

A QUESTIONNAIRE FOR EVALUATING PUPILS' COGNITIVE PATH ABOUT SYMMETRY AT PRIMARY SCHOOL

Simone Brasili, & Riccardo Piergallini

School of Science and Technology, University of Camerino (Italy)

Abstract

The fundamental role of symmetry has to be more closely enhanced in the interplay between mathematics and physics to foster the teaching of the Nature of Science. In our presentation, we explore the positive effect of introducing the modern concept of symmetry viewed as “*sameness within change*”. A teaching-learning sequence (TLS) was conducted to test the challenges of an interdisciplinary approach based on symmetry and invariance in the educational context, namely at the primary school level. In the design of this sequence, solving the game of closing special cardboard boxes provides a fundamental role. The study evaluates how the specific teaching action makes the modern concept of symmetry in principle appropriate for primary school students through manipulative games. We investigate the students' cognitive paths about symmetry during the sequence by analysing pre- and post-sequence questionnaires. The questionnaire comprises four questions with dichotomous choice, items text, narrative text, and open justification. It is structured mainly in three domains devoted to cognitive, affective, and psychomotor dimensions. The analysis is composed of a mixed method approach. The investigations incorporate qualitative data with Text Analytics and Natural Language Processing (NLP) statistics to identify and extract information from pupils' written reflections. Our study also explores whether the emotion experienced by students plays a role in the TLS. Results show that educational activities induce the increasing knowledge and skills of students. In particular, most students interpreted the lessons as experiences rich in stimuli and insights on symmetry and mathematics in general. The findings also bring important suggestions and contents of reflection that teachers can consider for exploiting the potential learning path on symmetry and invariance.

Keywords: *Nature of science, symmetry, questionnaire, cognitive path, primary school.*

1. Introduction

This study is based on an ongoing research project that investigates the concept of symmetry across mathematics and physics to foster the teaching of the Nature of Science. The unitary nature of symmetry across mathematics and science is definite throughout historical and scientific development. However, symmetry has a more far-reaching role in the educational context, even if its role in the teaching/learning process is a story seldom told (Dreyfuss & Eisenberg, 1990). Indeed, symmetry intervenes in different disciplines, and too often, it is presented as a collection of disconnected concepts. According to Darvas (2015, p. 328), symmetry has an important role in mathematics education as a vehicle of a general scientific method applicable throughout the sciences as a set of principles. We introduced an interdisciplinary approach based on the connection between symmetry and invariance, besides consolidating pedagogical-educational practices. Our educational project designed and implemented a teaching-learning sequence (TLS) based on introducing the modern concept of symmetry viewed as “*sameness within change*” (Leikin et al., 1998) at primary school. Previous work (Brasili & Piergallini, 2020) indicates that the changed view of symmetry linked to the search for invariance is theoretically appropriate for primary school students through manipulative games. This mixed-methods study aims at investigating the pupils' cognitive path about symmetry during the sequence. We focus on the pupils' answers to the open question in the post-test. Open questions prompting students to reflect on the insight they gained and sensed during the teaching-learning unit can shed more light on their thought processes. Exploring the pupils' reflections and observations allows examining the extent to which the students have elaborated and mastered the contents and how conscious they were of the whole activity.

1.1. Theoretical framework

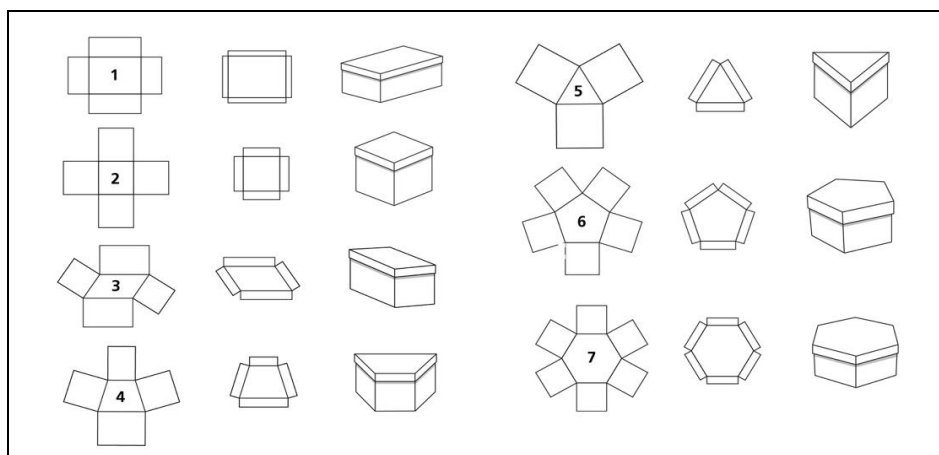
In addition to the visualization and construction process, the reasoning is one of three kinds of cognitive processes in geometry affecting practical functions (Duval, 1998). The activity of thinking about symmetry is not simply about recognizing symmetrical shapes but refers to the deep knowledge and regulation of the cognitive process of symmetry. Authentic learning occurs when there is a change in cognitive schemes (Piaget, 1964). Learning should be considered a dynamic process of modification and restructuring of representative schemes inherent in a given phase and its potential future development.

Moerk (1974) demonstrated close parallels between aspects of cognitive development theory and independent language development. According to Sfard (2008), learning occurs when a definite change in vocabulary, visual mediators, routines and narratives operates in a cognitive perspective. The theoretical framework displays the foundation for investigating the pupils' narratives in terms of keywords and signs showing interaction with the learning path's visual mediators as artifacts, actions and dynamics in the classroom social setting.

2. Materials and methods

Our teaching-learning sequence (TLS) is structured into five activities, three lessons of two hours, and a timeframe of three weeks. The sequence is proposed to a sample of 96 pupils (aged 10) of the primary school in Italy. The researcher proposes various activities, and the teachers participate in the lessons as observers. The main activity is centered on the task game consisting of closing cardboard boxes. The pupils, divided into small homogeneous groups of 4/5 pupils chosen by the teacher, carry out manipulative work following the instructions and filling out the task game questionnaire. The manipulative work consists of folding and taping these cardboard models to obtain boxes with their covers. The covers are designed in such a way that they can only be rotated. The task game is to find in how many ways it is possible to put the cover on each box. This game refers to the rotation symmetries of the differently shaped lids. We take, for example, the triangle-shaped cover. The symmetries of the equilateral triangle can be thought of as the transformations (i.e., rotative actions) that allow covering of its respective box. This way, pupils can play and, in the meantime, deal intuitively with triangle invariants under rotations. The activities are designed to promote discussion and construct the mathematical concept behind the game, comparing the symmetries and invariants for all shapes in Figure 1.

Figure 1. Models of cardboard boxes and covers of 7 shapes.



We evaluate the effectiveness of the TLS by examining the learning changes in the level of pupils' knowledge of symmetry resulting from the teaching-learning session by pre-/post-tests. The questionnaire comprises four questions: dichotomous choice, items text, narrative text, and a task with open justification each addressing a specific aim and learning object. It is structured mainly in three domains devoted to cognitive, affective, and psychomotor dimensions. Our present work focalizes the study on the third open question¹. It aims at letting students to reflect and shape on their personal experiences during the TLS. The resulting inner dialogue should help the student in the authentic personal construction of concepts and the relative regulation in the cognitive process.

¹Question 3: Tell the "Little Prince" the activities carried out in the classroom, what you liked, what you learned about symmetries and if you had any difficulties.

The analysis is composed of a mixed method approach. The investigations incorporate qualitative data with Text Analytics and Natural Language Processing (NLP) statistics to identify and extract information from pupils' written reflections. The use of text analysis technologies from NLP reconciles the qualitative character of semantic meaning with the quantitative aspect of their pertinence (Wiedemann, 2016).

The qualitative and quantitative approach for empirical investigation aims at broadening the dimension level of study understanding and reducing errors of interpretation. We addressed the analysis in three main dimensions: (D1) the emotional and affective dimension; (D2) the cognitive dimension of learning gain; (D3) the possible presence of a cognitive conflict in processing information on the difficulties encountered.

The first step is to identify the principal lexicographic information in the pupils' narratives, which form our corpus, using the text analysis tool *Textalyser* (Gregori-Signes & Clavel-Arroitia, 2015). The results obtained are illustrated in Table 1. The set of distinct words (i.e., lexemes or word tokens) determines the lexical size of vocabulary (V). We indicate the number of i-times recurring lexemes with the i-th vocabulary V_i . Indeed, the vocabulary V is the sum of each single vocabulary V_i starting from the words that occur only once in the text V_1 (i.e., hapax) up to the vocabulary with the maximum number of occurrences $V_{f_{max}}$ where f_{max} indicates the maximum frequency.

Table 1. Complete lexicometry measurements report of corpus (elab. *Textalyser*).

Total word count (Tokens, N)	4.452	Total number of characters	26.523
Distinct words (Types, V)	660	Character length without spaces	19.967
Complexity Factor Lexical Density (LD, V/N) [<20%]	15%	Average syllables per word	2,08
Lexical Richness (V_1/V) [<50%]	48%	Sentences count (S)	175
Readability (Gunning-Fog Ind.) [6-easy - 20 hard]	13,5	Average words per sentence (WPS)	27,1
General average frequency (FMG, N/V)	6,7	Max sentence length (words) (MW)	87
Guiraud's Index (V/\sqrt{N})	10	Min sentence length (words) (LW)	4

The corpus contains a total number (N) of 4452 words (types), 175 sentences with 660 unique forms (token) (V). The lexical density (LD) is the ratio between the vocabulary width and the corpus size (V/N). It expresses the lexical extension of the text or, better to say, the variety of the vocabulary. The LD index is one of the fundamental parameters for evaluating the corpus suitability since a language that presents a corpus with a wide lexical extension (i.e., rich vocabulary) should have an equally large number of occurrences. Empirically to extrapolate useful information in the textual analysis, the LD acceptability threshold should be less than or equal to 20%. In our case, this is a good value because the basic vocabulary represents 15% of the entire corpus with an average of words per sentence (AWPS) of 27.1. Another important parameter for evaluating the text statistical tractability is the percentage of hapax with the vocabulary size (V_1/V). This index represents a lexical richness measure since it specifies how many graphic forms appear once compared to distinct words. In general, the number of hapaxes should not exceed more than half the extent of the vocabulary; the corpus again respects the indication being V_1 equal to 320 with a percentage ratio of hapax equal to 48%. Therefore, our corpus is sufficiently large for quantitative analysis. Consequently, the lexical size of the text is deemed representative of the language for primary school students. On the contrary, a too rich vocabulary compared to the size of the corpus would not make the text a characteristic sample of the language (Corral et al., 2015).

The vocabulary distribution is not homogeneous because there are many words with a low frequency corresponding to the highest rank positions and few terms persistent with a lower rank. The most frequent different terms constitute the core of the keywords of the corpus. Indeed, identifying the keywords is essential for getting the meaning of the answer content. They can highlight the latent structure underlying the text by reducing the representation space of the linguistic variables. To obtain a list of thematic keywords of the corpus, we grouped the terms into three frequency classes: high frequencies (HF), medium frequencies (MF), and low frequencies (LF). The first six graphic forms in the high-frequency class have a cumulative frequency equal to 42.7%. By adding the classes of medium and low frequencies, we obtain 57.3% of the occurrences. It corresponds to 99% of the vocabulary. Consequently, we considered the union of the two groups MF and LF as a sub-corpus to extract the set of keywords.

Such selection of the most recurring terms should provide a synthetic representation of the peculiar semantic references even though they represent less than 10% of the vocabulary. These terms make up a descriptive structure not as a summary of the content but as key descriptive indices of the analysis' three dimensions. The term "simmetria" [symmetry] is the theme word par excellence (122 occ.) followed by "molto" [a lot] (76 occ.), "imparare" [to learn] (74 occ.), "piacere" [pleasure] (73 occ.), "cosa" [thing] (60 occ.), "difficoltà" [difficulty] (54 occ.), "scatole" [boxes] (51 occ.).

To investigate any specificities related to the central theme of symmetry, we highlight the presence of other critical keywords such as “film” (33 occ.), “palindromo” [palindrome] (32 occ.), “parola” [word] (27 occ.), “figura” [figure] (25 occ.), “mela” [apple] (17 occ.), “stella” [star] (14 occ.), “numero” [number] (13 occ.). These graphic forms constitute essential traces of the activities carried out in the classroom, connected especially to the main task of cardboard boxes. They define the broad spectrum of the “semantic trait” that revolves around the symmetry argument, even better framed through the most common verbs “vedere” [to see] (49 occ.), “fare” [to do] (47 occ.), “potere” [can] (29 occ.), “chiudere” [to close] (25 occ.), “capire” [to understand] (24 occ.), “scoprire” [to discover] (23 occ.), “costruire” [to build] (21 occ.), “leggere” [to read] (15 occ.), “divertire” [to enjoy] (16 occ.), “tagliare” [to cut] (13 occ.).

The techniques of lexical analysis can extend the effectiveness of analysis by finding other textual regularities such as repeated groups of words with a precise semantic meaning. Repeated text segments of three graphic forms (3-grams or trigrams) are very connotative for the content extraction. They are particularly suitable for highlighting syntagmatic connections between the graphic forms because they have more explicit significance than the same words would have individually. We extracted a list of 15 segments from the corpus using the Textalyser application. We identified the most predominant trigrams that offer precise indications on the analysis dimensions (D1, D2, D3).

Table 2. Summary of predominant trigrams for dimensions D1, D2, D3.

(D1) Emotional dim.	IS _r	(D2) Cognitive dim.	IS _r	(D3) Conflict dim.	IS _r
mi è piaciuto [I liked it]	0,92	ho imparato che [I learned that]	0,62	non ho avuto [I didn't have]	0,70
è piaciuto molto [I loved it]	0,36	in quanti modi [in how many ways]	0,48	ho avuto difficoltà [I had difficulty]	0,83
mi sono divertito [I had fun]	0,68	imparato molte cose [learned things]	0,48	po' di difficoltà [little bit of difficulty]	0,31
mi sono piaciute [I liked them]	0,59	abbiamo scoperto che [we found that]	0,50	avuto nessuna difficoltà [had no difficulty]	0,30
piaciuta di più [I liked most]	0,36	ho capito che [I understood that]	0,47	ho avuto nessuna [had not any]	0,26

To further evaluate the students' emotional and affective dimensions, the narratives were afterward analysed using natural language processing techniques called Natural Language Processing “NLP”. The textual statistical analysis used is the Sentiment Analysis “SA” to extract information from texts about associated attitudes (i.e., sentiment). The attitude can be of various types: judgment or evaluation of the topics and the pupils' emotional reaction. The fundamental task of the “SA” is the classification of the overall polarity of the answer (i.e., positive, negative, or neutral value about the topic in question) with API (Application Programming Interface) library functions provided by free software *Meaning Cloud*. Although the algorithmic analysis does not ensure objectiveness, adopting such procedures allows to orient the semi-structured question towards objectivity, reducing interpretation errors. The performance metrics used to evaluate the classification results are accuracy, precision, recall, and F-measure. The metrics are calculated on the values of agreement positive (TP), disagreement positive (FP), agreement negative (TN), and disagreement negative (FN) (Altrabsheh et al., 2014). The levels of processing metrics achieved on the data set are relatively good, showing that the model allows a fairly correct sentiment characterization to the corpus of pupils' answers.

Table 3. Sentiment Analysis report of corpus and polarity classification performance metrics (elab. *Meaning Cloud*).

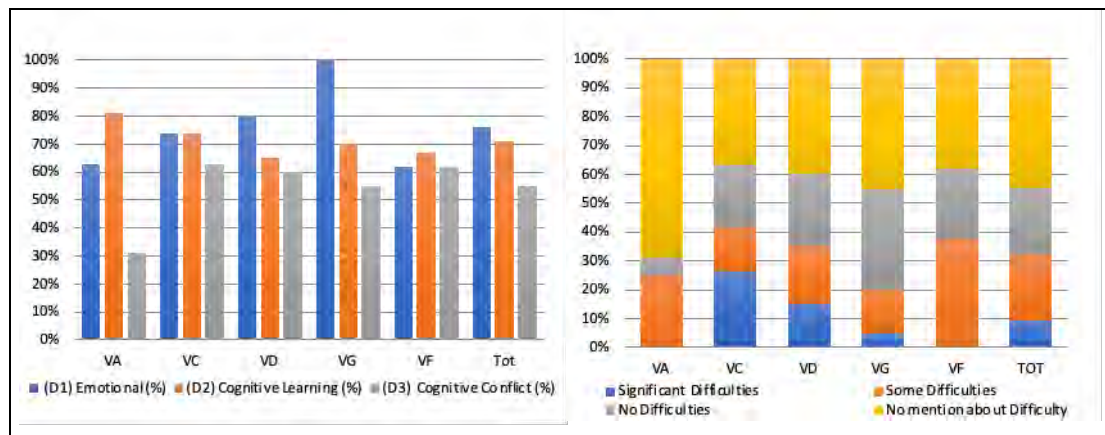
Sentiment	Count	Percentage	Agreement	Disagreement	Performance Metrics	
Positive	112	60%	103 (TP)	9 (FP)	Accuracy	89%
Neutral	15	8%	4	11	Precision	92%
Negative	17	9%	14 (TN)	3 (FN)	Recall	97%
No Sentiment	42	23%	42	0	F-measure	94%
Total	186	100%	163	23	Accuracy	89%

3. Results

As shown in Figure 2, the emotional component D1 is 77% (S.D. 16%). These indices show that the classes had a very positive but not regular emotional and affective attitude. The most positively involved class is the VG (100%), while the least involved are the VA and VF classes (63%). The cognitive dimension of D2 learning is right and relatively constant in the various classes, with an average

percentage index of 71% (S.D. 6%). In the corpus, the dimension of cognitive conflict is 55% (S.D. 14%). The detail of cognitive conflict distribution shows that 10% of pupils had difficulties, 22% some difficulties and 23% no difficulties, while 45% mention no perceived difficulties.

Figure 2. Bar graphs of the diffusion of dimensions (left) and the detail of cognitive conflict distribution (right).



A coherence of judgment is identified between the two estimates obtained with different methods. Indeed, it emerged that the polarity of attitude is positive for 60% (S.D. 5%) (Table 3), comparable with the D1 index shown in Figure 2.

4. Conclusions

The findings suggest that open questions are effective when exploring pupils' answers in the questionnaire. This study reports some results about the validity and potentiality of using Text Analytics and Natural Language Processing (NLP) in the mixed-methods analysis. Additional qualitative information expanding structured analysis responses can provide signs and key insight about students' cognitive path. Their positive feedback to the TLS indicates that the activities linked to the manipulative games motivate and advance the learning process on symmetry and mathematics in general. The findings also bring important suggestions and contents of reflection that teachers can consider for exploiting the potential learning path on symmetry and invariance.

References

- Altrabsheh, N., Cocca, M., & Fallahkhair, S. (2014). Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback. *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, 419-423. doi.org/10.1109/ICTAI.2014.70.
- Brasili, S., Piergallini, R., (2020) Nature of Science Interdisciplinary Teaching based on Symmetry and the Search of Invariants, *International Conference on Education and New Developments*, 394-398.
- Corral, Á., Boleda, G., & Ferrer-i-Cancho, R. (2015). Zipf's Law for Word Frequencies: Word Forms versus Lemmas in Long Texts. *PLoS ONE*, 10(7). doi.org/10.1371/journal.pone.0129031.
- Dreyfus, T., & Eisenberg, T. (1990). Symmetry in mathematics learning. *International Reviews on Mathematical Education*, 2, 53-59.
- Duval, R. (1998). Geometry from a cognitive point of view. In C. Mammana & V. Villani (Eds.), *Perspectives on the Teaching of Geometry for the 21st century*, 37-52. Dordrecht: Kluwer Academic Publishers.
- Gregori-Signes, C., & Clavel-Arroitia, B. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia - Social and Behavioral Sciences*, 198, 546-556.
- Leikin, R., Berman, A. & Zaslavsky, O. (1998). Definition of Symmetry. *Symmetry: Culture and Science: Order and Disorder*, 9 (2-4), 375-382.
- Moerk, E. (1974). Changes in verbal child-mother interactions with increasing language skills of the child. *Journal of Psycholinguistic Research*, 3, 101-116.
- Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of Research in Science Teaching*, 2, 176-186.
- Sfard, A. (2008). *Thinking as Communicating: Human Development, the Growth of Discourses, and Mathematizing*. New York: Cambridge University Press.
- Wiedemann, G. (2016). Text Mining for Qualitative Data Analysis in the Social Sciences. Wiesbaden: Springer VS.