

Chapter 44

From Subjectivity to Objective Evaluation: A Techno-rationalist Approach of Assessment Design for Art and Design Education

Lin Chou Cheng

Abstract In the field of art and design, it is an accepted norm for the marketplace's art appraisers to use subjective assessment in determining the quality of an artwork. Inherently, such subjective mode of judgmental practice is also somewhat formalized and being brought into the art education. The use of subjective assessment in the context of art is touted to have better validity among the practitioners, as the work of art has a unique emotional context that cannot be assessed in a fixated manner. However, from the perspective of institutional ranking purposes, the use of subjective assessment is posed to have a reliability issue, for it cannot objectively differentiate students' performance from the norms and the goods. In spite of the fact that art is customarily to be appraised subjectively, this paper would like to share that the work of art can be evaluated objectively, through the adoption of criteria-marking scheme that is techno-rational centric.

Keywords Assessment • Assessment of learning • Assessment for learning • Criterion reference • Norm-referenced assessment • Assessment for art and design • Techno-rationalist

44.1 Introduction

The work of creative art is elusive by nature as it is not easily defined [1]. In his article [2], Dallow had reflected that art is like philosophy, and it is difficult to be contained or conceptually explained. Such elusiveness is all because of subjectivity, where the judgment of a creative work lies in the eye of beholder. This sort of tantalized perception had its impact toward the assessment of creative art, as in the context of education, there is no room for subjectivity especially when an assessment is related to public accountability. All assessments for creative art should be

L.C. Cheng (✉)

Game Design & Development, School of Informatics & IT, Temasek Polytechnic,
21 Tampines Avenue 1, 529757 Singapore, Singapore
e-mail: hi@linchoucheng.com

just like any other exams where they can be objectively valid and reliably discreet toward the abilities of a given student population. However, achieving objectivity during an event of assessment in visual art remains to be difficult and challenging, as there is no standard model answer to it [3, 4]. The challenge in the assessment of creative works lies in how to identify and adopt agreeable criteria as the standard measure among different assessors. It is a common struggle among all art educators and assessors to identify such exemplary criteria that can reliably describe the students' attainment at a particular period of time [5].

However, the actual challenge in the assessment design for art and design education is actually not so much about the identification of exemplary criteria. Rather, it is the procedure of a particular mode of assessment that posed the problems. In practice, most visual art modules would have their learning outcomes being examined through a series of deliverable product. These delivered student products can be a mix of different tangible and intangible forms of art, produced and self-curated over a given period of time. These deliverables are commonly referred to as portfolio. A portfolio of art is usually open ended; it is a collection of evidence that demonstrates one's creative performance. There is no definite model answer to measure one's portfolio, as the assessment is usually subjected to the interpretation and evaluation of its team of assessors through the manner of studio-based dialogue [6]. Such discourse mode of assessment procedure tends to lack reliability, as the judgments can vary from different groups of assessors who have different perceptual beliefs, identity, and expectation.

44.2 Research Objective

This study begins with the objective to establish a common evaluation framework that can help to guide a team of assessors to minimize the bias when assessing art and design modules. From the perspective of an educator and a creative practitioner, a reliable assessment is possible if there is a calibration procedure in place for the assessors who are part of the teaching team. By and large, in the teaching of creative arts, the assessors are teachers who are or have been practitioners themselves [7]. As such, this group of assessors would act with their own professional judgment during an assessment [8]. In most situations, the assessors would be the authority that is central to this process as they are being empowered to act on behalf of an educational system [9]. Inevitably, an event of hegemony like "harsh" marking would occur, when one of these professional-practitioner teachers used their own yardstick of "industrial" standard to gauge a young learner's work. Regardless of intention, these aberrant assessors simply acted according to the standard of their domain, where excellence is expected. This has invited the question of whether the judgment of an experienced practitioner is valid and fair, as their basis of measurement can be subjectively broad or too narrow at times [1]. Thus, in order to comprehend how an art educator would appraise one's creative work, Cowdroy and Williams had conducted an interesting study which shockingly revealed that most tutors in creative arts tend to rely on their intuitive

understanding of a creative ability, and their evaluation was emotionally based on the manner of “what we teachers like” about the work during the assessment [10]. The findings in [10] were not much of a surprise in the landscape of art and design education, but rather a formal confirmation about the existence of such issues where the current assessment practices are far from being objectively reliable. Nevertheless, in higher education the opinions of professional assessors are still central to assessment practice for art education as their judgment is regarded as amenable [8].

Thus, there is a need to develop a reliable assessment framework that can support the creative art assessors to regulate their own marking decisions objectively within a reasonable pedagogical context. Hence, in the effort to reduce inconsistency between assessors within a teaching team, this paper is to propose the adoption of criteria-marking scheme that is techno-rational centric, as the basis of measurement for assessing any creative art subjects.

44.3 Research Background

The practice of assessment is one of the complex yet important processes that are central to every academic curriculum. The methods and beliefs about assessment are similar to any teaching and learning strategy, where both practices are highly dependent on institutional procedures and their policies [11]. To date, there are two distinct types of academic assessment: “assessment for learning (AfL)” and “assessment of learning.” Each of these assessment types has its own methodological belief that would influence students’ behaviors in learning. As pointed out by Biggs [12], “assessment determines what and how students learn more than the curriculum does.” For instance, the assessment of learning is about public accountability; it is a macro evaluation system that ranks the academic performance of a given student population in a particular period of time against the national standard [13]. The assessment of learning uses norm referencing, where ranking of the students is top priority rather than the individual attainment of students. As such, the outcome of norm referencing can be visually modeled and represented by a bell-shaped curve, shown in Fig. 44.1. The mechanic of norm referencing uses a standardized test to classify a broad range of students into dependable rank order. Based on the belief of this system, high achievers would always be limited to a handful. Norm referencing works in a way that if a group of distinctive students were to be banded together and put through another set of test, the system can further differentiate this group of distinctive students into a whole new rank of “A, B, C, and D.”

As such, the outcome of a norm referencing type of assessment should always retain a graph of normal distribution, similar to Fig. 44.1. However, if the results had far too many high achievers, the results can be voided for it cannot differentiate between the norms and the goods. In such circumstances, the assessment might need to be redesigned and readministered. However, this seldom happens as the policy makers can decide whether to skew the graph to the right to maintain the pattern of a normal distribution.

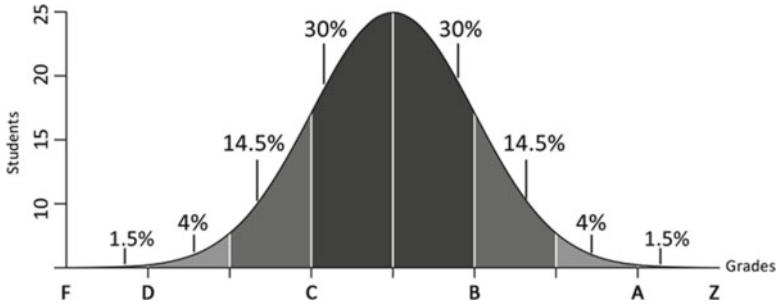


Fig. 44.1 Sample of bell-shaped curve distribution

Unlike the assessment of learning, AfL uses the system of criteria referencing to measure what the learners can do and know, without benchmarking them against their peers [14]. Criteria referencing stresses on the attainment of every individual learner within the predetermined learning outcomes of a curriculum. These learning outcomes are clearly stated criteria for a particular level of practices, and a learner's achievement is to be compared against these standards. As such, there will be a high percentile of learners within a particular cohort to receive either very high or very low grades [15]. Both the methods for assessment of learning and AfL have been used interchangeably, and again, the adoption is subjected to the policy of the respective institutions.

From a holistic viewpoint, criterion referencing has higher degree of validity when assessing art and design modules, as the practices of creative art are all unique performances of individuals. On the other hand, the norm referencing system is better at reliability to gauge a student's performance in the exams; peer benchmarking is much more objective than measuring the creative abilities of an individual, which can be abstract at times. Nevertheless, it is possible to use criterion referencing within the method for assessment of learning. In criterion referencing, students are evaluated individually based on a set of criteria and standards than being graded and compared among their peers. The aim of criterion referencing is to differentiate what students know, understand, and can do, as the result is a formative feedback for teachers and students on their future teaching and learning needs [16]. In this system, students are to compare against descriptions of expected standard across a range of criteria without the need of referencing against the performance of others [17]. However, to maintain reliability among different assessors who use the same exact marking criteria is challenging. The reliability issue in a criteria-based assessment is largely hinged on the involvement of human interpretation of a given criterion. According to Green, a "true" criteria-based marking scheme can be relatively narrow, as highly reliable criteria would not allow other range of interpretations by its user [17]. These micro criteria were meant to guide and restrain the assessors from any bias interpretation. Nevertheless, such criteria-based marking scheme is too time-consuming to prepare and be followed [16].

As discussed, the assessment issue in creative art is duly caused by the accepted norm of either having a loosely crafted criteria-marking scheme or no use of criteria

at all. However, all these can be improved if the assessors were to adopt a rigorous criteria-based marking scheme as their basis of measure. By holding on to this hypothetical view, no work of art is too elusive to be evaluated objectively. Hence, in an effort to validly assess every student's capability with criteria referencing, while achieving a reliable normal distributed outcome that fits the institutional ranking purposes, this research had proposed and implemented the use of criteria-based marking scheme that is techno-rational centric. The proposed assessment practice has been implemented on the subject – *Game Modeling* (GAM). GAM is an introductory technical art subject offered to polytechnic students who studied in the course of game design and development. The subject has been designed and taught in a studio-based approach, with the aim to equip students with technical 3D modeling skills through the use of a 3D graphic application. Through the subject, students will be able to create a series of three-dimensional (3D) digital model for real-time use.

To craft the techno-rational marking scheme for this study, the performance descriptors of the rubric were mainly a detailed extension of the general learning outcomes of the GAM syllabus, which was later being fused with quantitative requirement. In the context of education, techno-rational is about the systematical use of quantitative measure to ensure accountability, and as such, the approach of techno-rational has strong emphasis on validity, objectivity, processes, and procedures [8, 18]. Besides the extrinsic motivation of being a conscientious educator who is obliged to public accountability, the intrinsic motivation of this study and its implementation was largely a personal reflection and a statement of belief in AfL. Prior to the commencement of the subject, GAM was assigned to a team of instructors who were veteran creative practitioners being teamed up for the first time. The issue of reliability was foreseeable among the instructors if the teaching team were to rely on a standard criteria-based marking scheme and the use of a studio-based dialogue. It would take quite a while for the teaching team to calibrate itself. Hence, instead of risking and bracing for unreliable outcome with abnormal distribution of grades, a formal method had to be put in place to direct this group of individual instructors who were going to be the assessors, to be objective during the assessment. From the literature [16, 17], a criteria-based marking scheme can be easily subjected to various interpretations. By recognizing this issue, the techno-rational approach was proposed; all the written assessment criteria for GAM are injected with quantifiable requirements, and any flimsy qualitative statements that might be dubious to interpret are avoided. A study of effectiveness has then been conducted on this proposed assessment practice, and the findings were reported in the following sections.

44.4 Research Method

As aforementioned, this study is intended to find out the reliability of the proposed criteria-marking scheme, which is based on techno-rationalism. The hypothesis of this research is that all biases in the assessment of art can be reduced if all the art assessors were to adopt the same quantifiable marking criteria. The assessment

- Construct basic 3D model by utilizing appropriate Digital Content Creation (DCC) tool;
- Demonstrate the application of polygon modelling techniques to create 3D objects;
- Apply appropriate low-poly modelling techniques to create 3D game object;

Fig. 44.2 General learning outcomes for GAM

would then be able to achieve better inter-rater and intra-rater reliability when there is less involvement of human judgments and dubious interpretations. Quantitative methods were used to conduct the study. The study began with part 1 of the first GAM's assignment: modeling a 3D game level. The assignment is set to measure students' capabilities in crafting 3D objects with a set of prescribed requirements, which is stated in the project brief. The assessment was designed to coach students incrementally to attain three general learning outcomes stated in the GAM's syllabus (Fig. 44.2).

The study involved second year polytechnic students in the Game Design and Development course at a middle-sized polytechnic. The course has good transparency, as the subject's information, such as teaching plan, marking schemes, and teaching materials, was made available in advance for all students to access. Seventy-eight students took the subject (GAM), and this research had sampled 49 students who were randomly selected from three different tutorial groups (P1, P2, and P3). The age of the participants ranged from 17 to 21 years old. The study began by first distributing the proposed criteria-marking scheme to these 41 participants and the researcher. The participants were asked to assess their own work individually based on the given criteria-marking scheme in 30 min. After that their self-assessed result with the given mark sheet was submitted to their tutors. The researcher will evaluate all the participants' works. All the assessments conducted by the researcher were made independently and in reference to the same marking scheme. Toward the end of the process, two sets of test scores were provided by the two parties: the student participants and the researcher.

To analyze the collected data, the research used Microsoft Excel and applied "Pearson correlation coefficient" analyses. As stated, the aim of this research is to determine the inter and intra-reliability of a criteria-marking scheme that centered on techno-rationalism. The study stood on the hypothesis that the higher the correlation (agreement) between the students and researcher during an assessment, the more reliable of the suggested criteria-marking scheme. To interpret the level of correlation coefficient, the research followed the guidelines given by Cohen [19] and Hopkins [20]. There are about seven ranges of correlation coefficient values, as shown in Table 44.1.

Table 44.1 Guideline of correlation coefficient value

Correlation coefficient descriptor	Value (positive)
Trivial	0.0–0.09
Low	0.1–0.29
Medium	0.3–0.49
High	0.5–0.69
Very high	0.7–0.89
Neatly perfect	0.9–0.99
Perfect	1.0

Table 44.2 Assessment correlations between students and researcher for tutorial groups P1, P2, and P3

Assessors	Students	Researcher
Students	1.000	
Researcher	0.651	1.000

44.5 Findings and Discussions

Based on the initial analyses, the result in Table 44.2 had shown that the scoring between the students and the researcher was at 0.651. By consulting the guideline in Table 44.1, the finding was to be interpreted as high. This preliminary study had somewhat positively supported the hypothesis of this research, where a techno-rationalist type of criteria-marking scheme has better inter-rater and intra-rater reliability. According to Sabol [5], assessment that has high correlation among two different groups of assessors, such as the students and the researcher, which in this context was possibly the result of pre-imposition of criteria into the learning and teaching strategies before the assessment. In fact, it is a positive sign that teachers are being serious in developing their students with a specific range of knowledge and skills [4]. Although this study had shown a significant level of reliability with the correlation analyses, there are gaps that need to be closed so as to ensure absolute reliability.

This is because according to the research by Pitts, Coles, and Thomas [21], the average level of reliability between the individual assessors is usually only moderate, but the preliminary result of this research had suggested otherwise. Thus, this researcher has decided to conduct a second correlational analysis, by extending the initial study with additional set of inputs from another independent assessor who is the instructor for tutorial group of P2. The new extended study had paired the instructor of P2 to evaluate the artworks from the tutorial group of P3 with the same exact criteria-marking scheme. With an additional input from this new assessor, the research now has three sets of correlative data to compare and analyze.

In this extended study, the sample size constituted of eleven students from tutorial group of P3, the researcher, and the new assessor himself. The first analysis showed that the correlations between the students and the researcher were at 0.546 and were regarded as high (See Table 44.3). However, the newly obtained value is

Table 44.3 Assessment correlations between students, researcher, and assessor for tutorial group P3

Assessors	Students	Researcher	Assessor
Students	1.000		
Researcher	0.546	1.000	
Assessor	0.451	0.930	1.000

significantly lower than the previous study which was at 0.651 with the sample size of forty-nine. The researcher noted that some students in P3 might be a little bit overconfident about their work and they had rated themselves much higher than they should. Interestingly, the correlations between the new assessor and the students were at medium value of 0.451. This suggested that there were more disagreements between the students and the new assessor within the exact criteria-marking scheme. Based on observation, the new assessor came in as an artist-educator and was different from the researcher who is the archetype of an educator-artist. Hence, the new assessor would be more critical in his judgment during the evaluation. Nevertheless, the correlations score between both of the researcher and assessor was neatly pitched at 0.930.

This significant correlations finding between the researcher and assessor has implied that both markers who have the background of professional practice tend to be highly reliable in their judgment when a techno-rational centric criteria-marking scheme is being employed.

44.6 Conclusions

In conclusion, the assessment system of criteria referencing can be highly reliable and can function like the norm-referenced assessment in differentiating and ranking students for public accountability. A distribution of bell-shaped curve would naturally occur without any skewness if a system of assessment is highly valid and reliable. For instance, the final subject statistic for GAM was able to closely achieve normal distribution while objectively assessing the attainment of every individual student without peer benchmarking. Furthermore, the implemented approach was regarded as more holistic and liberal, as it allows more room for deserving students to receive the commendable grade (Fig. 44.3).

With such an assessment system, it would cultivate a different competitive culture, by comparing to oneself than with others. Nevertheless, all this can only happen when a techno-rational centric criteria-marking scheme is in place. The work of creative art can indeed be objectively evaluated. Through this study, the implemented research method and the use of correlational analyses can actually be adapted into other techno-rational approach for the purpose of testing and calibrating any newly crafted marking rubrics.

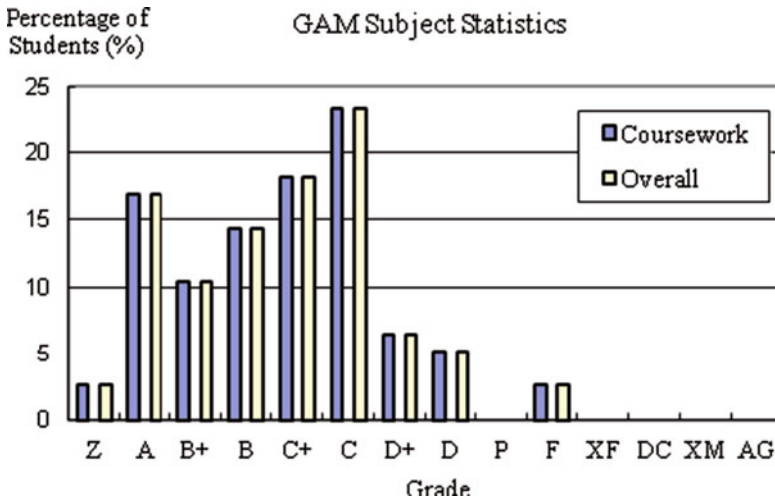


Fig. 44.3 Final subject statistic for GAM

References

- Harris, J. (2008). Developing a language for assessing creativity: A taxonomy to support student learning and assessment. *Investigations in University Teaching and Learning*, 5(1), 80–86.
- Dallow, P. (2003). Representing creativeness: Practice-based approaches to research in creative arts. *Art, Design & Communication in Higher Education*, 2(1&2), 49–66.
- Orr, S. (2006). Assessment practices in art and design. *Art, Design & Communication in Higher Education*, 5(2), 79–81.
- Dikici, A. (2009). An application of digital portfolio with the peer, self & instructor assessments in art education. *Eurasian Journal of Educational Research*, 36, 91–108.
- Sabol, F. R. (2006). Identifying exemplary criteria to evaluate studio products in art education. *Art Education*, 59(6), 6–11.
- Orr, S. (2007). Assessment moderation: Constructing the mark and constructing the student. *Assessment and Evaluation in Higher Education*, 32(6), 645–56.
- Gordon, J. (2004). The 'wow' factor: The assessment of practical media and creative arts subject. *Arts Design & Communication in Higher Education*, 3(1), 62.
- Orr, S. (2011). 'Being an artist you kind of, I mean, you get used to excellence': Identity, values and fine art assessment practices. *Art & Design Education*, 30(1), 37–44.
- Leach, L., Neutze, G., & Zepke, N. (2001). Assessment and empowerment: Some critical questions. *Assessment & Evaluation in Higher Education*, 26(4), 293–305.
- Cowdroy, R., & William, A. (2006). Assessing creativity in the creative arts. *Art, Design & Communication in Higher Education*, 5(2), 97–116.
- Biggs, J. (1999). *Teaching for quality learning at university*. Buckingham/Philadelphia: Society for Research into Higher Education and Open University Press.
- Biggs, J. (2003). *Teaching for quality learning at university* (2nd ed.). Buckingham: SRHE and Open University Press.

13. CEA@Islington. (2003). Quality statement on assessment practice (primary). Retrieved from CEA@Islington: http://www.european-agency.org/agency-projects/assessment-resource-guide/documents/2008/10/CEA_Islington.ok.pdf/view
14. Anastasi, A. (1998). *Psychological testing*. New York: MacMillan Publishing Company.
15. James, R., McInnis, C., & Devlin, M. (2002). *Assessing learning in Australian universities*. Melbourne: The University of Melbourne.
16. Green, S. (2002). Criterion referenced assessment as a guide to learning – The importance of progression and reliability. *Study of Evaluation in Education in Southern Africa International Conference*. Johannesburg: ASEESA.
17. Brown, S. (1988). Criterion referenced assessment: What role for research? In H. Black, & W. Dockerell (Eds.) (pp. 1–14), *British Journal of Educational Psychology*, Monograph series no. 3. UK: British Psychological Society.
18. Kappler, K. (2004). NCATE: Wolf in shepherd's clothes. In T. Poetter, T. Goodney, & J. Bird (Eds.), *Critical perspectives on the curriculum of teacher education* (pp. 19–40). Lanham: University Press of America.
19. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates, Inc.
20. Hopkins, W. G. (2002). A scale of magnitudes for effect statistics. Retrieved from A new view of statistics: <http://www.sportsci.org/resource/stats/effectmag.html>
21. Pitts, J., Cole, C., & Thomas, P. (1999). Educational portfolios in the assessment of general practice trainers: Reliability of assessors. *Medical Education*, 33(7), 515–520.