PRIMARY STUDENTS’ REASONING IN PROBLEM SOLVING
AND TEACHERS’ EVALUATION OF THEIR ARGUMENTS

Annita Monoyiou *, Xenia Xistouri ** & George Philippou ***
Department of Education, University of Cyprus

We examine 5th and 6th grade students’ ability to reason during problem solving activity and teachers’ evaluation of their arguments. Three tasks were distributed to 236 students asking them to decide on the conclusion and justify their decisions. Indicative examples of the students’ responses were given to 16 teachers for assessment during semi-structured interviews. The results suggest that a considerable proportion of students provide no mathematical justification and another proportion supported their argument on numerical examples. Some teachers were found to value justifications based on numerical examples as equally good and occasionally even better than mathematically valid statements. It seems that any effort for improvement should start from changing teachers’ views and didactical processes.

INTRODUCTION

Principles and Standards for School Mathematics NCTM (2000) draw attention on developing of students’ mathematical reasoning, as well as on the assessment of this competence. Teachers should encourage students to justify their assertions and statements, and search for new methods and means to develop students’ mathematical reasoning. However, it is not easy to specify the type of arguments that should be expected by students and the kind of reasoning that should be taught to primary students. Research shows that not all students’ statements and arguments in mathematical problem solving (MPS) are mathematically valid arguments (see e.g., Evens & Houssart, 2004). Students often reason according to their personal experiences, and teachers who seek to understand what is actually behind an argument should escape their “egocentricity” and think through a child’s perspective (Tang & Ginsburg, 1999). Therefore, teachers’ assessment of students’ arguments is essential to developing of students’ mathematical reasoning. However, no piece of research seems to have investigated how teachers appraise students’ arguments.

THEORETICAL BACKGROUND AND AIMS

Mathematical reasoning or justification is a type of “weak proof” for a mathematical assertion. Russel (1999, p.1) argues that reasoning refers to “what we use to think about the properties of these mathematical objects and develop generalizations that apply to whole classes of objects”. Recent studies (e.g., Pehkonen, 2000) suggest that primary students have difficulty in mathematical reasoning. It is, however, important in Mathematics teaching to let students develop the habit to ask for reasons and provide arguments in their mathematical activities as a preparation for the ultimate goal, which is to produce formal proofs in high school.

Monoyiou, Xistouri & Philippou

In this study we adopt the categorization of students’ arguments proposed by Evens and Houssart (2004), which refers to reasoning in MPS; they propose four types of responses: 1) wrong or irrelevant, 2) restatement or reinforcement, 3) providing numerical examples and 4) justification. The first type refers to responses that are irrelevant to the solution of the problem, either due to incorrect course of solution or to arguments that are not rationally connected to the problem. The second type refers to mere restatements of the data, most likely in ones’ own words, without any substantial addition to already given information. The third type concerns arguments limited to direct or indirect use of examples, a type of justification that might be accepted for primary students, as non-well articulated inductive reasoning. The last type refers to responses that have the element of generalization, without being based on testing examples. The same authors found that a large percentage of 11-year olds (42%) managed to give some form of valid mathematical reasoning, even with weakness in expression. On the contrary, similar studies (i.e. Healey & Hoyles, 2000) suggest that students support their responses on testing examples.

Assessment in mathematics is the process of gathering information concerning students’ mathematical abilities, to be used for various educational purposes (Lappan & Briars, 1995); it should not be considered as the final part of teaching. Assessment is an integral part of teaching, giving feedback to teachers about the efficiency of the teaching/learning process; it is an aid to adjust and redesign their teaching in view of the outcomes (Cooney, Badger & Wilson, 1993). Since problem solving is at the heart of mathematics, it should also be at the heart of assessment (Lester & Kroll, 1990). Assessment of MPS gives a measure of the level of success of the learning process, though assessment tasks are frequently limited to routine-problems and short-answer questions asking reproduction of knowledge (Webb, 1992). Teachers rarely ask students to give written reasons, due to time pressure and students’ difficulties to express their thoughts in writing (Philippou & Christou, 1997).

In the light of the above discussion, the aim of this study was to examine primary school students’ ability to reason in problem solving and to investigate how teachers assess students’ reasons in MPS. The research questions were:

1. How able are 5th and 6th grade students to reason in MPS and what kind of arguments do they give?

2. How do the teachers conceive and appraise students’ arguments in MPS?

METHODS

Participants were 236 primary school students of the 5th and 6th grade, 120 boys and 116 girls from six schools. Students were given about 40 minutes to consider the following three tasks, state whether they agree or disagree with Mary and explain their reasons in writing:

**TASK 1:** Consider a rectangle with 6 cm in length and 4 cm wide. If we half the width of the rectangle and double the length, we see that the area remains the same. Mary says: “This does not stand for all rectangles.”
TASK 2: Mary tried several examples to check the sum of two odd numbers. She tried: $1+3=4$, $3+5=8$, $7+3=10$, and concluded: “If you add two odd numbers, you will never have an odd sum.”

TASK 3: The rule for generating the number sequence: $1, 4, 7, 10, 13, 16 ...$ is “add 3 each time”. Mary says: “No matter how far you go, there will never be a multiple of three in the sequence.” (Evens & Houssart, 2004).

Based on the categories proposed by Evens and Houssart (2004), students’ responses were assorted in five types: nothing on script, wrong or irrelevant, restatement, numerical examples, and justification. Each category is presented with progression from the least to the most sophisticated answers, when applicable.

Semi-structured interviews with 16 teachers of the participating schools were conducted. They were asked to mark some of the students’ arguments, on a scale 0 to 5. The arguments presented to the teachers were examples that covered each of the categories for each of the three tasks. The quotes were given one after the other from the simpler one to the most sophisticated.

FINDINGS

Table 1 summarizes the frequencies of students’ responses on each of the tasks. Clearly, in each of the three tasks, about one third of the students either provided no justification or gave wrong or irrelevant answers.

<table>
<thead>
<tr>
<th>(N=236)</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasons</td>
<td>f</td>
<td>%</td>
<td>f</td>
</tr>
<tr>
<td>Nothing on script</td>
<td>19</td>
<td>8.1</td>
<td>22</td>
</tr>
<tr>
<td>Wrong or irrelevant</td>
<td>65</td>
<td>27.5</td>
<td>51</td>
</tr>
<tr>
<td>Restatement</td>
<td>32</td>
<td>13.6</td>
<td>56</td>
</tr>
<tr>
<td>Examples given/tested</td>
<td>98</td>
<td>41.5</td>
<td>99</td>
</tr>
<tr>
<td>Some degree of justification</td>
<td>22</td>
<td>9.3</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 1: Frequencies of students’ reasons given in each task

In all tasks, the highest proportion of students justified their answer on the basis of numerical examples (41.5%, 41.9%, 34.3%). It is noteworthy that some students simply restated the information already given in the question (13.6%, 23.7%, 8.5%). The lowest proportion of mathematically acceptable explanations was given in Task 2 (3.4% of the students), while the highest proportion of acceptable explanations was given in Task 3 (22.5% of the students) and less than ten percent of students gave valid arguments in Task 1 (9.3%).

The above five categories were next analysed on the basis of specific students’ responses in each task separately. Following the analysis of each task we present and discuss the teachers’ appraisals of each response type.

Task 1

Wrong or irrelevant answers: In task 1, some students gave explanations that were irrelevant with Mary’s statement for example: “This does not stand for all rectangles
because simply the length gets larger and the width gets smaller” (S1). Many students
failed to work out the example correctly, for instance “I don’t agree because 4×6=24 but
2×18=36”. Others worked Mary’s numerical example right, but argued that this does
not stand for all rectangles because: “The area of the rectangle with sides 6cm and 4cm is
24. If the sides are 12cm and 2cm the area is again 24 but this does not happen in all
rectangles”. Other students thought that since the dimensions of the rectangle change,
the area would also change: “Since the two sides change the area will also change”.

Restatement: Some students simply restated the information already given in the
question: “I agree because if we split the width and double the length the area will remain
the same” (S2).

Numerical examples: Some students tried to explain their answer by using Mary’s
numerical example: “Because 6×4=24 and 12×2=24, so the area is the same”. Some
children also drew Mary’s rectangle: “I drew a rectangle, I multiplied the one side and I
divided the other and the area remained the same”, while other students went beyond the
example already given applying their own examples. Some students gave additional
examples “If the one side is 8 and the other 10 the area is 80. If 8 became 4 and 10 became
20 then 4×20 is again 80” (S4), while others just mentioned that they worked some
“This stands for all rectangles because I tried others as well” (S3).

Justifications: Some children justified their answer by the argument we multiply the
length and divide the width with the same number, though not making finite mention
why the area remains constant: “This stands for all rectangles because we multiply one
side and divide the other with the same number”. Other children moved further, making
the general statement: “Because multiplication and division are reverse operations,
divided by two and times two” (S5).

Table 2 shows that most teachers (N=11) gave no marks for irrelevant answers. Half
of the teachers gave more than 3 points to simple restatement, arguing, “It’s correct. It
seems that he/she understands Mary’s statement”. The most accredited response seems to
be Statement 4 (S4), which gets 3 points or more, from all the teachers. It is
noteworthy that teachers who gave high grade to S4 argued, “The student gave a clear
example. He/she explained very well. S5 is not clear. It needs an example”. This shows
that all teachers accept reasoning by arithmetical examples, as even better than actual
justification, which received less than 3 points; one teacher gave zero to S5 arguing,
“I can’t understand this thought. He/She must explain better by giving an example like in
S4”. It needs to be noted that although both S3 and S4 justify by examples, S3 was
graded worse because the examples were not given.

<table>
<thead>
<tr>
<th>Statements</th>
<th>Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 1 2 3 4 5 N</td>
</tr>
<tr>
<td>S1: Irrelevant</td>
<td>11 1 4 0 0 0 16</td>
</tr>
<tr>
<td>S2: Restatement</td>
<td>0 2 3 3 7 1 16</td>
</tr>
<tr>
<td>S3: Examples (not given)</td>
<td>0 1 3 5 0 7 16</td>
</tr>
<tr>
<td>S4: Examples (given)</td>
<td>0 0 0 4 5 7 16</td>
</tr>
<tr>
<td>S5: Justification</td>
<td>1 0 2 1 5 7 16</td>
</tr>
</tbody>
</table>

Table 2: Teachers’ Assessment of Task 1
Task 2

Wrong or irrelevant answers: Some students gave irrelevant answers such as: “It can’t be an even number all the times”. A common mistake among the students was to add an odd number with an even number: “Because 3+2=5... so it’s not always an even number”. One student added three odd numbers, found an odd sum and rejected Mary’s statement: “Because if I add 3+3+3 then the sum is an odd number, 9” (S1).

Restatement: Again some students restated the data of the task: “Because that’s how it always goes” (S2), “If we add two odd numbers the sum will be an even number” (S3).

Numerical examples: In task 2, few students gave reasons using the examples already given in the task: “I did the additions given and the sum is always an even number”. Other students gave their own examples such as: “I added many odd numbers 9+9=18, 9+5=14, 1+3=4 and the sum is always an even number” (S4). Some students provided numerical examples by adding two-digit odd numbers such as: “35+35=70, 53+57=110 so when you add odd numbers the sum is even number”. Other students simply mentioned that they tested numerical examples but they did not provide any: “I did several additions with odd numbers and the result was always an even number.”

Justification: Some students gave a form of valid justification by arguing that: “Because each odd number is one more than even” (S5). One student tried to explain it more extensively by using a numerical example as an aid to express the general rule s/he had in mind: “I said 7+5 take away 1 will became 6 and another 1 from 5 will become 4. If we add them the sum will be even and if we add the two it will be even again” (S6).

Table 3 shows that again most teachers gave no marks to irrelevant responses arguing, “This student didn’t understand the problem. He/She added three instead of two odd numbers”. As far as restatements are concerned, S2 was granted no marks from most teachers, while some teachers gave marks and one teacher gave full marks. The teacher who gave full marks argued that “The student seems to understand the problem but he/she can’t express his/her thoughts” while the teachers who did not give marks argued “He/she does not explain. His/her justification is not mathematical”. A longer restatement (S3) received better marks, while there appears again lack of homogeneity in teachers’ grading. Almost half of the teachers gave 3 or more points to this statement arguing, “It is correct. He/She could explain better or give an example but he/she is in a correct path” while the others gave less than 3 arguing, “He/she doesn’t explain at all. He/She simply restates the data given in the task”. Reasoning by example was highly received in this task, even higher than justification. Many teachers mentioned that although S5 and S6 are correct, students could enrich their justification by the use of examples like those in S4. Others did not understand students’ arguments, due to poor language expression.

<table>
<thead>
<tr>
<th>Statements</th>
<th>Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>S1: Irrelevant</td>
<td>10</td>
</tr>
<tr>
<td>S2: Restatement (short)</td>
<td>10</td>
</tr>
<tr>
<td>S3: Restatement (long)</td>
<td>4</td>
</tr>
</tbody>
</table>
Wrong or irrelevant answers: In task 3, a few students gave irrelevant answers such as: “I agree with Mary because three is not an even number so there isn’t a multiple of three in the sequence”. Some students argued that if the sequence continues there will be a multiple of three without giving an explanation: “Mary is wrong. There are multiples of three in the sequence”. Some students made numerical mistakes such as: “Because 1, 4, 7, 10, 13, 16, 19, 21 (3×7=21), so there is a multiple of three. Mary is wrong” (S1).

Restatement: Restatements in this task were of the type: “If you add three each time there will never be a multiple of three in the sequence” (S2).

Numerical examples: In task 3, some students provided explanations using the numbers already given: “The numbers 4, 7, 10, 13, 16 are not divided by three. There aren’t multiples of three”. Some students continued the sequence to justify that there are no multiples of three: “I continued the pattern 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52. There aren’t multiples of three” (S3). Other students mentioned that the sequence continues without mentioning the numbers: “I continue the sequence and there isn’t a multiple of three” (S4). One student continued the sequence but he/she stopped at 31, arguing that there is a pattern at the unit digit: “I continued the sequence but I stopped at 31. There is a pattern at the units 1, 4, 7…. Until 31 there isn’t any multiple of three” (S5).

Justification: Some children focused at the starting point of the sequence providing a valid form of justification: “I have to start from 3 or 0 to have a multiple of 3” (S6). Other students focused on comparing the numbers of the sequence with the multiples of three: “Because it is one bigger from multiples of 3”. Some students mentioned both the starting point of the sequence and the comparison of numbers: “Because it starts from 1 not from 0, the numbers will always be one more than the multiples of 3” (S7).

Table 4 summarises the teachers marking of students’ arguments. Clearly, most teachers give no marks to wrong response, though some of them would appreciate students’ efforts arguing, “He/she tried to continue the sequence… He/she just made a numerical mistake”. There is again lack of homogeneity in teachers’ grading as far as restatement is concerned. Some teachers referred that the student simply restated the data while others argued, “He/she understands the problem. It’s correct” giving 3 or more points. Examples received high marks again with S4 receiving relatively lower marks because the examples were not given. Once again teachers supported numerical examples as valid forms of justification and not many expressed the need for a general rule. Although in this task justification received higher marks than numerical examples, most of the teachers referred “Students in statement 6 and 7 could give an example. They could continue the sequence like student in S3”.

Table 3: Teachers’ Assessment of Task 2

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>5</th>
<th>5</th>
<th>5</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>S4: Examples</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5: Justification</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>S6: Justification (General rule)</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 3: Teachers’ Assessment of Task 2

Task 3
### Table 4: Teachers’ Assessment of Task 3

<table>
<thead>
<tr>
<th>Statements</th>
<th>Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1: Wrong (numerical mistake)</td>
<td>11 2 1 2 0 0 0 16</td>
</tr>
<tr>
<td>S2: Restatement</td>
<td>5 3 3 3 1 1 1 16</td>
</tr>
<tr>
<td>S3: Examples (given)</td>
<td>0 0 1 5 6 4 16</td>
</tr>
<tr>
<td>S4: Examples (not given)</td>
<td>2 0 4 5 4 1 1 16</td>
</tr>
<tr>
<td>S5: Examples (pattern at the units)</td>
<td>0 0 0 3 9 4 16</td>
</tr>
<tr>
<td>S6: Justification (Starting point)</td>
<td>0 0 1 1 6 8 16</td>
</tr>
<tr>
<td>S7: Justification (Starting point/Comparison)</td>
<td>0 0 0 1 4 11 16</td>
</tr>
</tbody>
</table>

#### CONCLUSIONS

The findings of this study indicate that an alarmingly large proportion of the students were unable to give a relevant response, while the majority of the remaining gave arguments based on numerical examples. The latter is in line with results by Healey and Hoyles (2000), who argue that preference to using numerical examples, as opposed to accepted forms of proof, is found even amongst older students. Arguing by example should not be surprising, as it may form the basis of inductive reasoning, provided one guards against overdue generalization. Though Mathematics is renowned as prime area that offers the chance to develop students’ ability to reason, the outcomes seem to fall short of objectives. This is in line with earlier findings (Evens & Houssart, 2004; Pehkonen, 2000), though our findings indicate a wide variation of students’ arguments within each category.

The situation seems to be more complex regarding teachers’ assessment. Apparently, teachers’ appraisals are based on subjective criteria and differ far from one another. This was evidenced in the range of points they proposed in responses classified as restatement, where some teachers found them as good answer giving high grades, while others gave low grade because students simply rephrased. In the case of numerical examples, most teachers gave high grades. It is noteworthy that they were graded evenly with mathematical justification and occasionally higher. It is important that, even in the case of actual justification, students’ responses did not receive high marks, due to poor expression, which made their statements not explicable to the teachers. The teachers’ trend to accept as valid, argumentation by example may contribute to and enhance the students’ conception about the validity of this type of argument.

A point of possible focus in teaching and assessing ability to reason is to engage students’ in group discourse asking the classical question “why”, drawing distinction between general properties and special cases, providing simple examples, preferably from everyday life, and counterexamples. So far, it seems that the goal for an early appreciation by students of the meaning and value of reasoning, and the process of “proving” seems to remain simply an ambition. As in most cases change should start from developing and testing in practice paradigms directed to teachers needs; how to initiate discussion, to build on false students’ arguments, encourage analysis of examples, draw attention on possible obstacles, etc.
Monoyiou, Xistouri & Philippou

References


