An Assessment Tool for Student Design Projects

Filippo A. Salustri, PhD, PEng
Ryerson University
salustri@ryerson.ca

Abstract

This paper presents a method developed by the author to assess student design projects. The method involves identifying assessment elements, weighting the elements, and grading the work as if had only one author. Data supplied by the students is used to identify the relative work put into each element. This allows individual grades to be calculated in a fair and consistent means. Sufficient background, context, and detail is given to allow anyone to use the reported method. In over eight years of use, no significant problems have been detected with this method.

1. Introduction

This paper describes a systematic way developed by the author to assess the final written reports of student design projects. The technique has been developed by the author over several years, during which time he has taught at four different universities. The author has found that the technique (a) lessens the amount of time spent administering courses – especially large ones, (b) works equally well for freshman and senior design courses, and (c) helps ensure consistency and fairness of *individual* student assessment.

2. Context and background

The author has taught design courses in engineering since 1989, including large first-year classes intended to introduce design processes, as well as senior undergraduate capstone courses and graduate courses. The undergraduate courses have involved up to 200 students at once; senior courses between 50 and 100 students; and graduate courses with as few as eight students. In all these courses, students were expected to work in teams on semester-long design projects culminating in written and oral presentations. The courses have always stopped short of prototyping or mocking up student designs due to limitations of time and resources.

In all cases, students are introduced via lectures and presentations to a generic design process covering product strategy, problem analysis and specification, ideation and concept design, and systems design. Short design exercises are used to let students practice the methods presented in lecture, in prelude to their application to the major project. Every student keeps a design journal, which must be submitted for grading at the end of the course. Students are expected to execute CAD models of their designs for inclusion in their reports. The differences between the courses have been largely a matter of the degree of detail, analysis, and optimisation expected of the students – one would expect seniors to execute better designs more properly grounded in engineering knowledge than juniors.

Projects typically focus on product design and include aspects of user-centred design, manufacturability and cost, sustainability and environmental concerns, as well as functionality and usability. The project descriptions are deliberately vague and open-ended. One typical project might be given as follows:

Ladder for Home & Light Construction Purposes. *Users of ladders often suffer muscle strain and risk falling from the ladder because of the positions they must assume to work on the ladder. Design a ladder for use by handymen and construction workers working on small projects (e.g. home renovations). The ladder should be safe for all users, regardless of age or training, and should be appropriate for typical mechanical and electrical jobs. The ladder must minimise risk of occupational injury as well as accidents.*

Students are assigned to teams based on results of an MBTI-like Personality Type Indicator (PTI) implemented by the author. Using the results of the PTI and rules for group formation taken from the literature [1], teams of between four and six students are created automatically by the software. Fine-tuning of the teams is done manually. (The author notes that the PTI is currently being developed by Ryerson's Digital Media Projects Group as a plug-in module for the Blackboard course delivery system; completion of this project is expected within a year.)

Students conduct several peer reviews of their teammates during the semester. The database

constructed by the PTI software is used to present each student with the opportunity to rank only his/her teammates. Students rank their teammates against ten characteristics: attendance, contributions to discussions, level of communication skills, commitment to team goals, ability to listen carefully, responsible, accepts criticism gracefully, performs significant work, level to technical understanding, and meets deadlines.

Some teams assess themselves quite harshly; other teams are more liberal. Surprisingly, the author has not yet been confronted with a team only one or two team members of which are harsh assessors. It is not clear why this is so. In any event, it is not fair to the teams who assess themselves harshly to simply let the peer review marks stand *raw*. Instead, all peer review raw scores are adjusted with respect to the overall class average on a given peer review. The adjusted scores are then scaled to an interval of {-2, +2}. Thus a student who does poorly on a peer review can lose up to 2% of his/her final mark and a student who does well can get up to 2% more.

Teams also present the status of their projects at two or three design reviews during a semester. Each design review is typically worth as much as a peer review. For junior classes, a design review involves the instructor (a) reviewing the design journal, (b) interviewing the team regarding their progress, and (c) answering any pertinent questions. A junior review typically takes 15 minutes per team.

In senior courses, milestone reports with pre-defined goals are prepared by each team. In each milestone report is a *workload distribution form* (WDF) wherein each team member rates the amount of work done by the member with respect to the one's teammates. Each WDF is specific to its milestone and identifies key tasks and deliverables for the milestone's period. Students must sign the WDF to indicate their agreement with the assessment of their contributions. The instructor explains to the students that they are not necessarily expected to do *a lot* of work on *all* topics listed in the WDF. The scale used by the students in this rating is as follows.

0: did little or no work done on the indicated milestone component.

1: did an average amount of work done on the indicated component.

2: assumed a leadership role on the indicated component.

The milestone report is assessed by the instructor with respect to each task and deliverable for the milestone, as if one person had written it. The instructor does this without looking at the WDF, thus helping to ensure that the report is marked on its own merits and is not biased by what the instructor might think about the team members' performance as indicated in the WDF. These scores are put into a spreadsheet, along with all the data from the WDF. The spreadsheet uses the WDF work contributions to establish a relative measure of the contribution of a team member to each task and deliverable. Marks are assigned to team members based on the relative amount of work done by the team member on each task and deliverable. These marks are then totalled for each team member.

The use of relative work done, via the WDF, has the effect of developing a different score for each team member. For example, someone who did most of the work on a task that received a high score from the instructor should and does benefit in this scheme. Similarly, a student doing most of the work on a task that was poorly executed will get a lower grade. Team members who do more work get proportionally higher grades; students who do less work get proportionally lower grades. Furthermore, since relative work measures are used, it is not possible for students to inflate their marks artificially by always assessing themselves as a group at the highest level.

Also of note here is that milestone reports can be graded quite quickly – more quickly than if the instructor had to differentiate individual team members' performance himself. This allows the milestones to be returned promptly to the students to give them timely feedback on their performance.

The author also notes that students tend to endorse this grading scheme, once it is explained to them. They generally appreciate the degree of detail and "fairness" that is intended by the instructor.

Problems have arisen, however, when teams are highly dysfunctional. These matters will be addressed in another section below.

3. Final report assessment method

At the end of the semester, written reports are submitted by each team. Scheduling usually requires the reports to be graded quite quickly, necessitating a "low overhead" assessment method.

The goals of the assessment technique are: minimise administrative overhead during the assessment process; ensure fair and consistent grading to all students; allow individual student contributions to team projects to be recognised; and account for variation in team sizes and other special circumstances.

Over several years, the author has developed a simple, structured method for assessing these reports; what is presented here is the culmination of many years of trial and error. The method has five steps.

The first step is to **identify the elements of the design work that need to be assessed**. This is done based on the topics covered in the course. Typical topics include problem analysis, requirements definition, ideation, concept generation, concept evaluation, systems architecting, systems interface definition, and detailed design. One should identify specifically the deliverables for each topic. For example, for concept evaluation, the author expects teams to use pair-wise comparison [2] showing how relative weights for evaluation criteria were derived, a weighted decision matrix [3] showing the evaluation of all the team's concepts, and up to five pages of explanatory text.

Furthermore, other collateral topics (spelling and grammar, composition and clarity of presentation, etc.) are defined as distinct elements to be assessed.

Next, the instructor **assigns weights to each assessment element**. This is necessary to take into account the relative importance, difficulty, and complexity of the elements with respect to the overall goals of the course. The author uses pair-wise comparison to establish relative weights of the assessment elements, but in general, an instructor may use any method that seems reasonable – including simply assigning the same weight to every element.

These two steps can be done at any time before the written reports are due. The author normally does this at the beginning of a course, so that the manner of assessment can be explained to students at any time. However, the author has found that most students are not particularly interested in these details – unless they end up with a grade lower than they expected!

The third step is that **students include a WDF with their written report**. The WDF included in a final report is typically longer than those included in milestone reports, but has the same structure and uses the same ranking method. The WDF data is transcribed eventually by the instructor into a spreadsheet that uses the information to assign fractional grades to each team member. This is described in detail in the next section.

The fourth step is that **the instructor assesses the teams' reports**. Each report is assessed as if it had been written by a single author; the instructor intentionally ignores the WDF during this part of the assessment process, because seeing how a team rated its own performance could bias (consciously or otherwise) his assessment. This is the same as grading milestone reports (discussed above). Details and examples of this step are given in Section 4.

The last step is to **scale team members' grades using data in the WDF**. Again, as with the milestone reports, the WDF is used to establish the relative amount of work done by each team member on each task/deliverable for the project. The WDF data is transcribed into the spreadsheet by the instructor. The spreadsheet does all the necessary calculations automatically. The measure of work done is used to adjust the instructor's overall report grade (step 4, above) for each team member. This allows the development of individual student's grades for the project. The details of the calculations are given in Section 4.

The result is a spreadsheet that gives appropriately scaled grades on a per-student basis, taking into account the relative amount of work done by each student, as reported by the students themselves.

4. The project assessment spreadsheet

A sample project assessment spreadsheet is shown in Figure 1. The values in the cells are test values that highlight how the spreadsheet treats certain degenerate cases.

The first row shows the *assessment elements* that the instructor evaluates in the report.

The row marked WEIGHT contains the relative weights for each element as determined by the instructor, using qualitative pair-wise comparison. At the end of that row, the value of 600 is the maximum possible score that a 100% report can attain. OUT OF indicates what the scaled maximum score per student, and SIZE shows the number of members in the team.

The column marked RATING holds the score given by the instructor to each element. Each rating is on a 0-10 scale, with the following correspondences:

0: missing or entirely inadequate.
1-4: inadequate.
5: minimum acceptable/passing score.
6: below expectations/average.
7: acceptable/average.
8-10: excellent

Note that this scale corresponds closely to typical letter grades.

The author used to use a far simpler scale, from 0 to 4 only, that corresponded to letter grades and American-style GPAs. However, that scale is non-linear and led to a surprising amount of confusion among students ("No, Bobby, 2 out of 4 does *not* mean 50%...."). In an effort to help students understand their grades and to lessen the "administrative" burden on instructors, the author changed to the 0-10 scale. While the 0-10 scale did not substantively change the assessment results, students believed the results were more "accurate" (based on informal interviews conducted by the author with students who had completed a course where this method was used).

From a practical point of view, using a linear 0-10 scale also helped simplify tracking grades with Blackboard™, the web-based course management tool that Ryerson University has adopted. Blackboard does not seem to handle non-linear grading scales.

Many other scales could be used; the spreadsheet can be adapted easily to any of them. However, the author has found that a coarse scale (even from 0-4) is sufficient to differentiate student performance. This is because of the many crisply identified elements and the relative weights calculated with pair-wise comparison.

The row marked SCORE in the spreadsheet is the product of the each element's WEIGHT and RATING. The two numbers in red at the end of this row are the sum of the SCOREs and the sum scaled to the value of OUT OF. So, in the sample spreadsheet, the team's report was rated at 397 out of 600, or 6.62 out of 10.

The next five rows (labelled A-E) contain the transcribed values from the team's WDF, one row per student; the labels would normally be the students' names. The work values for the first three elements (labelled ABSTRACT, REFERENCES, and CONCLUSIONS in Figure 1) do not come from the WDF – students do not rate their work on these items, because the author cannot reasonably justify how one might distinguish consistently individual work for these elements. Thus, the instructor arbitrarily assigns a value of 2 to all these work values. While this leads to anomalous grades for students who do no work (e.g. student B), this case does not occur in practice.

Below the WDF data rows are totals (in red) for each element; these values are measures of the total work done by the team on each element of the project.

The column labelled CUMULATIVE POINTS contains the sum of each SCORE multiplied by the WDF entry for a student, normalised by dividing by 2. (Normalisation is not necessary, but helped the author think through the calculations for the spreadsheet.) Thus, student C received 205 CUMULATIVE POINTS, which is a measure of (a) the difficulty of each of the assessment elements, (b) the rating of each element by the instructor, and (c) the amount of work done by the student on each element.

The column labelled RAW SCORE scales the CUMULATIVE POINTS value to the report's grade (6.62 in the example). Thus, a student who worked intensively on every element of the project (e.g. student A) would get the same raw score as the report itself. Students who did no real work (e.g. student B) receive an extremely low raw score. At the bottom of that column, in red, are the average and standard deviations of the RAW SCOREs. So the "average" student RAW SCORE is only 3.2/10, even though the report rated 6.62/10.

Finally, the column labelled FINAL GRADE shows an adjusted score for each student. The adjustment is based on the principle that (a) an average student should get the score given the report as a whole and (b) students who do more work/better than average should get a higher grade. The CUMULATIVE POINTS, however, does not do this.

The adjustment implemented by the author is based on a sliding linear multiplier arranged to raise the actual average RAW SCORE (3.2 in the example) to the report's score (6.62 in the example). The adjustment is also constrained to ensure that students who have RAW SCOREs of 0 or 10 get FINAL GRADEs of 0 and 10 (out of 10). The two values on the right side of the spreadsheet, named *A* and *B* are just partial values used to facilitate entering equations in the FINAL GRADE cells.

Mathematically, the adjustment is calculated with the following equations:

Let:

x: Individual student RAW SCORE

R: Report score (6.62 in the example)

μ: Team average (3.2 in the example)

M: 10 (max allowable grade)

$A = R/\mu$

$B = (R-M)/(\mu-M)$

If a student's RAW SCORE x is <= the average RAW SCORE μ, then the FINAL GRADE is:

$G = Ax$

Otherwise, the FINAL GRADE is:

$G = B(x-M)+M$

The FINAL GRADEs are the grades reported to students and used to calculate course grades.

5. Assessing final oral presentations

Each team must make an oral presentation of their project at the end of the semester. Information and advise on making good presentations is distributed and explained to the students during the semester, and web-based materials are made available for reference.

The author uses the oral presentations to assess the abilities of students to make presentations, *but not to assess the quality of the work done*. As the author often says to his students: "I could have you read the phonebook and still assess your presentation skills."

Each team is typically given 10 to 15 minutes (depending on class size) to make their presentation. A teaching assistant times the teams to ensure no team goes over their allotted time. This frees the instructor to focus on assessing the presentation of each student.

The instructor assesses each student with respect to the following criteria: loudness, eye contact, body language, delivery, attire, use of presentation

materials, and composition. The instructor uses the same 0-10 scale for each criterion as for the written reports. All criteria are given equal weight. The total score for each student is scaled as appropriate, with respect to the value of the oral presentation as part of the students' final course grade.

6. Assessing design journals

Students submit their design journals when they submit their team reports. The journals are checked twice during the semester by the instructor, and feedback on them provided to the students, so by the end of the semester the author expects quite reasonable performance. Web-based material about design journals is provided at the beginning of the semester, so students have many opportunities to read about formatting rules, etc.

Journals are marked against the following criteria.

Format: proper binding and size of journal; page numbers and dates on every page; no blanks pages; etc.

Neatness: not a matter of proper penmanship, is information neatly organised on the page?

Spread: are journal entries evenly distributed over the semester?

Quantity: is there a sufficient amount of material in the journal?

Quality: do the journal entries accurately represent typical activities that students might undertake? Is the journal a true notebook or a narrative written for the grader's benefit? Is there consistency between the journals of teammates?

Again, the author uses the 0-10 scale noted previously to assess each journal against each criterion. For journals, however, the quality criterion is given a weight of 4 and all other criteria a weight of 1, because the quality of journal entries is obviously the most important factor. The weighted total is scaled as required to define a final journal grade.

7. Dealing with dysfunctional teams

No matter what methods are used, there will always be dysfunctional teams – teams that are unable to collaborate, follow schedules, behave professionally, resolve internal conflicts, etc.

Team remediation is a matter separate from that of this paper and so will not be treated directly. However, the author has perceived (albeit only anecdotally) an indirect impact on team performance that is noteworthy.

First, there is the practical matter of using a *personality type indicator* to form teams. It is the author's experience at two Canadian universities that the PTI significantly lowered the number of dysfunctional teams, especially in large, junior classes. This freed more of the instructor's time to help those few teams that were dysfunctional.

Second, when students are made aware of the assessment method and its details, most of them are quite surprised that such an effort is made to devise a method that treats all students equally – especially in a discipline as subjective as design. This seems to motivate students. Many students over the past several years have indicated to the author that the assessment method was a significant driver for them to work well, to be patient, and to collaborate with their teammates.

Thirdly, explaining the assessment method early in the semester lets students understand exactly what will be graded, and which elements of the course are considered most important. In a surprisingly rational twist (especially in the case of 1st year students), many teams plan their projects to ensure that every team member is responsible for one "highly ranked" element, to ensure that every student has a chance to do well. This requires forming a bond with one's teammates that can counteract the typical interpersonal forces that lead to dysfunction.

Thus, it would appear that the assessment method described here helps ensure well-functioning teams at least indirectly.

Another common problem in first year classes is that of "disappearing students" who begin a project with their team but drop the course very late in the semester. One can adapt the team size in this case, which impacts the expected workload per student and alters the distribution of grades in favour of the students remaining in the team.

A similar strategy has been used successfully to address situations in which new students join the course late.

The dysfunctional behaviour discussed above is generally unintentional; it results from inherent personality traits, lack of training, inexperience and, sometimes, ignorance.

However, other kinds of dysfunction are intentional. Some students invest significant time and effort attempting to subvert the assessment scheme. So far, none has been successful, to the best of the author's knowledge. The most typical strategy of these students is to evaluate themselves very well in the WDF. However, this does them no good; the most it does is assure them of an "average" grade, because workload data is normalised with respect to the total amount of work done by the team.

A more malicious activity that the author has been identified on a few occasions is the intentional

marginalisation of particular team members. However, the marginalised students eventually seek out the instructor, and based on their documentation and a careful study of the actual content of the reports, it is usually evident what happened and corrective measures can be taken relatively easily.

The one truly problematic situation is that in which a student who should have done well receives a lower mark than the instructor expects directly because of the nature of the calculations performed by the spreadsheet.

To address this kind of problem, which the author has found happens only *very* rarely, it is essential to keep constant vigil on the teams themselves. The instructor must be able to get a "sense" of how each team is doing, to ensure that the spreadsheet calculations correspond with the instructor's expectations.

In such cases, the instructor will artificially adjust the grades of every team member such that the student who has been under-evaluated receives a more appropriate score. This means that other team members may get higher grades than they deserve, but the instructor deems it more important to ensure that students get *at least* the score that they deserve.

Although no rigorous analysis of the technique has been carried out to date, anecdotal evidence is promising.

Qualitative analysis says that poor students will tend to do poorly on assessments like peer reviews. The author's assessment is in keeping with the team's own assessments. Individual student grades on reports rarely vary more than +/- 10% unless there are extenuating circumstances.

The instructor has found grading using this technique has consistently been simpler and quicker. Furthermore, students appear pleased with the results, at least insofar as no student has successfully argued against this approach. There have been no formal appeals on the grounds of this assessment method. Indeed, most students comment favourably on the *level of detail* of the assessment and the *fairness* of the approach. While there are no assurances that others might benefit as much as the author has, it is possible that some will, and it is in that spirit that this paper was written.

8. Discussion

There are a number of possible improvements that the author is considering to the method reported in this paper. Two particularly significant ones are discussed here.

First, is the obvious procedural improvement of putting the WDF forms on-line. This would require developing a CGI script that would use Web forms to allow individual students to submit their rating data. The script would then use the PTI database to create a downloadable spreadsheet (like that in Figure 1), probably in CSV format. This would be quite straightforward to implement, and should help ensure that students are rating themselves without coercion from other team members. This has not been done yet in anticipation of a more robust online system being developed at Ryerson University (mentioned in Section 2).

Second, the author is considering changing the scale used on WDFs from 0-2 to 0-3, with the following meanings:

0: no work done on an assessment element
1: some work done
2: a significant amount of work done
3: a leadership role was assumed for that element.

Additionally, the author would impose a constraint that any one assessment element can have no more than one student in a leadership role.

However, the author is concerned that students will make tend to overlook the constraint. It is not clear what could be done to correct a situation where more than one student alleges to have assumed a leadership role. In the interests of fairness, this change will not be implemented until this matter can be resolved.

The author has found that using the assessment method described here has freed more time to focus on teaching and maintaining contact with students. This is because (a) a significant amount of preparatory work can be done "offline" and before a course actually starts, and (b) that the grading tasks required of the instructor are simpler owing to the careful use of computer tools like the Web and Excel.

The author bases this on his own experiences in teaching project-based team-oriented design courses for eight years at two Canadian Universities. In courses where assessment was done entirely manually, the author found it necessary to spend much more time on each report and concentrating both on the quality of the student work while *simultaneously* trying to distinguish the contributions of individual students. This essentially produced an *unnecessary coupling* of assessment tasks, which made the assessment more difficult. Without the guidance of a method such as the one described here, one can never be sure that every report is being assessed in a reliable, consistent way.

Furthermore, the author has found that expecting students to report on their own involvement via peer reviews and especially with the WDFs is no less reliable than expecting the instructor and teaching assistants to gauge student involvement. The author

has regularly "double-checked" the assessment method by re-evaluating random student projects using a more conventional, manual method. No significant differences have been found to date.

We note that this method is only part of an overall assessment process including conventional homework, in-class assignments, tests, and exams. It is unclear if the method described here is suitable as a *universal* method.

9. References

[1] D.D. Jensen, M.D. Murphy and K.L. Wood. 1998. "Evaluation and refinement of a restructured introduction to engineering design course using student surveys and MBTI data." *ASEE Conf.*, Session 2666

[2] C.L. Dym. 2002. "Rank ordering engineering designs: pairwise comparison charts and Borda counts." *Research in Engineering Design*, 13:236-242.

[3] G. Dieter. 1983. *Engineering Design*. Mc Graw-Hill, New York.

	ABSTRACT	REFERENCES	CONCLUSIONS	DOC PREP S&G	REPORT ORGAN	INTRODUCTION	BACKGROUND	PDS	PWC	DM	PAS	COMPLEXITY	DETAIL DESIGN	APP A: DWGS	SKETCHES	ORIGINALITY	CUMULATIVE PTS	RAW SCORE	FINAL GRADE	TOTAL		
WEIGHT	**2**	**2**	**1**	**8**	**8**	**2**	**2**	**4**	**4**	**4**	**4**	**1**	**4**	**8**	**4**	**2**				**600**	**OUT OF**	**10**
TEAM X																					**SIZE**	**5**
RATING	**7**	**6**	**6**	**7**	**6**	**7**	**5**	**6**	**8**	**8**	**6**	**3**	**7**	**7**	**6**	**7**						
SCORE	**14**	**12**	**6**	**56**	**48**	**14**	**10**	**24**	**32**	**32**	**24**	**3**	**28**	**56**	**24**	**14**				**397**	**6.62**	
A	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	**397**	**6.6**	8.3		A	2.1
B	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	**32**	**0.5**	1.1		B	0.5
C	2	2	2	1	1	1	0	0	2	0	1	1	1	1	1	2	**205**	**3.4**	6.7			
D	2	2	2	1	0	1	1	2	0	1	1	1	0	0	1	2	**152**	**2.5**	5.3			
E	2	2	2	0	1	1	1	0	0	2	2	0	1	0	1	2	**164**	**2.7**	5.7			
	10	**10**	**10**	**4**	**4**	**5**	**4**	**4**	**4**	**5**	**6**	**4**	**4**	**3**	**5**	**8**	**949**					
																av	**190**	**3.2**				
																sd	**132**	**2.2**				

Figure 1: Sample spreadsheet for a hypothetical team including degenerate cases.