

USING SOLO TO ANALYSE GROUP RESPONSES

Chris Reading and Christine Lawrie

Centre for Cognition Research in Learning and Teaching

University of New England, Australia

The increased use of group work in teaching and learning has seen an increased need for knowledge about assessment of group work. This report considers exploratory research where the SOLO Taxonomy, previously used to analyse the quality of individual responses, is applied to group responses. The responses were created as part of an activity dealing with weather data and analysed for the quality of the description of variation. The research indicated that the hierarchy developed for coding the individual responses could also be used to code the group responses. Group responses were generally found to be of poorer quality than was suggested as possible by the responses of individuals within the group. Factors that may have contributed to the differing quality of the group responses are considered.

INTRODUCTION

The use of group work in teaching and learning is an important preparation for the collaborative expectations of project work to be undertaken in the 'real world'. As group work becomes more recognized as a legitimate form of learning (Edwards, 2002), and as assessment is an important aspect of teaching and learning, the assessment of group responses is becoming increasingly important. The Structure of the Observed Learning Outcome (SOLO) Taxonomy is being used increasingly to code individual responses: but what about responses submitted by groups? This paper reports on the analysis of group responses, part of a research project where students were required to respond both individually and then as a group.

GROUP ASSESSMENT

When assessing group work there are two major components to be considered, product and process, that is, what is produced and how it is produced. Assessment schemes related to group work often include either or both of these components (Devlin, 2002; Michaelson, 2003). While process is acknowledged as an important factor with group work, the focus in this report is assessing the group product. Various combinations of assessment of group and individual contributions to the group work can be created to produce final assessment grades to be allocated to individual students. Devlin (2002, pp. 7-12) outlines a variety of these combinations, one of which assesses the group product by allocating an overall group mark to each student in the group. To implement such an approach would necessitate assessing the group product and being confident that this was a reasonable representation of each group member's capabilities. However, care would need to be taken when assigning a group mark to individuals, as scores from group assessment are not always valid

indicators of students' individual responses (Webb, 1993, p. 21). The question also needs to be raised as to whether group products can be assessed in a similar manner to individual products to give meaningful information about the group members.

CODING INDIVIDUAL RESPONSES USING THE SOLO TAXONOMY

Assessment of students based on the quality of their understanding and learning is a basic principle of Developmental-Based Assessment (DBA). This approach to assessment, with the mental structure of understanding as paramount, differs from outcomes-based assessment that focuses on what students are expected to know. Importantly, the developmental-based approach to assessment rests on the empirically established cognitive developmental SOLO Taxonomy model (Pegg, 2003, p. 238-239). This taxonomy consists of five modes of functioning, with levels of achievement identifiable within each of these modes (Biggs & Collis, 1991). Although these modes are similar to Piagetian stages, an important difference is that with the SOLO earlier modes are not replaced by subsequent modes and, in fact, often support growth in later modes. A series of levels have been identified within each mode. Three relevant levels are: *unistructural* (U) with focus on one aspect; *multistructural* (M) with focus on several unrelated aspects; and *relational* (R) with focus on several aspects in which inter-relationships are identified. These three levels form a cycle of growth that occurs in each mode and recurs in some modes, with each cycle being identified by the nature of the element on which it is based.

CODING GROUP RESPONSES USING THE SOLO TAXONOMY

SOLO is now widely used to code individual responses to tasks in mathematics (Pegg, 2003). Given the growing popularity of group work the question arises as to whether SOLO can also be used to code group responses. However, no reported research using SOLO to analyse group responses could be found. Given this lack of research, there are two possible approaches that could be taken when coding group responses. The first is to collect group responses to a task and then develop a hierarchy using SOLO as a framework to code the responses. The second is to code the group responses by using a SOLO-based hierarchy that has already been developed for individual responses to a similar task.

RESEARCH FOCUS

The reported research takes the latter approach, posing the question: can a hierarchy developed for individual responses be used to code group responses? Also, if such coding of the group responses is possible, then how useful is the information gained from such a coding of group responses? Especially important is the comparison of the level of group response to the levels of individual responses within the group.

METHODOLOGY

This was an exploratory study with research targeting students in Grades 7, 9 and 11 (aged 13 to 17) in a Australian secondary school undertaking a Weather Activity

consisting of a Scenario based around choosing the most suitable month for a proposed Youth Festival. The stages in this activity, which spread over a number of lessons, included a rainfall segment, a teaching intervention and finally a temperature segment. In each segment the students were asked to describe individually the data for one particular month and then, as a group, compare their individual months to decide on the most suitable month. The teaching intervention involved the introduction of box-and-whisker plots to Grades 7 and 9 and a fuller statistics teaching-sequence for Grade 11. Students worked in self-selected groups of three to five. For more detailed description of the context of the research see Reading (under review). Although students were not specifically asked to describe the variation within the data, the key focus of the assessment of both individual and group responses was the description of variation that occurred within the response.

Detailed analysis of the individual responses is presented in Reading (under review), including discussion of the levels in the hierarchy developed, based on the SOLO framework, and sample responses from the various levels. Two cycles of U-M-R levels were identified representing increasing quality of response, the first cycle U1-M1-R1 based on qualitative descriptions of variation and the second cycle U2-M2-R2 based on quantitative descriptions. This hierarchy was used to code the group responses and then the performance of the groups was considered in relation to the performance of the individual members of the group.

RESULTS

The Reading (under review) description of variation hierarchy, developed from analysis of the individual responses, was found to be appropriate to code the group responses. Both researchers coded the responses independently and then coding levels were compared. Checking inter-coder reliability for the group responses, 17% of responses were initially coded at a different level by the two researchers but ensuing discussion was able to bring agreement on the level in all cases. This compared favourably with 13% initial coding disagreement for the individual responses. Coding for the individual responses and group responses are presented, by group, in Table 1 for the rainfall segment and in Table 2 for temperature. The grade of a group is indicated at the beginning of the group number, e.g., group 1104 is the fourth group in Grade 11. An 'nc' coding indicates that the response could not be coded because it contained no description of variation.

Previous analysis of the individual responses indicated that students often included an inference rather than just describing the data as required (Reading, under review). This resulted in the individual and group tasks both having an inference focus. Also, noticeable during the coding of the temperature group responses was the lack of use of statistical tools introduced in the intervention. This observation is similar to the trend observed when analysing the individual responses (Reading, under review).

Group	Student 1	Student 2	Student 3	Student 4	Student 5	Group
701	nc	U1	M2			M1
702	U1	U2	M1			U1
703	M2	U1	M1			M1
704	U2	M2	U1	nc		M2*
705	M1	M2	U1	U2		M1
706	M1	U1	M1	NA		U1
901	U1	U2	U1	U1		U1
902	R1	U1	M1	U2	nc	U1
903	U2	M1	M1			M2*
904	R1	M1	M1			U1
905	M1	nc	M1			R1*
906	M1	M2	R1			U1
907	M2	U1	M1	M1		U1
908	M1	U1	M1	M1		U1
1101	U2	nc	U1	M1		U1
1102	U1	U1	U2	U1		U1
1103	M1	M1	M1	M1		U1
1104	M1	U1	M1	R1		U1

* indicates group responses 'better than' or 'equal to' the best individual response within the group

Table 1: Rainfall segment response coding

Group	Student 1	Student 2	Student 3	Student 4	Student 5	Group
707	M2	M2	U1	nc		M2*
708	R2	M1	nc	U1		M2
709	M2	nc	U1	U2		U1
710	U2	nc	M2	nc		M2*
711	M2	U1	U1	U1	U2	U1
909	U2	U2	M2			U2
910	M2	M2	M2			U1
911	U2	M1	U1	U1		U2
912	U1	U2	U1			R1
913	M1	U2	U2			M1
914	U2	U2	M1	M2	U2	U1
915	M2	M1	M2	M2		M2*
1105	M2	R2	R2	M2		R1
1106	M2	U2	M2			M2*
1107	U2	U2	M1	M2	U2	nc
1108	U2	M2	U1	U1		M2*
1109	R1	M2	M2			U1

* indicates group responses 'better than' or 'equal to' the best individual response within the group

Table 2: Temperature segment response coding

Once coding of the group responses had been completed the level was compared to that of the highest level of individual response within the group, considering whether the group response was 'better than', 'equal to', or 'poorer than' the best individual response. The 21% (8/35) of group responses that were better or equal (see asterisks in Tables 1 and 2) comprise 27% (3/11) in Grade 7, 20% (3/15) in Grade 9 and 22% (2/9) in Grade 11, suggesting similar performance across the three grades. However, the better or equal group responses comprise 17% (3/18) for the rainfall segment compared to 29% (5/17) for temperature suggesting better performance. The number of groups is too small to test for statistical significance. This suggestion of better performance after the teaching-intervention is supported by a comparison between the number of individual students and groups giving second cycle responses. For rainfall 20% (13/66) of the individual responses and 17% (3/18) of the group responses were coded as second cycle. However, for temperature, 60% (39/65) of individual responses and 47% (8/17) of the group responses were in the second cycle.

Of the eight groups with equal or better responses only two gave results better than the best individual response. Both groups were from Grade 9 (rainfall), each comprising three students. Of the six groups, spread across the three grades, where the results were equal, allowing for individual responses that were unable to be coded, there were effectively three responses contributing to the group discussion. This suggests that the smaller groups produced the better quality responses. By comparison, the four largest groups, each of five students, all produced a U1 or nc response much weaker than the best individual response.

In an attempt to explain why some group responses were better than the best individual response within the group while others were not, researchers then considered the original wording of the responses, rather than just the coded level. Although space prevents detailing of such observations here, some instances are mentioned in the following discussion.

DISCUSSION

Having determined that the hierarchy developed for coding individual responses was useful for coding the group responses, the next issue to be addressed is the usefulness of the coded level of the group responses. What is the relationship between the quality of response produced by a group compared to the responses produced by individuals within the group? The results of this study support Webb's findings that "scores from group assessment may not be valid indicators of many students' individual competence" (1993, p. 21). However, whereas Webb's results show that many of the group responses are of a higher level than that indicated by post-testing of individuals, group responses in the present study tend to be of poorer quality than those produced by the individuals involved.

When considering what might influence the quality of group responses, the following factors reported by other researchers as having improved the quality of group products are relevant. While, students have indicated the positive influence of putting

ideas together with others, especially those well known to them (Edwards & Jones, 1999, pp. 284-285), researchers have identified as important, the need for students to be more experienced in group work (p. 286) and the need to receive instruction in working as a group (Edwards, 2002, p. 318). The current research acknowledged the positive influence of these factors by encouraging students, with previous group work experience, to work together in self-selected groups. However, the students did not have specific instruction in how to work as part of a group.

The results of the current study suggest that there were few group responses that were equal or better and the following factors may have influenced the quality of the response made by the group.

1. Homogeneity in the thinking of group members

It was noticeable that in some groups the individual responses were more homogenous, i.e., the students were responding at similar coded levels as well as using similar reasoning (based on the original wording of the responses). For both better groups (903 and 905), the students' individual responses displayed homogeneous thinking that may have enabled them to build on their ideas and to work through to a higher level. By contrast, inspection of the many heterogeneous groups (e.g., 1109) indicates the production of a 'residue' response reflecting only what was common or agreeable to all individuals, i.e., the 'popular view', thus producing a poorer quality response.

2. Lack of recognition of the need to cite supporting statistics

It appears that poorer group responses resulted when students did not perceive a need to support discussions with statistics already cited in their individual responses. For example, in group 1105 all four individual responses were higher quality second cycle responses but the group response, although relational, contained no supporting figures and so was coded as first cycle. In another example, group 709 (with two stronger and two much weaker students) calculated the averages for the maximum temperatures for all four months to find the highest; but then apparently group members decided it was unnecessary to include these figures in their response, being satisfied with just describing what they had done.

3. Nature of the data

The better quality responses for temperature produced by individuals may have been due to the different ways that students view the two variables, with rainfall often treated as a dichotomous variable (rain/no rain), students thus perceiving less of a need to include quantitative support for their inference (Reading, under review). Analysis of group responses suggests that the nature of the data also influences the quality of the group response compared to responses of individuals involved. There are more of the better or equal group responses for temperature. A complicating factor when investigating such a notion is that the rainfall data was the focus in the segment prior to the teaching-intervention while temperature was the focus afterwards. However, as no group responses made any reference to the statistical

tools introduced in the teaching-intervention it must be assumed that there was no direct influence on the quality of the group responses. The observed greater proportion of better group responses for the temperature segment is more likely attributable to the differing nature of the data than the teaching-intervention.

4. Group size

As detailed in the results, better responses were given by the groups with only three, or three effective, members. There is the possibility that with fewer contributing members these groups can more easily come to a consensus and perhaps improve the quality of response. However, with smaller group sizes it needs to be noted that that there is also more chance of homogeneity of thinking, and this has already been identified as contributing to a better quality group response.

It might also be considered that the differing purposes of the tasks for individuals (description) and groups (inference) could have an effect on the quality of response to each task. Although students may see the need to describe variation in their individual responses, they may not see the need to include such information for the inferences drawn in the group responses. However, any possible effect of this factor on the quality of the group response is less likely because many of the individual students made inferences as well as descriptions, before working together within the group.

There is clear evidence from this research that the group responses are not reflecting the best of the capabilities of the individuals within the group. Factors most likely to be causing such a trend include homogeneity in the thinking of the groups, the lack of recognition of the need to cite supporting statistics, the nature of the data and the group size. A previously noted limitation of this study, though, is that only the group product was considered, and not the process. However, the importance of process was noticeable. For example, an apparent lack of effort was evident in the non-serious attempt of the group 117 response that was not able to be coded, while the five individual responses in the group were all coded in the second cycle. This suggests that the stages of maturity of the student, or other social factors such as the effect of the dominant student, may be affecting the group responses. However, such factors should be considered as part of the process, the other important aspect of group work but not the focus of this report. Importantly, assessment of the process could also help to shed more light on the factors affecting the relatively poor quality of group responses in comparison to responses of the individuals involved.

IMPLICATIONS

This research has implications for educators as well as researchers. When group work is being assessed the factors identified above need to be taken into consideration and care must be taken with the relationship of that group assessment to individual assessment. Educators also need to remember that overall assessment of both groups and individuals should include aspects of the group process as well. The SOLO levels previously developed for individual responses have proven suitable for coding group

responses but in situations for which a coding hierarchy has not yet been developed, researchers should consider using group responses to directly develop a hierarchy for coding based on the SOLO Taxonomy. Future research could also consider using the SOLO Taxonomy to develop levels to assess the quality of how individuals work together within a group, i.e., the group process.

References

- Biggs, J., & Collis, K. (1991). Multimodal learning and the quality of intelligent behaviour. In H. Rowe (Ed.), *Intelligence, Reconceptualization and Measurement* (pp. 57-76). New Jersey: Laurence Erlbaum Assoc.
- Devlin, M. (2002). Assessing group work. In R. James, C. McInnis, & M. Devlin (Eds.), *Assessing Learning in Australian Universities*, Australian Universities Teaching Committee. Retrieved January 2, 2004 from <http://www.cshe.unimelb.edu.au/assessinglearning/03/group.html>.
- Edwards, J. (2002). Learning mathematics collaboratively – learning the skills. In A. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 313 - 320). Norwich: University of East Anglia.
- Edwards, J., & Jones, K. (1999). Students' views of learning mathematics in collaborative small groups. In O. Zaslavsky (Ed.), *Proceedings of the 23rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 281 - 288). Haifa: Technion – Israel Institute of Technology.
- Michaelson, R. (updated December 2, 2003). Assessing group work, *Briefing papers of the Business Education Support Team (BEST) - Scotland and Northern Ireland*. Retrieved January 2, 2004, from http://cbs1.gcal.ac.uk/lts/RMichaelson_assess_group_work.htm.
- Pegg, J. (2003). Assessment in Mathematics: A developmental approach. In J. M. Royer (Ed.), *Advances in Mathematical Cognition* (pp. 227-259). New York: Information Age Publishing.
- Reading, C., & Shaughnessy, M. (in press). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.) *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Reading, C. (under review). Student description of variation while working with weather data. Submitted to *Statistics Education Research Journal*.
- Webb, N. (1993). *Collaborative group versus individual assessment in mathematics: Group processes and outcomes*. CSE Technical Report 352. Los Angeles: University of California.