INTERNATIONAL CONFERENCE ON ENGINEERING AND PRODUCT DESIGN EDUCATION
8 & 9 SEPTEMBER 2011, CITY UNIVERSITY, LONDON, UK

DEVELOPMENT OF THE SYSTEMATIC GRADING PROCEDURE

Anders BERGLUND and Phillip TRETTEN
Division of Industrial Design, Luleå University of Technology

ABSTRACT
A challenge in design education is the grading of students work when they show applied knowledge. Due to this difficulty the grading criteria needs to be and is most often subjectively focused. As previous research has shown the assessment results can vary between teachers, thus, increasing the chance for varied and possibly inconsistent grades. Much has to do with the level of experience and knowledge a particular teacher has and specific individual preferences. The Systematic Grading Procedure (SGP) is a method, which breaks down an assignment and assists the teacher in grading a task based upon the assessment areas. The aim of this paper is to validate the SGP as a grading method for teachers in 3D modeling and 3D visualization and further develop the SGP as an assessment tool for lesser-experienced teachers. The SGP was previously shown to be a helpful tool for giving and receiving feedback. Previous work did not show significant differences between SGP and the commonly used methods but this more extensive study did show that the SGP did significantly help reduce the variation in grading for both experienced 3D teachers and inexperienced 3D teachers but not for non-teachers. There seems to be a good potential for the SGP method to help teachers give more consistent grades and at the same time assist students in the learning process through more specific feedback.

Keywords: Assessing 3D-art, subjective values, systematic grading procedure

1 INTRODUCTION

Grading students' work can be a challenge, especially when the evaluation criteria for the task are largely based on subjective assessments. The optimal situation would be when the evaluation criteria can be specified with a defined value based upon objective criteria. Although, creative work, such as 3D generated images and movies, requires that individual teachers are knowledgeable in the 3D methods since specific answers are not possible. You could say that there are an infinite amount of correct answers because the assessment criteria are often of subjective nature based upon the whole product. For example, the subjective or qualitative criteria could be that the composition in the image is to give a harmonic impression or that lighting and shadows should give a realistic feeling. Lawson [1] states that it is easier to define and enforce quantities than to legislate qualities in the design process. In the case of grading quantitative criteria, usually a pass or fail grade is given while qualitative criteria must be graded on an undefined scale based upon the teacher’s analysis and overall impression of the work. Meadows and Billington [2] say that no assessment has any validity if the grade a student gets varies significantly from time to time, or, on the other hand, the grade is dependent on a specific individual who assesses the work. Since the qualitative criteria are of a subjective nature the grades can vary between examiners.

The Swedish grading system used in this test contains the levels U (fail), 3 (average), 4 (above average) and 5 (excellent). It is a rather imprecise grading scale and sometimes a student's grade can balance between two numbers making it difficult for a teacher to choose the most appropriate grade. It is also possible that two teachers have different opinions on a students work. The difference in such cases is usually only one grade level, i.e. between a 3 and 4 or a 4 and 5. Leung [3] used the *Biggs SOLO taxonomy* to assess design and technology students and he argued that the design problems students face are complex and that there are no fail-safe methods for assessing the students. In an attempt to systematize the grading process in our 3D courses we have developed a model in which the student’s task is assessed in four important assessment areas, *modeling, lighting, materials,* and *composition* (Table 1). The SGP method attempts to breakdown the grading procedure into smaller and more specific parts. Each assessment area contains specific grading criteria. For example, the criteria for the assessment area *material* is lights and shadows and the criteria can be given different weights depending on the purpose of the task, e.g. *modeling* can be 50%, *lighting* for 25% and *materials* 25% of the assignment grade. The overall goal is to use more objective specifics in grading thus giving the teachers more specific objectives without removing from them the overall subjective judgment of the 3D object.

Table 1. The SGP scheme used in the test with example grades in the grey cells. Each image had its own scheme

Modeling		**What is assessed?**
Modeling (Grade 3-5)	4,5	The complexity and details of the model
Assessment area grade	**4,5**	

Lighting		
Light (Grade 3-5)	4	The realism and how well the lighting harmonizes with the objects and the environment in the image
Shadows (Grade 3-5)	4,5	The realism, softness and quality of the shadows
Assessment area grade	**4,25**	

Material		
Material feeling (Grade 3-5)	4	The sense of the materials, e.g. reflections, refractions, highlight size
Texturing (Grade 3-5)	3	How the texture is used, placement, resolution and quality
Assessment area grade	**3,5**	

Composition		
The image composition (Grade 3-5)	5	The composition of objects in the scene and camera angle
Assessment area grade	**5**	

The image grade	**4,313**

The image grade rounded	**4**

After the student's image is assessed based upon the criteria, the grade from each criterion is then weighed together to produce a final assignment grade. Finally, the grades of all assignments are then weighted together to the final course grade, which is rounded off to a U, 3, 4 or 5. An example of the SGP flow is found in Figure 1. The greatest benefit of the SGP, according to Berglund and Tretten [4], is that it is a good tool for giving students feedback on their work. This is true because the teachers are given specific guidelines to follow when meeting with the students thus the validity of ratings increase, as well as, providing a good overview of how the examiner grades. However the grading method of the SGP had not been thoroughly tested and only the 3D teachers interaction with students was previously studied. According to Elisabeth Ahlstrand [5] students should know in advance what is expected and how the student work will be assessed to ensure high reliability for subjective assessments, so that, both the students and the teachers have the same frame of reference to what level the work was assessed. Welldesigned and well-planned assessments are, according to James et al. [6], a strategic tool to clarify what learning is rewarded and also guides students in an effective approach to study. To facilitate the grading process for new teachers Ponn et al. [7] says that a standardized method for the systematic grading is to be preferred. The approach chosen for this paper was that three groups, 3D-teachers, experienced teachers, and non-teachers, would be given the SGP to use when grading students work to see if it does significantly help reduce variation in grading. The aim of this paper is to validate the SGP as a grading method for teachers in 3D modeling and 3D visualization and further develop the SGP as an assessment tool for lesser-experienced teachers.

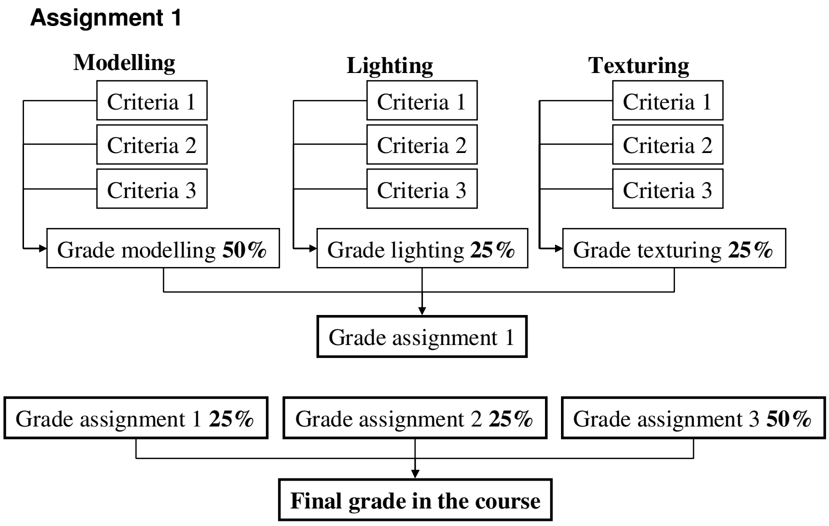

Figure 1. Example of the flow in the SGP

2 METHOD

The test was conducted as an experiment with 3x2 within-subjects design [8] using 15 participants split into three groups. The groups contained 5 participants each, experienced university teachers with 3D modeling and 3D visualization background (Group A), experienced university teachers with no 3D modeling and 3D visualization background (Group B), and individuals with no university teaching experience and no 3D modeling and 3D visualization background (Group C). Since it was difficult to recruit suitable 3D art teachers only five participants were chosen for each group. In order to detect any difference in the grading with or without the use of SGP, the test was divided in two parts, a baseline and experiment part. In the baseline part each person graded the same 10 images without using the SGP, and the second part, the experiment; each person graded 10 new but similar images using the SGP. Each participant was given the assessment criteria along with the instructions in written form before each part of the experiment. The criteria was experienced by 3D art teachers and only images from previously approved 3D course work were used. The test participants were asked to grade each criterion with increments of 0.5, from 3.0 to 5.0. An example of a 3D image used in the test is shown below in Figure 2.

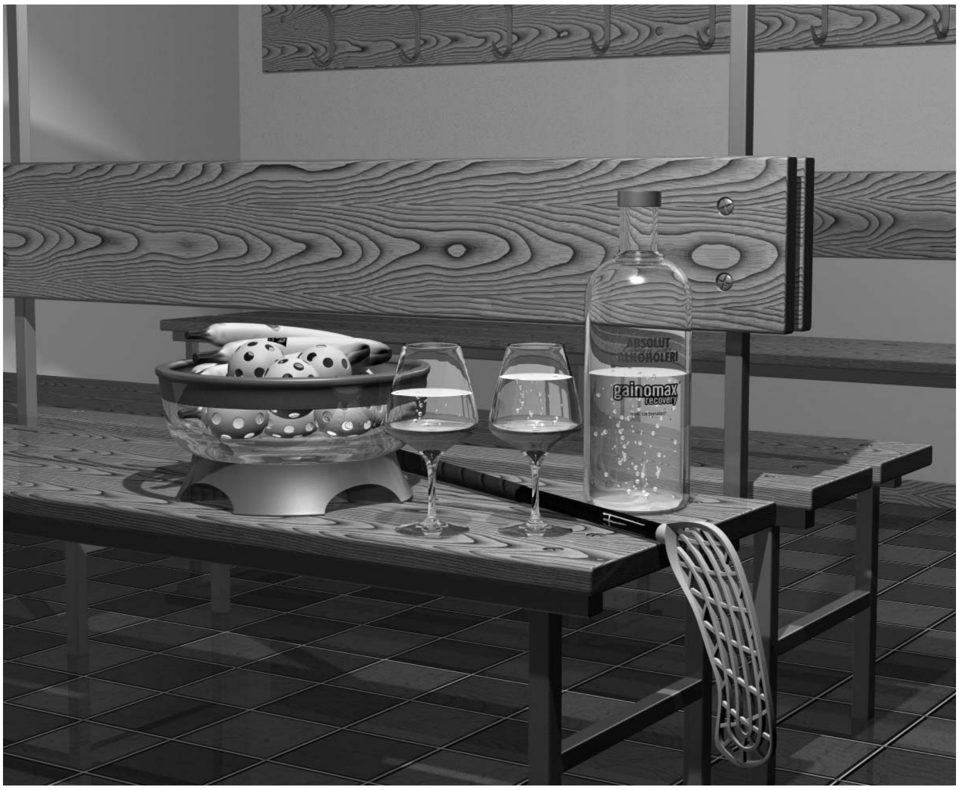

Figure 2. Still life modeled and visualized by Daniel Oliv (student work)

The results were analyzed in SPSS and the grades were placed into six matrices, one for each Group and baseline/experimental. An example is given in Table 2 below showing how the grades differed. The average is shown along with the difference, shown as greatest difference. The black area shows that the difference between the highest and lowest grade is 2, in the grey area, the difference is 1 and in the white area have all subjects given the same grade. In Table 3 the overall results are shown with the average reported grade for each image and examiner, along with the Standard Deviation (*SD*). The groups with the lower *SD* show that they graded the images more consistently, but not necessarily more correctly. The Table 1 shows that the average score differed between the methods. In the experiment only images meeting the basic requirements were used.

Table 2. Example of the matrix for group A for the baseline block of the experiment

	A1	A2	A3	A4	A5	Average	Greatest Difference
Image 1	4	4	5	5	3	4,2	2
Image 2	3	3	3	3	3	3,0	0
Image 3	3	3	4	3	3	3,2	1
Image 4	5	3	3	4	3	3,6	2
Image 5	5	5	5	5	5	5,0	0
Image 6	3	3	3	3	4	3,2	1
Image 7	4	3	3	4	3	3,4	1
Image 8	5	3	4	4	3	3,8	2
Image 9	4	3	4	4	4	3,8	1
Image 10	5	4	4	4	4	4,2	1
Average	**4,10**	**3,40**	**3,80**	**3,90**	**3,50**	**3,74**	**1,1**

3 RESULTS & DISCUSSION

The results (Table 3) show that there were significant differences between the methods used. For both methods, all three groups, and the interaction between methods and groups, only the method showed a significant difference ($F(1) = 12.973$, $p = .001$). The SGP method helped to produce more consistent results for Group A ($F(1) = 12.174$, $p = .025$) and Group B ($F(1) = 36.737$, $p = .004$), but not for Group C. This shows that the method was beneficial in helping teachers grade the student's 3D work more consistently but it did not help non-teachers grade more consistently. It is quite possible that this method is most applicable for those with teaching experience. As a further development of the SGP more explicit criteria may be given to the teachers and non-teachers to find out if the teaching experience was the factor that helps teachers grade similarly or if it was the type of instructions, which were written by teachers, and thus teachers possibly understood them better than a non-experienced reader (non-teachers). This can be supported by the fact that the Groups without the SGP (baseline) did not significantly differ from each other. The SGP method could to be further developed and adapted for non-experienced teachers. Although the Swedish university system requires that certified teachers grade students work and non-certified individuals may assist in other capacities. This common denominator of teaching experience did result in significantly more consistent responses.

Table 3. Average grades and standard deviation of grades for each group

	Non-SGP		**SGP**	
	M	***SD***	***M***	***SD***
Group A	3.74	.762	3,92	.578
Group B	3,70	.807	4,00	.549
Group C	3,64	.745	3,82	.659
M	3.69	.771	3.91	.595

For future work some 3D images that do not fulfill the requirements should be used so that the grade *not passed* also would be included in the participants grading scheme. This would be a further extension of testing the method, although, images that are not considered to pass are not usually graded since the students are required to make the necessary improvements before a final grade is to be given. The method of using the four specific assessment areas seemed to be sufficient for 3D images when assessing the images. A future study would be to test the SGP on a larger scale, testing different types of 3D art courses and different university programs with 3D art programs. It may be possible that the assessment areas need to be refined even more with additional or refined subcategories.

4 CONCLUSIONS

It was expected that the experienced test persons would not assess the images the same way but the SGP helped teachers grade more consistently. Although teaching experience showed to help reduce grading variation the method could be further developed to help 3D teachers and even other teachers grade student work even more consistently. This could help reduce confusion and fulfill many students' needs by giving a more through explanation of what they need to improve in their work.

REFERENCES

[1] Lawson B. *How Designers Think: The design process demystified*, 4th ed., 2005 (Elsevier Linacre House, Jordan Hill, Oxford).

[2] Meadows M. and Billington L. A review of the literature on marking reliability. National Assessment Agency, 2005, p.13.

[3] Leung C. F. Assessment for Learning: Using Solo Taxonomy to Measure Design Performance of Design & Technology Students, *International Journal of Technology and Design Education, 2*, 2000, p.153.

[4] Berglund A. and Tretten P. (2010). Systematic grading procedure based on subjective values. *International Conference on Engineering and Product Design Education,* Trondheim, Norway, September 2010, pp.96-101.

[5] Ahlstrand E. Examination and assessment / teacher training. *Quality Conference*, October 2007, Umeå University.

[6] James R. McInnis C. and Devlin M., *Assessing Learning in Australian Universities, Centre for the Study of Higher Education*, Centre for the Study of Higher Education, The University of Melbourne, Victoria, Australia, 2002, Retrieved from www.cshe.unimelb.edu.au/assessinglearning

[7] Ponn J. Kreimeyer M. and Lindemann U. Methodical Evaluation of Single and Group Projects, *International Conference on Engineering and Product Design Education*, September 2007, Northumbria University, Newcastle upon Tyne, United Kingdom, p.190.

[8] McQueen R. A. and Knussen C. *Introduction to Research Methods and Statistics in Psychology*, 2006, (Pearson Education Limited, Essex, England).