

Searching for Agreement. On the Difficulty in Assessing Artworks – Own Research Report

DOI: 10.15804/tner.2019.55.1.10

Abstract

The main objectives of the long-term research, the partial results of which are presented in this paper, were related, among other things, to identification of the influence of experimental artwork classes on children's artistic skills. The adopted research method based on analysis and interpretation of artworks regarding their compositional modality turned out to be the factor unmasking the difficulty in assessing children's artworks. Therefore, the paper presents fragmentary results of complex research and demonstrates the divergence of marks given by a group of competent judges who used a structured tool.

Keywords: *artwork assessment, visual research, assessment subjectivity*

The role of art and sight in science

Post-modernism is undeniably saturated with sight-centrism, which “refers to the apparent privilege of fusion over all other senses in contemporary Western society (increasingly elsewhere)” (Banks, 2007, p. 14). The era dominated by eyesight was mentioned already in 1993 by Martin Jay, who coined the term *oculocentrism* (adjective: *oculocentric*) to point at the dominant position of the visual sphere (Jay, 1993, p. 3). In the present times, eyesight is becoming the dominant sense and simultaneously takes the form of a cognitive and communication medium which enables people to convey messages, perceive, identify and, consequently, understand

and get to know the world as it really is (Fyfe, Law, 1988, p. 2). This phenomenon is a certain initiation of a tendency for integrating the spheres of art and science (Statford, 1991), which are interpreted as specific forms of experiencing and getting to know the world surrounding the individual. Thus, art aspires to be the companion of scientific activities, becoming a research space and methodology, broadening the boundaries of cognition and weakening the methodological discipline; it is situated in one of the three relational areas distinguished by Ryszard Kluszczyński, which is called *art for science* (Kluszczyński, 2011). Certainly, one should remember that the use of visual materials in research is strongly related to anthropology and sociology and must not be seen from the angle of methodological novelties (Harper, 2002). It would be more appropriate to state that using visual materials in research processes is an expansive phenomenon which gradually creates new possibilities and broadens the areas where visual methodology could be applied. The indisputably broad scope of application of visual methods (Bagnoli, 2009, Frith et al., 2005; Guillemin, 2004) is obscured by the conviction that they are unreliable because the analyzed empirical material of optical nature generates the problem of interpretation ambiguity, which excludes objectivity. Therefore, research based on analysis of visual materials is frequently marginalized and meets with skepticism expressed by scientists (Guillemin, 2004). Every image, visualized idea or materialized, recorded form of artistic undertakings in the optical aspect or any visual aspect of culture may be interpreted many times and in various ways depending on the individual and social context (Banks, 2007).

In view of the above findings, reliability and objectivity of the assessment of children's artworks in scientific research seems very interesting. Therefore, this text aims to present the results of the research and stimulate a discussion about using visual materials in scientific research.

Methodological introduction

In the presented article, research results are only a small part of more extensive empirical material collected from long-term research activities¹. The aim of the research is determination of the impact of experimental activities inspired by contemporary visual art on children's creative potential (the leading area) and their artistic skills (additional, complementary and complementary areas). Due to the number of people in experimental and control groups (each group contains

¹ The research activities are currently being finalized in the form of a doctoral dissertation.

16 respondents), research activities were a quasi-experimental form with pre-tests and post-tests. It is worth highlighting that the pupils' age was between 7 and 9 years, and both groups were similar to each other, which is reflected by the calculated averages and standard deviations. The experimental classes were organized in one of the educational institutions in Zabrze, during the club activities when children had already finished their obligatory lessons and waited for their parents.

As mentioned before, this article applies only to a fragment of the research related to the artistic area. An original tool, called the Artwork Assessment Sheet, was used to identify the potential changes. The tool contains subcategories which refer to analyzing images by using compositional interpretation (Rose, 2001). That special kind of interpretation, called the "good eye" method, has not been made methodologically precise yet, as Gillian Rose remarks. However, it allows for assessing the compositional structure of an image; that structure is related to the content, spatial organization, colors and expressive content (Rose, 2001). In the presented research fragment, a group of three competent judges assessed the works by the children from the experimental and control groups, made during the first and last classes (for the experimental group) and during additional classes (for the control group). They used a four-mark scale (0, 1, 2 and 3), giving their marks in several subcategories:

- practical application of the means of artistic expression (line, texture, color, shape);
- interpretation of the topic and solving an artistic problem;
- the ability to express one's own thoughts and feelings;
- the ability to shape perspective and composition;
- the ability to shape the relationship of the artwork with reality;
- the ability to provide a sufficient amount of details and ensure completeness.

It must be stressed that the competent judges undoubtedly understood the defined and categorized assessment areas presented above because they are closely connected with artwork and visual arts, both in terms of theory and practice. Every person assessing the children's works demonstrated specific competence (Rose, 2001). The judges worked individually, without consulting one another. They assessed forty works² (20 pre-experimental works in the experimental and control groups as well as 20 post-experimental works in the experimental and control groups) made using the same artistic technique, i.e., collage, which allows for experimenting freely with tools, materials and means of artistic expression. The artistic activities were planned in such a way as to take into account the necessity

² The works were officially finished and randomly selected.

and possibility of free interpretation of the topics, hence the suggested class topics and the resulting artworks: *Labyrinth of thoughts* — the experimental work, and *Something out of nothing* — the post-experimental work. The artworks were positioned and ordered in the sheet in such a way as to prevent identification of their authors from the experimental and control groups.

After collecting the filled-in sheets, counting the score and conducting an analysis using the Wilcoxon signed-rank test, which is a non-parametric alternative to Student's *t*-Test, it turned out that no (statistically significant) positive changes had taken place concerning the artistic skills of the children from the experimental group; thus, the previously formulated alternative hypothesis assuming the occurrence of such changes had to be deemed wrong and rejected. It was also confirmed by an individual analysis of the results of pre- and post-experimental works, which revealed that only four participants in the classes demonstrated positive changes. While ordering the results and working on the empirical material, it was additionally decided to estimate the agreement of the marks given by the competent judges using Kendall's *W* coefficient, which is used to measure judge reliability strength (Ferguson, Takane, 1989). This was done because judge non-agreement might have become the factor disturbing the analysis of the obtained material and interpretation subjectivity could have made it impossible to record the approved changes concerning the artistic skills of the children from the experimental group. Analyzing the obtained empirical data in terms of potential divergence of marks was tremendously important because when the three judges were filling in the sheets, they frequently provided feedback concerning the impossibility of formulating a reliable assessment or giving a reliable mark due to the excess of the visual material (which was insufficient to allow for conducting a statistically correct analysis) and due to the necessity of assessing the artworks alone, completely separated from the creation process (which the judges did not watch). The reservations concerning the visual material assessment manner are truly puzzling and raise a number of questions and uncertainties. I hope that the data analysis which follows will initiate considerations about the possibilities and limitations of applying visual methods in pedagogical research.

Searching for divergence

I will begin the analysis of the gathered empirical material with a presentation of the data which directly influenced making the decision to estimate judge reliability strength. As mentioned above, the idea of using non-parametric statistics in the

analysis emerged during ordering the data and making preliminary calculations. Therefore, the material included in the introductory part presents the issue and justifies further calculations rather than explaining the existing state of affairs. The tables below order the average marks given by the three competent judges and present the dispersion value for the individual areas previously distinguished in the tool and subjected to assessment. Table 1 presents the results obtained by pre-experimental artworks, while Table 2 shows those of post-experimental ones, which concluded the cycle of unconventional artwork classes.

Table 1. Average marks given by competent judges to pre-experimental artworks

Categories assessed in pre-experimental artworks	Average marks Judge 1	Average marks Judge 2	Average marks Judge 3	Average marks given by three Judges	Standard deviation
Practical application of the means of artistic expression (line, texture, color, shape)	3	2	2.5	2.5	0.50
Interpretation of the topic and solving an artistic problem	2.6	1.9	2.2	2.23	0.35
The ability to express one's own thoughts and feelings	3	1.05	2.2	2.08	0.98
The ability to shape perspective and composition	2.33	0.48	1.53	1.45	0.93
The ability to shape the relationship of the artwork with reality	3	1.1	1.2	1.77	1.07
The ability to provide a sufficient amount of details and ensure completeness	3	2.5	2.4	2.63	0.32

Table 2. Average marks given by competent judges to post-experimental artworks

Categories assessed in post-experimental artworks	Average marks Judge 1	Average marks Judge 2	Average marks Judge 3	Average marks given by three Judges	Standard deviation
Practical application of the means of artistic expression (line, texture, color, shape)	2.95	1.9	2.3	2.38	0.53

Categories assessed in post-experimental artworks	Average marks Judge 1	Average marks Judge 2	Average marks Judge 3	Average marks given by three Judges	Standard deviation
Interpretation of the topic and solving an artistic problem	2.78	1.68	1.99	2.15	0.57
The ability to express one's own thoughts and feelings	2.95	1	2.2	2.05	0.98
The ability to shape perspective and composition	2.88	0.58	1.08	1.51	1.21
The ability to shape the relationship of the artwork with reality	3	1.05	1.65	1.90	1.00
The ability to provide a sufficient amount of details and ensure completeness	2.88	1.92	2.37	2.39	0.48

The data in the above tables point at a significant dispersion of the mark values around the average values, especially in the area concerning the ability to shape the relationship of the artwork with reality (the deviation of the average for the marks given to pre-experimental artworks: 1.07; for post-experimental artworks: 1.0), the ability to shape perspective and composition (the deviation of the average for the marks given to pre-experimental artworks: 0.93; for post-experimental artworks: 1.21) and the ability to express one's own thoughts and feelings (the deviation of the average for the marks given to pre- and post-experimental artworks: 0.98).

Certainly, the above findings are introductory and approximate. It must be remembered that they cannot be a basis for making conclusions about the judges' agreement or its lack, especially because the above descriptive statistics is based on the most popular measure of central tendency (classical average measure), i.e., arithmetic mean (Bobiński, 2004, pp. 30–33), the value of which does not allow for verifying interpretation convergence. Therefore, further analyses utilized a non-parametric statistical test — Kendall's *W* test, during which the agreement coefficient was calculated separately for either type of artworks (pre- and post-experimental) and either group (control and experimental) as well as for each type of skills. The results of these analyses are presented in Tables 3 and 4.

Table 3. Kendall's *W* coefficient value for the assessment of the pupils' works before starting the experiment

Parameters	PRE-EXPERIMENTAL WORK ASSESSMENT						
	Groups	Practical application of the means of artistic expression (line, texture, color, shape)	Interpretation of the topic and solving an artistic problem	The ability to express one's own thoughts and feelings	The ability to shape perspective and composition	The ability to shape the relationship of the artwork with reality	The ability to provide a sufficient amount of details and ensure completeness
N=3							
Kendall's <i>W</i>	E.	0.317	0.425	0.335	0.445	0.363	0.383
	C.	0.292	0.432	0.298	0.437	0.333	0.401
Chi-squared test	E.	8.571	49.728	9.456	30.011	9.814	21.853
	C.	7.875	50.578	8.048	24.901	9.000	28.112
df	E.	9	39	9	19	9	29
	C.	9	39	9	19	9	29
Asymptotic significance (p)	E.	0.478	0.117	0.447	0.059	0.366	0.292
	C.	0.547	0.101	0.529	0.164	0.472	0.113

C. – the control group, E. – the experimental group

Table 4. Kendall's *W* coefficient value for the assessment of the pupils' works after the experiment

Parameters	POST-EXPERIMENTAL WORK ASSESSMENT						
	Groups	Practical application of the means of artistic expression (line, texture, color, shape)	Interpretation of the topic and solving an artistic problem	The ability to express one's own thoughts and feelings	The ability to shape perspective and composition	The ability to shape the relationship of the artwork with reality	The ability to provide a sufficient amount of details and ensure completeness
N=3							
Kendall's <i>W</i>	E.	0.458	0.528	0.333	0.636	0.324	0.501
	C.	0.364	0.564	0.487	0.665	0.222	0.391
Chi-squared test	E.	12.375	61.820	9.000	36.229	9.000	49.541
	C.	8.378	65.934	13.162	37.889	6.000	34.011
df	E.	9	39	9	19	9	29
	C.	9	39	9	19	9	29
Asymptotic significance (p)	E.	0.193	0.011	0.437	0.010	0.446	0.011
	C.	0.365	0.009	0.115	0.006	0.740	0.239

C. – the control group, E. – the experimental group

The results presented in the above tables clearly point at the divergence of the marks given by the competent judges. Kendall's agreement coefficient value ranges from 0.222 to 0.065 for the distinguished categories, which means that the results in all the areas are lower than 0.7 — a value which could be deemed satisfactory and confirm the judges' agreement in giving the marks. The lowest result, which points at the dissimilarity of the marks in the analysis of the pre-experimental works, was obtained by the areas related to practical application of the means of artistic expression and the ability to express one's own thoughts and feelings — in the control group (0.292 and 0.298). In Table 4, which presents the results concerning the post-experimental works, the lowest result belongs to the ability to shape the relationship of the artwork with reality (0.222 in the control group and 0.342 in the experimental group), while the highest one belongs to the category determining the ability to shape perspective and composition (0.665 in the control group and 0.636 in the experimental group). The latter value seems interesting in view of the average values and deviations presented earlier and the individual manner of assessment. Therefore, the table below orders and displays the partial marks given by the competent judges in this category, observing the split into two separate sub-areas focusing on the interpretation of perspective (*The attempt to apply perspective and show image depth*) and of composition (*The ability to distribute the individual elements of the image on the plane*).

Table 5. Partial marks given by competent judges for perspective and composition (post-experimental works)

The ability to shape perspective and composition								Average for the module
The attempt to apply perspective and show image depth				The ability to distribute the individual elements of the image on the plane				
THE EXPERIMENTAL GROUP								
J1	J2	J3	THE AVERAGE	J1	J2	J3	THE AVERAGE	
3.00	1.00	0.00	1.33	3.00	1.00	2.00	2.00	1.67
2.00	0.00	0.00	0.67	3.00	1.00	2.00	2.00	1.14
3.00	0.00	0.00	1.00	3.00	1.00	3.00	2.33	1.43
3.00	0.00	0.00	1.00	3.00	1.00	3.00	2.33	1.43
3.00	0.00	0.00	1.00	3.00	1.00	2.00	2.00	1.29
3.00	0.00	0.00	1.00	3.00	1.00	3.00	2.33	1.43

The ability to shape perspective and composition								
The attempt to apply perspective and show image depth				The ability to distribute the individual elements of the image on the plane				Average for the module
THE EXPERIMENTAL GROUP								
J1	J2	J3	THE AVERAGE	J1	J2	J3	THE AVERAGE	
3.00	0.00	0.00	1.00	3.00	1.00	2.00	2.00	1.29
3.00	0.00	1.00	1.33	3.00	1.00	3.00	2.33	1.57
3.00	0.00	0.00	1.00	3.00	2.00	2.00	2.33	1.43
3.00	0.00	0.00	1.00	3.00	1.00	1.00	1.67	1.14
THE CONTROL GROUP								
J1	J2	J3	THE AVERAGE	J1	J2	J3	THE AVERAGE	Average for the module
3.00	0.00	0.00	1.00	3.00	1.00	2.00	2.00	1.50
0.00	0.00	0.00	0.00	3.00	1.00	2.00	2.00	1.00
3.00	0.00	0.00	1.00	3.00	1.00	3.00	2.33	1.67
3.00	0.00	0.00	1.00	3.00	1.00	1.00	1.67	1.33
3.00	0.00	0.00	1.00	3.00	1.00	2.00	2.00	1.50
3.00	0.00	0.00	1.00	3.00	1.00	1.00	1.67	1.33
3.00	0.00	0.00	1.00	3.00	2.00	2.00	2.33	1.67
3.00	0.00	0.00	1.00	3.00	1.00	2.00	2.00	1.50
3.00	0.00	0.00	1.00	3.00	1.00	1.00	1.67	1.33
2.00	0.00	1.00	1.00	3.00	1.00	2.00	2.00	1.50

It turns out that Judge 2 and Judge 3 gave very similar marks, especially in the sub-category concerning application of perspective and showing image depth, and their marks suggested a complete absence of those skills (an indisputable dominance of “0” mark) both in the control group and the experimental group, which might justify the high yet unsatisfactory value of Kendall’s *W* coefficient. Judge 1 gave completely different marks (exceeding 3); that person rated the ability to shape perspective and composition very highly, which influenced the average value of all the judges’ marks and the standard deviation for that area. The marks presented in the above table, except the marks given by Judge 2 and Judge 3 in the sub-category related to perspective identification, often seem extreme and this impression complies with the previous findings concerning the impossibility

of confirming the agreement of the marks given by the three-person group of competent judges.

Assessing the assessment: conclusions regarding the use of visual materials in research

The conducted and finalized research activities allowed me to capture the problems related to ambiguous assessment of children's artworks. The initial ordering of the empirical material, the first calculations and the obtained results, which caused an unambiguous rejection of the hypothesis initiated further analysis. The latter aimed to show the agreement degree using Kendall's W agreement coefficient. However, its final value must be deemed fairly unsatisfactory because it ranges from 0.222 to 0.665, thus pointing at the divergence of the judges' marks.

The use of children's artworks in research allowed me to finalize the verification research effectively. However, it simultaneously generated and identified a certain area of ignorance, which revealed the difficulty in analyzing and assessing the compositional modality of images using the "good eye" method. Due to the possibility of working on the visual material and the difficulties in interpreting the marks given to the works as well as the emerging areas requiring deepened exploration, it would be necessary to broaden and intensify the research concerning the research tools, adopting the direction determined by the questions bothering the researcher. Those works may concern, among other things, the following issues:

- development of an appropriate, reliable interpretation tool allowing for analyzing children's artworks in the scope of their compositional modality and simultaneously useful in practical situations (e.g., during debates of artistic competition committees);
- checking if the marks given by competent judges vary when they are enabled to record and observe the creation process;
- controlling the reliability of the developed tool by maximizing the number of assessing persons (with visual competence) and a significant reduction of the number of assessed works, which could be divided into groups and assigned to a specific version of the tool.

The works aimed at optimizing the tools allowing for analysis of images reliably using compositional interpretation and assessing artworks (especially those made by children) will make it possible to use visual materials effectively in research conducted on various cognitive planes. Therefore, the first conclusion refers to the necessity of intensifying the works in the area connecting science with art.

The second one takes the form of a specific declaration postulating the creation of an educational space allowing for conducting activities aimed at shaping and developing interpretation skills concerning visual resources as the latter are nowadays becoming a communication monster. Due to the size of that monster, it is necessary to ensure that the undertaking in question takes the form which can be called the basic course in visual communication (Foss and Kanengieter, 1992).

To conclude the above considerations, it must be stated that intensification of research activities in the area integrating art with science will be desirable and will probably result in broadening the research tool set for the interpretation of visual works (not only in the scope of analyzing the compositional modality of an image, but also the technological and social scope) as well as that taking actions preventing the expansion of visual ignorance in the world ruled by images is absolutely necessary.

References

- Banks, M. (2007). *Using Visual Data in Qualitative Research*. Los Angeles, CA: Sage.
- Bagnoli, A. (2009). Beyond the standard interview: The use of graphic elicitation and arts-based methods. *Qualitative Research*, 9(5), 547–570. doi: 10.1177/1468794109343625.
- Bobiński, Z. (2004). *Wybrane metody statystyki opisowej i wnioskowania statystycznego*. Wałbrzych: Wydawnictwo WWSZiP.
- Ferguson, G.A. & Takane, Y. (1989). *Statistical analysis in psychology and education*. New York: McGraw-Hill.
- Foss, S. & Kanengieter M. (1992). Visual Communication in the Basic Course. *Communication Education*, 41, 312–323.
- Frith H., Riley S., Archer L. & Gleeson K. (2005). Imag(in)ing visual methodologies [Editorial]. *Qualitative Research in Psychology*, 2(3), 187–198.
- Fyfe G., Law J. (1988), Introduction: on the on the invisibility of the visible. In: *Picturing Power: Visual Depiction and Social Relations*, London: Routledge, 1–14.
- Guillemin, M. (2004). Understanding Illness: Using Drawings as a Research Method. *Qual Health Res*, 14(2), 272–289. doi: 10.1177/1049732303260445.
- Harper, D. (2002). Talking about pictures: a case for photo elicitation. *Visual Studies*, 17(1). doi: 10.1080/14725860220137345.
- Jay, M. (1993). *Downcast Eyes. The Denigration of Vision in Twentieth-Century*. London: French Thought.
- Kluszczyński, R., W., (2011). Art@science. About relations between Art and Science. In: *Towards the Third Culture. The Co-Existence of Art, Science and Technology*, Gdańsk: Wydawnictwo: Narodowe Centrum Kultury, 32–44.
- Rose, G. (2001). *Visual Methodologies: An Introduction to the Interpretation of Visual Materials*. London: Sage.
- Stattford, B.M. (1996). *Good Looking: Essays on the Virtue of Images*. London: Mit Press.