study were adults enrolled as pre-service trainee teachers at the Singapore Institute of Education. A range of mathematical abilities for this adult level was chosen so that all possible behaviours could be expected to occur. The nine subjects were from three levels of mathematical background. Two had General Certificate of Education (GCE) 'O' (Ordinary) level secondary school mathematics with low credits, two GCE 'A' (Advanced) level post-secondary mathematics with average credits and five had university degrees in mathematics with two having upper class honours. They were recommended by their respective mathematics education course lecturers who considered that the subjects' mathematical backgrounds were suited to the investigator's purpose. These volunteers were briefed by the investigator on the aim of the project and given training in "thinking-aloud".

All the participants were proficient in the use of the English Language. This proficiency was both a prerequisite for their selection into the teacher training programme and had been the language of instruction throughout their mathematical education. This was an important consideration as this investigation utilised verbal reports as data and the subjects, who were not native speakers of the English Language, needed to be confident in the use of the language in order to participate in the "thinking-aloud" procedure.

The "Thinking Aloud" Technique

As the processes during cognitive performance tend not to be directly observable, the first task was to find a methodology which would make it possible to identify and monitor valid elements, in terms of operationally defined behaviours. A promising method was the "thinking aloud" (TA) technique which generates data from the concurrent verbal reports of subjects while engaged in problem solving. The main concern was that the TA instructions should be given to the subjects so that they could vocalise all their problem solving activities with minimal interruption while performing the tasks.

Previous research, using a similar methodology, commonly gave training to subjects before the study began so that during problem
solving they could "think aloud" in an ongoing manner and without time delay. It was found that "thinking aloud" was relatively easy to learn given sufficient practice and as long as the subjects were motivated to co-operate. A standard procedure was essential for all subjects and tasks. The experimenter should not say more than absolutely necessary; neither direct the subject's approach, ask leading questions or reinforce while the subjects were "thinking aloud".

The TA instructions were adapted from previous studies such as Duncker (1926), reported in Ericsson and Simon (1984). The subjects were asked to vocalise their thoughts as though they were talking to themselves. They were told to talk out loud constantly whatever came to mind and that the experimenter was not primarily only interested in their final solution but also in their thinking processes. To prevent subjects from explaining their solutions aloud, the experimenter followed Krutetskii's (1976) warning to subjects at the beginning against confusing the instruction to think aloud with that of explaining the solution:

Do not try to explain anything to anyone else. Pretend there is no one here but yourself. Do not tell about the solution but solve it.

(Krutetskii, 1976, p.93)

Procedures for data collection

The subjects were arranged in individual recording booths in a language laboratory to take part in the experiment. The experimenter started each subject off with a problem presented on a typed sheet of paper. The subject was asked to begin work on each task by reading the problem statement aloud. Once this had begun, the experimenter moved on to the next subject to start on the same problem. The subjects were told to put up their hands when they had finished with a problem and the experimenter would come to present them with the next problem. Each problem was presented on a separate sheet of paper and subjects were also provided with pencil and paper for their scratch work. Five problems (Appendix A) were presented, one at a time in the same order, for every subject. No time limit was imposed and the subjects were reminded at the beginning of the session of the TA instructions.
All the subjects attempted to solve the same five problems. Each problem solving session was recorded in its entirety and the audiotape protocols were transcribed verbatim. Forty-five protocols were collected, each protocol consisted of one complete transcription of a subject's audiotaped recording of his or her solution process to a problem. Each transcribed protocol was divided into segments of behaviours. Each segment was then classified and encoded according to a predetermined taxonomy of behaviours and a process-coding scheme. For each subject and task, the data consisted of the types, the frequencies and the sequence of coded behaviours used during the problem solving process.

**Protocol Analysis**

Protocol analysis refers to the use of a systematic event-by-event record of an individual's behaviours while engaged on a cognitive task, as a source of data. The analysis of the "think aloud" protocols in this study were based on a framework that consisted of two components:

a. a taxonomy with a process-coding scheme to code all observable problem solving behaviours in actual sequence of their occurrence.

b. a working model for episode-analysis of the protocols to identify global patterns of metacognitive processes.

**The Taxonomy**

The term taxonomy used in this research refers to a predetermined set of problem solving behavioural variables that are classifiable into five distinct categories. It serves as a dictionary for characterising the mathematical problem-solving behaviours in the subjects' protocols. Encoding decisions made by a coder were based on this dictionary of coding categories. A review of the problem solving literature provided the investigator with a list of frequently recurring behaviours that could be classified into five major categories:
1. **Problem-Orientation Heuristics**: strategies through which a problem-solver attempts to analyse and understand the problem situation.

2. **Problem-Solution Heuristics**: rule-of-thumb strategies through which a problem-solver moves towards a solution.

3. **Domain-Specific Knowledge**: the inventory of mathematical facts, procedures and skills that a problem solver is able to use in the solution process.

4. **Metacognition**: a problem-solver’s awareness and monitoring progress of his or her own thinking during the task.

5. **Affective Behaviours**: self-related expressions and emotional responses that are aroused in the problem solver.

A preliminary list consisting of all the possible behaviours in mathematical problem solving was prepared from the study of the literature. Altogether 40 behaviours were identified as relevant to the tasks used for this research. They formed the initial taxonomy which was used as a preliminary guide to identify observable behaviours in the protocols of the subjects in this study.

After a series of modifications through tests of reliability a final taxonomy (Appendix B), with a coding system was derived. Although it was initially based on prior assumptions about problem-solving processes from previous research, the taxonomy was modified empirically through the consensus of different coders after they had applied it to actual problem-solving protocols of different subjects across different problems. In order for the coding judgments of any coder to be as objective as possible, the definitions of each specific behaviour in the taxonomy were stated in operational terms and were supported by examples from the protocols (see Appendix C for some definitions and examples).
Process-Segmentation Encoding Procedures

Encoding was carried out on the transcripts of the subjects' audiotaped protocols, together with their written work, to provide a more comprehensive analysis of the solution path. The analysis of the protocols was based on the taxonomy. The overall strategy was to divide a protocol into segments of behaviour. Each segment was matched against the behaviours in the taxonomy and assigned a code. A whole protocol was eventually recorded as a sequence of encodings in a horizontal string corresponding to the order of the actual problem solution of an individual. Conversely, a string of encodings could provide a relatively clear description of what had actually happened during the problem-solving process. (A sample of the protocol analysis is shown in Appendix D)

The coding of individual segments of the protocols was intended to facilitate later comparison of encodings by different coders to establish the reliability of the taxonomy. Each segment of a protocol corresponded to a statement. According to Ericsson and Simon (1984), if the verbalisations were completely grammatical, a statement would essentially be a clause or a sentence, but in the thinking aloud protocols which were close to normal speech, statements were often abbreviated to phrases or even single words. It was also found in the protocols that some subjects often verbalised a single behaviour or operation ungrammatically as an aggregate of phrases. Each segment or an aggregated segment of behaviour was then encoded from the information contained in it.

A Working Model for the Episode-Analysis of the Protocols

The flow chart in Figure I represents a working model of a global problem solving process where "chunks" of consistent behaviours identified in the taxonomy could be parsed into six types of episodes which were modified from Schoenfeld (1983) for the purpose of this study as:

1. SCANNING  2. ANALYSIS  3. EXPLORATION
4. IMPLEMENTATION  5. REVIEW  6. VERIFICATION
FIGURE 1
A WORKING MODEL FOR EPISODE-ANALYSIS
OF THE MATHEMATICAL PROBLEM SOLVING PROCESS

PROBLEM POSED

SCANNING
the information

Re-scan

Give-up

ANALYSIS
of the problem

Re-analyse

Give-up

able to?

REVIEW

Yes

No

RE-explore

Give-up

EXPLORATION

stuck?

REVIEW

Yes

No

IMPLEMENTATION

Re-do

Done

works?

REVIEW

Yes

No

DONE

VERIFICATION
An episode defined for this research is a period of activity during which the problem solver is engaged in a set of consistent behaviours that can be described as a certain event, such as analysis of the problem or exploration of strategies. The episodes can occur in any sequence or they can be recurrent. For example, some problem solvers may, after scanning the problem, go straight to implementation without analysis or exploration and in review became "stuck" or "does not work" might go back to scanning or analysis and work down again, or they may "give-up", as illustrated by the variety of possible paths in the flow chart. The review episodes are junctures where evidence of metacognitive activities are most apparent.

Each episode consists of specific behaviours that are based on the Taxonomy. Table 1 shows the list of predominant behaviours in the six episodes which can be classified into three main components of the mathematical process: "orientation to the problem", "execute the solution" and "evaluation". Each protocol sequence was parsed by examining the behaviours and dividing them into episodes that were enclosed and labelled in boxes. An example is shown in Figure 2. The episodes were parsed in sequential order of occurrence and were presented in the form of a flow-chart. The behaviours within each episode were presented in code symbols and the interpretations of these codes are shown in the key. The coding procedures used for the behaviours in each problem solving protocol were those established in the taxonomy.
# TABLE 1

**PREDOMINANT BEHAVIOURS IN THE SIX EPISODES OF THE THREE MAIN PROBLEM SOLVING COMPONENTS**

<table>
<thead>
<tr>
<th>I. ORIENTATION TO THE PROBLEM</th>
<th>II. EXECUTE THE SOLUTION</th>
<th>III. EVALUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. SCANNING</strong></td>
<td><strong>2. ANALYSIS</strong></td>
<td><strong>5. REVIEW</strong></td>
</tr>
<tr>
<td>P2 - rereading of problem</td>
<td>P4 - examine dimensions</td>
<td>M1 - suggest a plan</td>
</tr>
<tr>
<td>P3 - paraphrase statement</td>
<td>P5 - construct representation</td>
<td>M2 - assess difficulty</td>
</tr>
<tr>
<td></td>
<td>H2 - draw diagrams</td>
<td>M3 - review progress</td>
</tr>
<tr>
<td><strong>3. EXPLORATION</strong></td>
<td><strong>4. IMPLEMENTATION</strong></td>
<td><strong>6. VERIFICATION</strong></td>
</tr>
<tr>
<td>H4 - look at cases</td>
<td>H7 - generalise</td>
<td>H9 - check computation and final answer; some metacognition and P, H and K behaviours</td>
</tr>
<tr>
<td>H5 - guess and check</td>
<td>H8 - make a deduction</td>
<td></td>
</tr>
<tr>
<td>H6 - search for patterns</td>
<td>K3 - use relevant</td>
<td></td>
</tr>
<tr>
<td>K1 - computation</td>
<td>mathematical procedures</td>
<td></td>
</tr>
<tr>
<td>K2 - state a fact or rule</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5. REVIEW</strong></td>
<td><strong>6. VERIFICATION</strong></td>
<td></td>
</tr>
<tr>
<td>M1 - suggest a plan</td>
<td>H9 - check computation</td>
<td></td>
</tr>
<tr>
<td>M2 - assess difficulty</td>
<td>and final answer; some metacognition</td>
<td></td>
</tr>
<tr>
<td>M3 - review progress</td>
<td>and P, H and K behaviours</td>
<td></td>
</tr>
<tr>
<td>M4 - recognise error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M5 - new development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1 - self-question</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 2
AN EXAMPLE ON THE EPISODE-ANALYSIS OF A PROCESS-SEQUENCE
A Protocol for the PALINDROMES-problem  (Time Taken: 7 min.)

EPISODES

Analyze
P1, P3, P5
recognise general term

Explore 1
H4, K1, H6

Review
M4, M1

Explore 2
H4, K1, H4, K1

Review
M4, M3, M3

Explore 3
P4, H4, K2, K1
search for 4 digits

Review
M4, M3, M5

Implement
H7, K3, Q1, K3
Ans.
state answer

EXCERPTS OF THE VERBAL PROTOCOL

"...that means in general a 4-digit palindrome can be written as ABBA..."

"...2 digit number, xy multiply by 11, something xy plus something xy..."

"...no it cannot be... let's find counter examples of palindromes."

"How about 3223... divide by 11... uhhmm no... 3443 by 11... 313... uhhmm"

"...wait... there must be some properties of 11 that I can make use of...
... let's go back to this 4... ABBA "

"...so we need a 2 digit number greater or equal to 91... so when multiply by 11 you get a 4 digit number..."

"...what I'm doing... no... not necessarily..."
"let's try again... I must start with 1001..."
"Oh! I see ABBA, I can do it this way..."

"ABBA is equal to 1000A + 100B + 10B + 1A, alright... then I group them together..."
"1000A + 1A + 100B + 10B = A(100) + 10B(10 + 1)"
"... I can take out a common factor... 11"

Key:
P1- first reading of problem
P3- rephrase or give meaning
P4- analyse given
P5- correct representation
K1- computation (division)
K3- use math procedures
M1- suggest a plan
M5- recognise new development

H6- search for pattern
Q1- self-question
H4- look at examples
H7- generalise
K2- state a fact
M3- review progress
M4- recognise error
Establishing the Reliability of the Taxonomy

Issues of reliability and replicability in the coding scheme based on the taxonomy are concerned with the questions: "Would others see the same behaviours in the protocol?" and "Would they ascribe to them the same coding categories?". To address these issues tests of reliability based on agreement between different persons encoding independently of each other were used. The correlation between the categorisations made by different coders in relation to the same task provides an index of the objectivity of the taxonomy for categorization of problem solving behaviours.

The protocols collected in this study were divided into individual segments in order that direct comparisons between different encodings of the same segments by different coders were made relatively easy. A percentage of agreement between coders was calculated to give an aggregate measure of reliability. However, this reliability mainly reflects the very frequent coding categories, while the encoding of infrequent categories may be more or less reliable than the average.

To further test the replicability of each of the individual behavioural variables in the taxonomy by different coders, a coefficient of agreement calculation modified by Putt and Pountney (1989) from Lucas et al. (1980) was used. Inter-coder agreement was assessed systematically in two stages. In the first stage, tests of reliability were performed for the five categorical constructs of the taxonomy. This led to a refinement of the preliminary taxonomy by either eliminating, combining, or redefining behaviours which were redundant or vague. The second stage involved tests of reliability for every of the 28 specific behavioural construct defined in the final taxonomy.

Inter-Coder Agreement: First Stage

Procedures. The first inter-coder agreement procedure was used to test the reliability of the five categorical constructs:

- Problem-Oriented Heuristics
- Problem-Solution Heuristics
- Domain-Specific Knowledge
- Metacognition
- Affective Behaviours
A random selection of the protocols of five subjects from the pilot sample were given to five other coders, each coding on one subject's transcript for five problems. Their encodings were then compared against the investigator's. They were instructed to code only the category to which they thought a segment of behaviour might belong. They did not need to code the specific behaviour. For example, if they thought a specific segment of behaviour was a strategy under "Problem-Solution Heuristics", they needed only to use the code "H" for that segment or if they thought a behaviour was an emotional response, then they used the code "A" for the "Affective Behaviours" category. In essence, the coders needed only to decide under which of the categories "P", "H", "K", "M", or "A" a certain behaviour should be classified.

Results and Discussion. Table 2 shows the aggregate agreement in proportion as well as in percentage for each coder against Coder F, who was the investigator. There were five different sets of protocols containing varying numbers of segments for the five pairs of coders and each pair encoded the same segments. There was a reasonably consistent agreement between each of the five other coders and the investigator.

<table>
<thead>
<tr>
<th>Coder Pair</th>
<th>Protocols</th>
<th>Proportion Agreement</th>
<th>Percentage Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>F - A</td>
<td>Subject 1</td>
<td>92/138</td>
<td>66.7%</td>
</tr>
<tr>
<td>F - B</td>
<td>Subject 2</td>
<td>56/78</td>
<td>71.8%</td>
</tr>
<tr>
<td>F - C</td>
<td>Subject 3</td>
<td>77/111</td>
<td>69.4%</td>
</tr>
<tr>
<td>F - D</td>
<td>Subject 4</td>
<td>63/98</td>
<td>64.3%</td>
</tr>
<tr>
<td>F - E</td>
<td>Subject 5</td>
<td>70/110</td>
<td>63.6%</td>
</tr>
</tbody>
</table>

Mean 67.2%

The aggregate percentage agreement for all the encodings between each of the other five coders and the investigator for each of the five protocols ranged from 63.6% to 71.8% with an overall mean agreement of 67.2%. In a similar study by Lucas et al. (1980), an overall agreement of 78% (agreement index ranged from 0 to 1.00) was obtained between two coders across fourteen different segments. Each of the segments might be an episode of behaviours that had more than one coding, depending on the interpretation of the coder.
The aggregate agreement percentages shown in Table 2, were not satisfactory at this early stage of the development of the taxonomy and it suggested a need for further improvement. The five coders had expressed difficulties in making judgement for some of the categories and had wished that they had been given more time to practice on some protocols before they proceeded with the assignments, which were to be completed within the three hour session. However, this had been intended, so that a general perception and interpretation from others for the items in the preliminary taxonomy could provide information and feedback to the researcher in refining the system.

Table 3 shows an analysis of the breakdown of the inter-coder agreements and discrepancies for each category of the behaviours across all the protocols. The columns of the table represent the encodings made by the investigator, Coder F, and the rows represent those made for the same protocols by each of the other five coders. The diagonal cells show the frequencies of agreement between Coder F and the other five for each category of behaviours. The last row X represents the frequencies of behaviours classified by Coder F for each of the five categories "P", "K", "H", "M", and "A". The last column, Y, represents the frequencies of corresponding categories by the other coders.

Large discrepancies were found in the "K", "M" and "A" categories. In the "A" category, Coder F identified a total of 30 behaviours as affective behaviours while the others coded 59 such behaviours. Of these the coders showed agreement on 26. Of the 59 affective behaviours identified by the five coders, 29 were perceived by the investigator as metacognitive behaviours. In the other categories, Coder F and the other coders identified almost the same number of "K" behaviours, 42 and 41 respectively, but they agreed on on 21 instances. Thirteen of the 42 behaviours identified by Coder F as belonged to the "K" category were perceived otherwise as "H" heuristic behaviours by the others. Similarly, 14 of the 41 "K" behaviours identified by the other coders were coded as "H" behaviours by Coder F. All these discrepancies and mismatches suggested that there were obvious overlapping in the interpretation of the behaviours in the various categories, possibly attributable to the vagueness in definition or misplacement of behaviours in each category.
Further tests of reliability of the coding were carried out for each of the five categories using a formula modified by Putt and Pountney (1989) from Lucas et. al (1980) to calculate a coefficient of agreement. The Lucas formula is:

\[ C = \frac{2A}{X + Y} \]

where 
- \( C \) - coefficient of agreement
- \( A \) - number of behaviours in a category with full agreement between Coder F and others
- \( X \) - number of behaviours in a category identified by Coder F.
- \( Y \) - number of behaviours in a category identified by other coders.

Table 4 shows the range of coefficients of agreement for the five categorical constructs. The moderate coefficient of agreement figures of 0.51, 0.57 and 0.58 for the "K-Knowledge", "M-Metacognition", and "A-Affective" categories respectively, led the investigator to review the constructs of all the behaviours under these classifications.
TABLE 4
COEFFICIENT OF AGREEMENT FOR EACH CATEGORY

<table>
<thead>
<tr>
<th>Category</th>
<th>Number Both Agreed</th>
<th>Number by Coder F</th>
<th>Number by 5 Others</th>
<th>Coefficient Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>X</td>
<td>Y</td>
<td>C=2A/X+Y</td>
</tr>
<tr>
<td>P-Orientation</td>
<td>90</td>
<td>118</td>
<td>114</td>
<td>0.78</td>
</tr>
<tr>
<td>K-Knowledge</td>
<td>21</td>
<td>42</td>
<td>41</td>
<td>0.51</td>
</tr>
<tr>
<td>H-Heuristic</td>
<td>157</td>
<td>201</td>
<td>218</td>
<td>0.75</td>
</tr>
<tr>
<td>M-metacognition</td>
<td>70</td>
<td>144</td>
<td>103</td>
<td>0.57</td>
</tr>
<tr>
<td>A-Affective</td>
<td>26</td>
<td>30</td>
<td>59</td>
<td>0.58</td>
</tr>
</tbody>
</table>

A further discussion was held with the five coders to clarify their perception on all the behaviours in the preliminary taxonomy that was used by them in the encoding process. It was found that the other coders felt that the three specific behaviours "Say don’t understand or don’t know how to do", "Express doubt about own knowledge" and some of the self-questioning responses in the "M" (Metacognition) category overlapped with the behaviour "Self-assessment" in the "A" (Affective) category. These were mainly comments about "self" by the problem-solver. Also, certain behaviours defined in the taxonomy as under the "K" Domain-Specific Knowledge category, were perceived under the "H" Problem-Solution Heuristics category. Some of the heuristic behaviours such as "Check a step" or "Review or summarise" in the taxonomy were perceived as metacognitive behaviours by the other coders. All these mismatches of judgment, owing to vague definitions and subtle interpretations by various coders, explained the low coefficient agreement figures of the categories already mentioned. Coefficient of agreement figures for the other two categories: "P-Problem-Orientation Heuristics", 0.78 and "H-Problem-Solution Heuristics", 0.75 were encouraging.

A rethinking of definitions and constructs was necessary to eliminate redundancies or to combine overlapping behavioural variables. Variables in the preliminary taxonomy were discarded if they had not been used, some were combined and recategorised for logical reasons and some redefined if there was inconsistency in judgement between coders and across coders.
Inter-coder Agreement: Second Stage

Procedures. After modification of the taxonomy in the first stage, tests of reliability were conducted in this second stage for all the behaviours finalised in the taxonomy. On this occasion two coders were used for the 45 protocols which yielded a total of 943 segments of behaviours across nine subjects on five different problems. One coder, the investigator who was by then experienced, applied the revised taxonomy and its coding system to the protocols. Another coder, who was an experienced secondary mathematics teacher and who had recently attended an in-service course on mathematical problem solving, coded the same 45 problem analysing the transcripts independently. Her encodings were then compared with those of the investigator. Aggregate measures of agreement between the two coders on the codings of behaviours were assessed separately for each subject across the five problems and for each task across all the nine subjects. Tests of reliability using the coefficient of agreement formula were also established for the codings of all behaviours that were identified between the two coders across all the subjects and tasks.

Results and Discussion

Table 5 shows the proportions and percentages of agreement between the coders. Of the total of 943 encodings across all problems and subjects both coders agreed on 786 instances which gave a relatively satisfactory aggregate agreement of 83.4%. The average percentages of agreement for tasks across subjects and for each subject across tasks were 84.18 (SD = 5.0) and 84.17 (SD = 4.3) respectively.

An inadequacy was indicated in the present study. Where there was a total of 28 different behavioural variables for coding decisions, a single aggregate measure of inter-coder agreement might not reflect coding reliability of those behaviours which occurred infrequently as this figure was only an average indication of inter-coder agreement for behaviours which occurred most often. Using the number of agreement and discrepancies of judgement between the coders for each behavioural variable summed across all the subjects and problems, the coefficient of agreement was calculated for each of the 29 behaviours to obtain a more objective inter-coder reliability index. Table 6 shows the coefficient of agreement calculated for each behavioural variable.
<table>
<thead>
<tr>
<th>Problem</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>8/8</td>
<td>13/15</td>
<td>23/25</td>
<td>12/13</td>
<td>68/90</td>
<td>124/151</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>92.0%</td>
<td>92.3%</td>
<td>75.6%</td>
<td>82.1%</td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>4/6</td>
<td>8/9</td>
<td>13/15</td>
<td>12/14</td>
<td>24/29</td>
<td>61/73</td>
</tr>
<tr>
<td></td>
<td>66.7%</td>
<td>88.9%</td>
<td>86.7%</td>
<td>85.7%</td>
<td>82.8%</td>
<td>83.6%</td>
</tr>
<tr>
<td>D3</td>
<td>6/6</td>
<td>14/15</td>
<td>14/16</td>
<td>21/25</td>
<td>41/51</td>
<td>96/113</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>93.3%</td>
<td>87.5%</td>
<td>84.0%</td>
<td>80.4%</td>
<td>85.0%</td>
</tr>
<tr>
<td>D4</td>
<td>5/5</td>
<td>9/9</td>
<td>3/3</td>
<td>5/6</td>
<td>21/24</td>
<td>43/47</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>83.3%</td>
<td>87.5%</td>
<td>91.5%</td>
</tr>
<tr>
<td>D5</td>
<td>6/8</td>
<td>8/9</td>
<td>8/10</td>
<td>11/14</td>
<td>67/85</td>
<td>100/126</td>
</tr>
<tr>
<td></td>
<td>75.0%</td>
<td>88.9%</td>
<td>80.0%</td>
<td>78.6%</td>
<td>78.8%</td>
<td>79.4%</td>
</tr>
<tr>
<td>C1</td>
<td>7/10</td>
<td>21/27</td>
<td>20/22</td>
<td>22/26</td>
<td>31/39</td>
<td>101/125</td>
</tr>
<tr>
<td></td>
<td>70.0%</td>
<td>77.8%</td>
<td>90.9%</td>
<td>84.6%</td>
<td>79.5%</td>
<td>81.5%</td>
</tr>
<tr>
<td>C2</td>
<td>4/6</td>
<td>14/16</td>
<td>13/13</td>
<td>21/27</td>
<td>38/45</td>
<td>90/107</td>
</tr>
<tr>
<td></td>
<td>66.7%</td>
<td>87.5%</td>
<td>100%</td>
<td>77.8%</td>
<td>84.4%</td>
<td>84.1%</td>
</tr>
<tr>
<td>C3</td>
<td>11/14</td>
<td>17/19</td>
<td>22/26</td>
<td>21/34</td>
<td>15/16</td>
<td>86/109</td>
</tr>
<tr>
<td></td>
<td>78.6%</td>
<td>89.5%</td>
<td>84.6%</td>
<td>61.8%</td>
<td>93.8%</td>
<td>78.9%</td>
</tr>
<tr>
<td>C4</td>
<td>8/10</td>
<td>25/25</td>
<td>22/22</td>
<td>18/22</td>
<td>12/14</td>
<td>85/93</td>
</tr>
<tr>
<td></td>
<td>80.0%</td>
<td>100%</td>
<td>100%</td>
<td>81.8%</td>
<td>85.7%</td>
<td>91.4%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>59/73</td>
<td>129/144</td>
<td>138/152</td>
<td>143/181</td>
<td>317/393</td>
<td>786/943</td>
</tr>
<tr>
<td></td>
<td>80.8%</td>
<td>89.6%</td>
<td>90.8%</td>
<td>79.0%</td>
<td>80.7%</td>
<td>83.4%</td>
</tr>
</tbody>
</table>
### TABLE 6
COEFFICIENTS OF AGREEMENT OF ALL BEHAVIOURAL VARIABLES

<table>
<thead>
<tr>
<th>Variable Behaviour</th>
<th>Frequency Both Agreed A</th>
<th>Frequency by Coder A X</th>
<th>Frequency by Coder B Y</th>
<th>Coefficient of Agreement C=2A/X+Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>1.00</td>
</tr>
<tr>
<td>P2</td>
<td>79</td>
<td>79</td>
<td>80</td>
<td>0.99</td>
</tr>
<tr>
<td>P3</td>
<td>45</td>
<td>45</td>
<td>60</td>
<td>0.86</td>
</tr>
<tr>
<td>P4</td>
<td>40</td>
<td>49</td>
<td>47</td>
<td>0.83</td>
</tr>
<tr>
<td>P5</td>
<td>25</td>
<td>39</td>
<td>27</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean = 0.89</td>
</tr>
<tr>
<td>K1</td>
<td>46</td>
<td>53</td>
<td>53</td>
<td>0.87</td>
</tr>
<tr>
<td>K2</td>
<td>17</td>
<td>19</td>
<td>24</td>
<td>0.79</td>
</tr>
<tr>
<td>K3</td>
<td>20</td>
<td>33</td>
<td>21</td>
<td>0.74</td>
</tr>
<tr>
<td>K4</td>
<td>11</td>
<td>13</td>
<td>21</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean = 0.77</td>
</tr>
<tr>
<td>Q1</td>
<td>68</td>
<td>69</td>
<td>69</td>
<td>0.99</td>
</tr>
<tr>
<td>M1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0.40</td>
</tr>
<tr>
<td>M2</td>
<td>13</td>
<td>15</td>
<td>13</td>
<td>0.93</td>
</tr>
<tr>
<td>M3</td>
<td>29</td>
<td>50</td>
<td>31</td>
<td>0.72</td>
</tr>
<tr>
<td>M4</td>
<td>41</td>
<td>49</td>
<td>47</td>
<td>0.85</td>
</tr>
<tr>
<td>M5</td>
<td>4</td>
<td>10</td>
<td>5</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean = 0.74</td>
</tr>
<tr>
<td>H1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0.50</td>
</tr>
<tr>
<td>H2</td>
<td>24</td>
<td>26</td>
<td>25</td>
<td>0.94</td>
</tr>
<tr>
<td>H3</td>
<td>42</td>
<td>45</td>
<td>47</td>
<td>0.91</td>
</tr>
<tr>
<td>H4</td>
<td>107</td>
<td>119</td>
<td>129</td>
<td>0.86</td>
</tr>
<tr>
<td>H5</td>
<td>15</td>
<td>26</td>
<td>21</td>
<td>0.63</td>
</tr>
<tr>
<td>H6</td>
<td>26</td>
<td>30</td>
<td>41</td>
<td>0.73</td>
</tr>
<tr>
<td>H7</td>
<td>18</td>
<td>20</td>
<td>27</td>
<td>0.77</td>
</tr>
<tr>
<td>H8</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>0.71</td>
</tr>
<tr>
<td>H9</td>
<td>17</td>
<td>22</td>
<td>21</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean = 0.76</td>
</tr>
<tr>
<td>A1</td>
<td>26</td>
<td>26</td>
<td>27</td>
<td>0.98</td>
</tr>
<tr>
<td>A2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1.00</td>
</tr>
<tr>
<td>A3</td>
<td>37</td>
<td>39</td>
<td>44</td>
<td>0.89</td>
</tr>
<tr>
<td>A4</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>0.90</td>
</tr>
<tr>
<td>A5</td>
<td>12</td>
<td>12</td>
<td>15</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean = 0.93</td>
</tr>
</tbody>
</table>
In Table 6, Coder B coded five instances of the behaviour "MS-recognise new development" as compared to 10 instances by Coder A and they agreed on only four to give a low coefficient value of 0.53. This behaviour was found to be difficult to code as it was not often explicitly vocalised when it did occur. It required a fine distinction during the protocol analysis and in many instances the coder had to make an inference based on the context of the entire solution path of the subject. Behaviour "HI-Recall similar problem" and "M1-Suggest a plan" had very low frequencies, a maximum of three and four respectively identified by Coder A. As a result it was meaningless to interpret anything from their coefficient values calculated as 0.5 and 0.4 respectively. In spite of this, the two behaviours were still retained because of their relevance to the study.

On the whole, the majority of the 29 behaviours had relatively high coefficients of agreement. The most frequent behaviour "H4-Looking at particular case(s)" had a coefficient of 0.86. The mean coefficient values of all the behaviours in each of the five categories were: P-category 0.89, K-category 0.77, H-category 0.76, M-category 0.74 and A-category 0.93. The higher values of the P-category and A-category indicated that these categories of behaviours by problem solvers to analyse the problem situation and behaviours that expressed emotions and self were easier to identify reliably.

Limitations and Conclusions

A number of limitations for this study have resulted directly from the chosen method of the "think aloud" data collection and protocol analysis based on a predetermined taxonomy for encoding behaviours. The contents of the subjects' problem solving protocols were restricted to what subjects were able to verbalise during the process. The possibility of observer bias in making inferences from the encodings of the subjects' responses cannot be denied. It is a well known fact that no matter how carefully a research strategy has been designed to be objective, the method of behavioural analysis through the use of verbal data and encoding categories has to contend with some distortions and incompleteness on the part of the subjects' verbalisation and the inconsistency in inferences on the part of the coder. The Taxonomy that was developed and used throughout this research, is just one of
the possibilities which could have been developed in such a study, depending on the position adopted by the investigator on the phenomena under discussion. This would eventually determine what the investigator observed and to some degree influence the regularities and patterns of behaviours that might be identified.

The task of coding was time consuming and required great concentration on the part of the coder. It was not feasible to engage an inexperienced coder for the task. A good level of mathematical background and teaching experience to detect the nuances and complexity of thinking in the students' problem-solving processes for a semantically rich domain like mathematics was found to be essential. The coding system developed for this study required coders to be able to identify a wide range of behaviours, including those that deviated from the ideal. Coders needed to be able to detect any errors in mathematical structure, strategies that could lead the problem solver astray or mathematical insights which an untrained person might not have.

Of the total 943 encodings across all problems for a sample of nine subjects of varying abilities, inter-coder agreement was relatively high at 83.4%. Except for some behaviours which were relatively difficult to code or which had few occurrences, the majority of the 29 behaviours in the taxonomy yielded consistent judgement between the different coders. The inter-coder agreements resulted from the two stages suggest that the problem solving behaviours investigated in the present study can be regarded as having been assessed reliably. The final taxonomy that evolved was intended to be used as a tool for investigating problem solving processes in a variety of non-routine mathematical tasks for a variety of problem solvers in the main part of the research project.
REFERENCES


APPENDIX A

THE PROBLEM SOLVING TASKS
FIVE NON-ROUTINE MATHEMATICAL PROBLEMS

1. LOANS
Sally loaned $7 to Betty. But Sally borrowed $15 from Estella and $32 from Joan. Moreover, Joan owes $3 to Estella and $7 to Betty. One day the girls got together at Betty’s house to straighten their accounts. Which girl left with $18 more than she came with?

2. PALINDROMES
A number like 12321 is called a palindrome because it reads the same backwards and forwards. A friend of mine claims that all palindromes with four digits are exactly divisible by eleven. Are they?

3. CHICKEN-COOP
The Smith family wants to build a chicken coop and they have bought enough wire for 19 metres of fence and one gate that is one metre wide. They have decided to make the coop rectangular or square. What width and length would they make the coop so that the chicken can have the largest possible area inside the coop?

4. DIAGONALS
A diagonal of this 5 x 7 rectangle passes through 11 squares (shaded). Can you find a way of forecasting the number of squares passed through if you know the dimensions of the rectangle? How many squares will the diagonal of a 1000 x 800 rectangle pass through?

5. Xmas Presents (this task was used only in the pilot study)
Mrs. Tan is buying Christmas presents for her six children to give one another. Each child gives a present to each of the others. How many presents must she buy?
APPENDIX B

THE FINAL TAXONOMY WITH A SYSTEM OF CODES

P-CATEGORY: PROBLEM-ORIENTATION HEURISTICS

P1 - First reading of the whole problem
P2 - Re-reading the whole or part of the problem
P3 - Paraphrase problem statement
P4 - Examine dimensions
P5 - Construct Representation

H-CATEGORY: PROBLEM-SOLUTION HEURISTICS

H1 - Recall similar problem
H2 - Draw diagram(s)
H3 - State an answer
H4 - Look at particular or simple cases
H5 - Make a guess and check
H6 - Search for a pattern
H7 - Generalise to a rule
H8 - Make a logical deduction
H9 - Check computation or final answer

K-CATEGORY: DOMAIN-SPECIFIC KNOWLEDGE

K1 - Apply arithmetic algorithm
K2 - State a fact, principle or theorem
K3 - Apply routine mathematical procedures
K4 - Mathematical misconceptions

M-CATEGORY: METACOGNITIVE BEHAVIOUR

M1 - Suggest a plan
M2 - Assess task facility
M3 - Review progress
M4 - Recognise error
M5 - Recognise new development
Q1 - Task-relevant self-question
N1 - Task-irrelevant rhetoric

A-CATEGORY: AFFECTIVE BEHAVIOUR

A1 - Negative self-evaluation
A2 - Giving up
A3 - Emotional expression
APPENDIX C

SOME OPERATIONAL DEFINITIONS OF PROBLEM SOLVING BEHAVIOURS

The operational definitions helped to specify the phenomena i.e., the problem solving behaviours under investigation in the present study, to provide a basis for distinguishing between various items of the taxonomy. For a better introduction to the general nature of the phenomenon to be observed, some examples from the subjects' problem solving protocols were extracted to provide a frame of reference for the operational definitions.

Every behaviour was assigned a code eg. P3 for "Paraphrase problem statement" and if the behaviour was incorrect, inappropriate or irrelevant to the solution then a bar was drawn over the code eg. P3 for "Misinterpret problem statement". The following are sample definitions of selected behaviours in the five categories of the taxonomy:

Problem-Orientation Heuristics

e.g. P3 - Paraphrase problem statement.
Attempt to give meaning to a certain word or to rephrase the problem in their own words.

P3 - Misinterpret problem statement.
A subject sometimes misinterpreted the meaning of a certain word or of a certain part of the text from the problem.

Some examples of code P3:

a. "...loan and borrow.....owes and borrow is the same."

b. "...so in other words, Sally again borrowed from Betty $7 and borrowed $15 from Estella.."

c. "..Can you find a way of forecasting ....forecasting means what.?...ah..Is it to predict .......... Yes.."

Problem-Solution Heuristics

e.g. H5 Make a guess and check
Subject may predict or expect a certain result while considering a particular case and then proceed to check it.

Some examples of code H5:

a. ".....so, I’ll expect 3 by 7 to be 9...let’s check that 1,2,3...1,2,3,4,5,6,7,...Ok, I expect 9 squares cut ...
   ...1,2,3,4,5,6,7,8,9...yes."

b. ".....6 by 7 ...uhm... Ok, I have a rectangle now, let me see if I am right...my guess will be 14."
e.g. H6

Search for pattern
This behaviour follows from H4 or H5 where subjects, after trying some cases, attempted to identify relationships for a pattern or common property among them.

Some examples of code H6:

a. "....Ok, let me see the relationship again...if it's 5 x7 is 11, 10 by 8 is 16, 6 by 4 is 8...."

b. "....let's try and get pattern...2 by 3 you have 4 squares, 2 by 4 you have 4 squares,2 by 5 ....6."

c. ".... there must be some properties of 11 that I can make use of.."

Domain-Specific Knowledge

e.g. K3 - Apply routine mathematical procedures
Subjects may apply domain specific routine procedures in algebra, geometry, calculus or others in their solution paths.

K3 Inappropriate application of procedures

Structural errors may stem from a misunderstanding of the problem or some of the principles necessary for its solution. However, any computational error made while manipulating these procedures is coded as K1.

Sometimes subjects may try to use certain mathematical procedures which are inappropriate to the solution of the problem.

Some examples of code K3:

a. "....I can do it this way...ABBA is equal to 1000A + 100A +10B +1A , alright...and then group them together and get 1001A + 110B, right ."

b. "....can use calculus to calculate...so that gives me ...differentiating with respect to small a ,we get 10 - 2a with zero, we get a=5 and b=5.."
Metacognition

e.g. Q1 - Self-questioning
In a problem-solving situation it is quite natural for subjects to ask questions of themselves. These self-questions could be task-specific and were often junctures at which decisions had to be made. They could also be rhetorical questions which were usually self-related and task-irrelevant.

Some examples of code Q1:

a. "....let's say if it's a rectangle in shape...can I have other possible numbers?"
b. "....Can I use any mathematical facts or not?"
c. "....If one of the side is odd...what happen if this odd side is the smaller one?"
d. "....But is this general enough?"

e.g. M1 - Suggest a plan or subgoal operation
Subjects might consider the possibility of using a particular plan or subgoal operation. This might be a global plan or hypothesis where the subjects could see the solution and sometimes it could be just an exploration of ideas.

Some examples of M1:

a. "So we need to generalise first and then find answer for the specific dimensions of 1000x800."
b. "....Ok, let's draw a table and see if I can generalise anything from there."
c. "....so at the beginning, if we can restraint...that is, if we can reduce squares into smaller sizes."
d. "Let's look at how much each person owes the other and then deduce from there."

e.g. M2 - Assess the task facility
When presented with the task some subjects would quite spontaneously perceive it as easy, difficult or confusing. Sometimes in the middle of the solution the subject may say a certain operation is easy or hard to do.

Some examples of code M2:

a. "let me see...it's simple...straightforward."
b. "If it's a square...then it'll be easier."
Review Progress
Subjects may review and evaluate progress at any stage of the solution path, in terms of adequacy and reasonableness of any results obtained; method used; or outcome of any action or decision made.

Some examples of code M3:

a. "What I'm doing is I'm adding it this way...Ok,.. no, that's not very good."

b. "This method is shorter."

c. "Actually I don't have to consider 6 by 12 ... I just take small numbers and try to see if I can generalise from a few specific examples."

d. "....Er...ummm...ABBA..let's go back to this...in general ......cannot find anything general..it doesn't follow any pattern."

e.g. M4 - Recognise error
For some subjects, evaluation is an on-going process in solving a problem and sometimes they would review a step and could than instantly recognise an error or failure before any final result was obtained.

Some examples of code M4:

a. "....Mm..no,not necessary true...100 and..uh..that’s wrong."

b. "....no..you can’t go on to a number that is larger.."

c. "....I think the first solution that I gave is wrong."

d. "....Umm..I think I make a mistake.

e. "....Yes, something is wrong here because it’s said that given 19m of fence..cannot..cannot.."
Affective Behaviours

**e.g. A1 - Negative Self-evaluation**
Subjects assess their own ability, confidence or personality. Express self-doubt, they say they don’t know how to do; can’t remember; don’t understand; not sure etc..

Some examples of code A1:

a. "....No, I don’t know how to do !"
b. "....Umm..not again!.I can’t do!"
c. "....Aiyah, my math is not working !"
d. "....If I can remember, I also can’t do it now because I don’t have the log book"
e. "....Hiyah, anyway, I just can’t remember how to solve.. If I revise my work, I think I’ll know ."

**e.g. A2 - Giving Up**
Lack of persistence. Subjects in this study were not given the chance to ask for help if they were stuck. This behaviour was coded when subjects indicated their intention to quit the task.

Some examples of code A2:

a. "I give up, lah..don’t do already."
b. "Hiyah, I think I give up, I just don’t know how to do!
c. "I surrender already...Ok, I surrender here."

**e.g. A3 - Negative emotional responses**
Emotions can manifest physiologically as muscle tension or rapid heartbeat. In this study using “think-aloud” protocols, such arousal cannot be detected, however evidence of emotions can be identified from their verbalization if they are explicit. Tonal expressions or exclamations of frustration could be heard as sighing or as irritated noise such as "Ich ! Ich! ", " Eee..", "Hiyah..", " Yuck ! ", " Oh, dear.." etc.

Examples of A3:

a. "....aiyah!.Let’s see..hiyah ! [sigh]...
b. "....Oh, my goodness...this is bad..
c. "....Where got such question ! Aiyah.[sigh]"
# APPENDIX D

## A SAMPLE OF THE PROTOCOL ANALYSIS FOR A SOLUTION TO THE LOAN-PROBLEM

<table>
<thead>
<tr>
<th>Segment Code</th>
<th>Verbatim Transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. P1:</td>
<td>&quot;Sally loaned $7 to Betty. But Sally borrowed $15 from Estella and $32 from Joan. Moreover, Joan owes $3 to Estella and $7 to Betty. One day the girl got together at Betty's house to straighten their accounts. Which girl left with $18 more than she came with?&quot;</td>
</tr>
<tr>
<td>2. P2:</td>
<td>&quot;Ok, one more time...Sally owes $7 to Betty...(repeats reading the question).&quot;</td>
</tr>
<tr>
<td>3. M2:</td>
<td>&quot;Umm...this is quite confusing.&quot;</td>
</tr>
<tr>
<td>4. M1:</td>
<td>&quot;Ok, I'll consider each person one at a time.&quot;</td>
</tr>
<tr>
<td>5. H2:</td>
<td>&quot;S...B...E...and J. (drawing a table)&quot;</td>
</tr>
<tr>
<td>6. P4:</td>
<td>&quot;Sally loaned $7 to Betty, so Sally will receive $7 in the end and Betty returns $7, so plus 7 here and minus 7 here.&quot;</td>
</tr>
<tr>
<td>7. P4:</td>
<td>&quot;But Sally borrow 15...borrow 15...so minus 15 from Estella...so from E put +15.&quot;</td>
</tr>
<tr>
<td>8. P4:</td>
<td>&quot;$32 from Joan...so minus $32 for Sally and Joan plus $32&quot;</td>
</tr>
<tr>
<td>9. P4:</td>
<td>&quot;Moreover, Joan owes $3, so she must minus 3 to E, E must plus 3 and owes $7 to Betty so another minus 7 to B, B must plus 7, alright&quot;</td>
</tr>
<tr>
<td>10. P2:</td>
<td>&quot;One day the girls got together and...which girl left with $18?&quot;</td>
</tr>
<tr>
<td>11. M3:</td>
<td>&quot;Now is easy..to add up for each one.&quot;</td>
</tr>
<tr>
<td>12. H4:</td>
<td>&quot;Sally, cannot because she's still in debt.&quot;</td>
</tr>
<tr>
<td>13. H4:</td>
<td>&quot;Er..Betty..is zero..ya..&quot;</td>
</tr>
<tr>
<td>14. H4:</td>
<td>&quot;Estella got $18 right, plus&quot;</td>
</tr>
<tr>
<td>15. H4:</td>
<td>&quot;and Joan has 32 minus 10, $22 to take home&quot;</td>
</tr>
<tr>
<td>16. H8:</td>
<td>&quot;So looks like Estella is the girl.&quot;</td>
</tr>
</tbody>
</table>
| 17. H3:      | "The answer is Estella left with $18 more than she came with."
Sequence of behaviours in string of codes in Protocol IA:

Pl --> P2 --> M2 --> M1 --> H2 --> P4 --> P4 --> P4 --> P4
----> P2 --> M3 --> H4 --> H4 --> H4 --> H4 --> H4 --> H8 --> H3.

Conversely the sequence of codes can be interpreted as:

Subject first read the problem (Pl), reread the problem (P2), found it confusing (M2), suggested a plan to organise the given information (M1), by tabulating (H2), then examined the dimensions of the problem statement by breaking down into the table drawn (P4, P4, P4, P4,), reread part of the question (P2) then reviewed to sum up the numerical relationships (M3) by looking at specific cases of the borrowers and lenders (H4, H4, H4, H4) and making a logical deductions (H8) and finally stated the correct answer (H3).

[ time taken : 2 min. 15 sec. ]