

EXPERT AGREEMENT - MYTH OR REALITY?

Andrea Kárpáti

ABSTRACT

**for the XIV. Congress of the International Association of Empirical Aesthetics,
Prague, 2001**

Between 1993-1995, in the course of the Dutch-Hungarian Project for the Modernization of Final Examinations, 500 students in 36 Hungarian secondary grammar schools completed art and design assignments selected from 24 topics offered by the Dutch examination board, Cito. In the course of three months, students worked for 28 hours individually and produced collections of plans, sketches, a research logbook, variations and a final work in two- or three-dimensional form. A 10-item set of criteria was developed by D. Schönau, Cito and the author of this paper according to which internal judges (the art teachers of the students) and external judges (a panel of artists and educational researchers) judged the collections.

60 collections representing 2 two-dimensional tasks were selected for an international jury session where 30 Hungarian and 6 foreign art teachers judged 20 randomly selected works each to validate and optimize the assessment instrument developed for judging portfolios of art work. The two tasks chosen for the validation of the instrument are labeled '*Gaze at infinity*' and '*Wallpaper*'. In the first task, only the title was given for an expressive drawing or painting. The second task contained a description: wallpapers had to be designed for a seaside hotel that represented both the local environment and the function of the rooms: nursery, bedroom and restaurant. Thus, the judgment of a free expression task could be compared with that of a design task. In this report the '*Gaze at infinity*' data will be analyzed in more detail, and the differences between the judgment results of the two tasks will be pointed out.

Four kinds of data are available in this project:

- (i) Ratings by judges or raters with respect to ten different criteria.
- (ii) Global judgments from the same raters who gave the ratings on the criteria.

These global judgments are marks on a ten-point scale (from 1 to 10), and were given after the ratings per criterion were given. They might be viewed as a kind of summary of the detailed ratings per criterion.

(iii) First impressions, given by different raters who did not judge the works using the criteria, also on a ten-point scale, ranging from 1 to 10. In fact, the raters of the '*wallpaper*' project gave their first impressions on the '*Gaze at infinity*' collections and vice versa. (This is why the two projects are not completely independent.)

(iv) All the collections used in the present research are examination works of students and have been judged in the context of the school examinations. The marks given at these examinations at the schools by the respective classroom teachers are also available in the present

project and will be considered in the second report. The school marks consist of three ratings (on a ten-point scale ranging from 1 to 10). The first is the mark given by the own teacher, the second is the mark given by an external judge, and the third (which is the only mark that counts for the examinations) is a compromise between the mark given by the own teacher and that given by the external judge. This final mark results from a discussion between the teacher and the external judge in case of disagreement.

The paper will introduce the list of criteria with illustrations of art work representing solutions of different quality for each criterion, explain the mathematical procedure specially developed to analyze juror bias and criterion validity and show to what extent jurors are able to give an objective judgment of student portfolios.