

Preliminary Validation of a Critical Thinking Rubric for Fluid Mechanics and Heat Transfer

Paul Golter
Ohio University*

Gary Brown
AAC&U AAEEL*

Bernard Van Wie
Washington State University (WSU)

* Completed while co-authors were at WSU

Abstract

The preliminary steps toward verifying that a critical thinking rubric has meaning and utility within and outside of classroom use have been undertaken. Analyses offer evidence that the rubric may be used to measure the development of critical thinking in the context of chemical engineering design projects from inception of a problem statement to the final project reports. It has been shown to be an especially effective tool for faculty, who take time to develop high interrater reliability, and for student use in rating each other's projects. In addition, through solicitation of ratings from industry representatives, there are indications that the constructs measured by the rubric have value in a broader, non-academic setting.

Introduction

One key task in developing or evaluating an educational intervention is that of identifying or developing assessment tools that can measure the intended improvements in student learning that will be produced by the intervention. To this end the authors have developed a rubric to measure critical thinking in the specific realm of chemical engineering fluid mechanics and heat transfer design and problem solving. We are certain that the rubric measures things that we, the educators who developed it, value for assigning grades and assessing competence. However, from the perspective that the purpose of undergraduate engineering education is to prepare students for their future careers, it is important to determine whether the skills and thinking characterized by the rubric are similarly of interest to our student's potential future industrial employers.

Specific need for validating our rubric

Our rubric is based on a critical thinking rubric (CTR) developed by an assessment group at Washington State University (WSU). The original rubric is broad and multidisciplinary, has been used throughout WSU (Brown, 2004), and has been a major assessment tool for an NSF grant at another institution (Damron & High, 2008). This original rubric is based on formal analysis, which is rooted in classical rhetoric, and has concurrent validity with multiple definitions of critical thinking (Facione, 1990, 2000;

Paul & Elder, 2005; Toulmin, 1979). Though we have modified the CTR in a way that maintains a logical connection to the construct validity of the original and performed the modification in a way that we believe builds content and expert validity for the purpose of assessing critical thinking in our particular course, we are only meeting a portion of the broader need. We have not ensured that the measurement matters to the future employers of the students. Specifically this is the sub-category of content validity known as content relevance (Chatterji, 2003); some frameworks also refer to this as criterion related validity (Moskal & Leydens, 2000).

In addition to relevance, we need to examine issues that relate to the usefulness of the rubric, both as a teaching tool (Wiggins, 1998; Wiggins & McTighe, 2005) and as a tool that can be distributed and used beyond the pool of individuals who helped with the development. Are there systematic biases in between different groups of raters? Is the measurement credible to all stakeholders? Is it both honest and fair?

In order to begin addressing these questions, the authors undertook a study to utilize our CTR as a major assessment component in a project based junior level chemical engineering course on fluid mechanics and heat transfer. The rubric was used for instructor assessments of student work at various milestones during the semester. As a means to help them understand the use and meaning of the rubric, students also used the rubric to assess a student project from the previous semester. The final project reports were rated by the instructors for grades, the students as their final, and by a group of alumni who were willing to take part.

In this article we begin to examine the content relevance of our rubric by examining the differences and similarities between how the different groups rated the final project reports. These results, and supporting survey responses, provide some insight into probable biases between groups, the credibility of the rubric to different groups, and the portability of the rubric to new users.

Background

Why Critical Thinking?

For at least the past 20 years academics have been

claiming that engineers need more training in what are frequently referred to as "soft skills" (Adams et al., 2011; Dickson & Grant, 2007; Felder, Brent, & Prince, 2011; Litzinger, Lattuca, Hadgraft, & Newstetter, 2011; Varma, 2003). Engineering educators generally think of soft skills as the non-technical items in ABET criterion 3 (ABET, 2003), which include items such as "an understanding of professional and ethical responsibility" and "the broad education necessary to understand the impact of engineering solutions in a global, economic, environmental and societal context." These claims have been echoed by industry, as evidenced by a recent study conducted by the American Association of Colleges and Universities (Peter D. Hart Research Associates, 2008), surveying 301 employers, which shows a frustration with the "broader skills" of recent graduates. Boeing has taken this frustration a step farther and implemented a process that aligns Boeing's hiring practices with their internal evaluation data to track back to college programs that provide them with successful hires (Baskin, 2008). Critical thinking is a key component that underlies many of the soft skills with which we are concerned (Anderson, Howe, Soden, Halliday, & Low, 2001) as it is the base upon which other soft skills such as teamwork, leadership, communication, engineering solution context, and professional and ethical considerations must be built. Without critical thinking all other skills, though they may be present to an extent, will lack the solid thoughtful and rigorous foundation upon which judgements are based.

Critical thinking is a broad term and multiple definitions have been proposed. For example, Facione (1990) gives a definition for critical thinking as follows:

"a purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based. . ."

Looking at this definition of critical thinking, we can see that it also describes, in broad terms, the process by which engineers solve ill-structured problems; namely coming to a solution (judgment) by a process of interpretation, analysis, evaluation and inference using principles (concepts), standards (methods), regulations (context),

and client criteria. Ill-structured problems are problems that do not have a single correct answer. In engineering this includes design and analysis of an existing design or proposal. For this reason problem solving has been selected as an appropriate, domain specific sub-set of critical thinking.

Why Rubrics?

Though multiple standardized tests exist that measure aspects of critical thinking (Anderson et al., 2001; Facione, 2000; King, Wood, & Mines, 1990; Terenzini, Springer, Pascarella, & Nora, 1995), what we are concerned with here is measuring critical thinking as shown in a student's written products. In education assessment terms this is a "performance" and rubrics are especially well suited for the systematic assessment of performances relative to criteria (Wiggins & McTighe, 2005). This allows us to embed the assessment in a task that resembles something that might be done by a professional in the field (Moskal & Leydens, 2000).

From a formative assessment perspective, rubrics allow for feedback that is focused and directly related to the assignment. This provides an opportunity to explicitly show a student where and why they went wrong, which provides an opportunity for them to correct their understanding. It is well established that formative assessment has many benefits (Bransford et al., 2000). Having a formal tool that provides a framework and structure for feedback should be helpful for both students and teachers.

Rubric History and Development

As was previously mentioned, our rubric is a modification of a broader, existing CTR. The original rubric measures critical thinking in 7 dimensions on a 3-point scale; emerging, developing and mastering. The dimensions are: identifies and summarizes the problem/question; identifies the student's own perspective, hypothesis, or position; identifies and considers other salient perspectives and positions; identify and assess key assumptions; identifies and assesses the quality of supporting data/evidence; identifies and considers the influence of context; identifies and assesses conclusions, implications, and consequences. With a well-designed rubric already existing, it may appear that there is no reason to make modifications; however, the authors believed that a more specific rubric, which used language and examples from the discipline of chemical engineering, would be more accessible to and usable by chemical engineering students and faculty.

Details on what was changed and why may be found elsewhere (Golter, Abdul, Thiessen, Brown, & Van Wie, 2010; Golter et al., 2008; Golter, Van Wie, & Brown, 2007; Thiessen et al., 2009), however a brief summary is presented here. The revisions were done over a series of 4 years, and were undertaken by a group consisting of 3 chemical engineering professors and 1 chemical engineering graduate student, with guidance from an educational assessor. The instructional subject matter expertise

of the group would ensure that the modified rubric has utility for measuring constructs that matter for academic use in engineering education. Furthermore, this revision was done in a way that maintained a logical connection with the parent instrument and the research pool it drew from. For example, the categories and criteria of the modified rubric maintained alignment with Facione's broad definition of critical thinking.

The modified rubric, appearing in Appendix 1, contains 9 dimensions: problem identification, fluid mechanics principles and equations, heat transfer principles and equations, fluid mechanics assumptions, heat transfer assumptions, other assumptions, equipment specification, solution quality, and organization/communication. These 9 dimensions correspond to a subset of the dimensions of the original rubric. In both cases, the first dimension is problem identification. The second dimension of the original rubric, presenting the student's own perspective became more specific to engineering problems in the form of presenting students' solution methodology, which became both "Principles and equations" and "Equipment specification" dimensions in the present rubric. This particular dimension, along with assumptions, was made more analytic by separating it out by subject. "Identify and assess conclusions, implications and consequences" became "Solution quality." The dimension of "Identify and consider other salient perspectives and positions" was incorporated into the "Equipment specification" section through metric descriptors "Thorough evaluation of alternatives" and "Solutions . . . can be extended to other situations"; the "Identify and assess the quality of supporting data/evidence" dimension was distributed among the dimensions that come out of presenting the student's perspective, especially in assessing assumptions, solution quality and conclusions, and the "Identify and consider the influence of context" dimension was incorporated into the new rubric's "Solution quality" dimension as represented in the rubric descriptor stating "Solutions are accurate, appropriate, thorough, and clearly linked to design problem." "Organization and communication" was added as a rubric dimension in order to encourage raters to separate the student's writing from their thinking.

The modified rubric works on a scale from 0 to 6 with half, 0.5, point increments rather than using the original rubric's 3 point scale. Conceptually this adds "absent" to the "emerging / developing / mastering" scale and further subdivisions for finer differentiation in rating student assignments. We further define a 4 as the desired competence level of graduating undergraduate engineers which if achieved places them in "developing", but on the cusp of "mastering" range. This allows alignment with the Educational Testing Services (ETS, 2010) convention of using an absolute scale, and anchoring at 2 points, absent (0) and competent (4), the understanding we would expect of the average graduating senior. The scale ranges from 0 to 6, which represents a range from little to no understanding

at levels 0 to 1 up to complete mastery, level 6. The choice of level 4 as an anchor is an important one as it calibrates the rubric at a level closest to the learning expectations in the undergraduate curriculum, and gives room for raters to assign a score for those who go over and above the average, somewhere above average performance but presumably less than the level of an experienced engineer.

Methods

Overview of Class and Assignments

The rubric was used in a required junior level chemical engineering course. This course, Fluid Mechanics and Heat Transfer, is two credits, is the second course in our "transport series", and follows a course titled Transport Phenomena. At this point the students have had the relevant theory in both momentum and energy transport. The course in question is focused on the application of these theories, specifically topics such as sizing pumps and heat exchangers. The instructor has, for many years, included a small group design project as a significant portion of the course. The class was split into groups using fairly standard criteria of mixing ability level, as measured by entering GPA, avoiding creating groups with solitary female or minority students, and trying to ensure that group members had compatible schedules. There were seven groups of five students and one group of four, for a total of 39 students.

For this semester, we used a unique semester-long team project-based pedagogy where we promoted learning based on goals, procedures and outcomes stated in the syllabus, in-class interactive exercises to inform project efforts, the CTR and its use, and well written, but not perfect, sample project artifacts from a previous course offering. We emphasized from the first day of class the program level outcome of creating chemical engineers that individually understand the conceptual and practical aspects of the discipline with the professional skills needed to develop a rigorous multi-component process design in a team setting. Students were told they would be treated as practicing engineers in industry who by the end of the course should be competent to analyze and design fluid mechanics and heat transfer systems in a simulated "real world" environment.

An ill-structured project statement was given in the syllabus to "design a fluid mechanics and heat transfer system to solve a practical problem: e.g., a hot waste stream from an ion exchange regeneration process needs to be treated before being discharged." However, a specific set of learning outcomes, Insert 1, was also given in the syllabus and students told they were to meet all of them. To buttress team efforts on a project proposal due at the end of the first week an example problem statement from a successful past project (see C2J2T team project proposal in Appendix 2) was discussed in class through an instructor-facilitated team-based think-pair-share exercise before the due date. The CTR was also given in the syllabus as a basis on which all project aspects would

- Design a piping system to accomplish a particular task.
- Size a pump to handle a specified fluid-transfer task.
- Analyze an existing heat exchanger and be able to predict its performance for specified inlet fluid properties and flow rates.
- Design a heat exchanger to perform a specified task in an optimum way. You should consider tradeoffs and different heat exchanger designs.
- Explain, in terms of fundamental principles, how a heat exchanger works.
- Understand and model pressure losses and heat transfer in packed and fluidized beds.
- Understand losses due to natural convection, radiation and evaporation.
- Consider safety, and environmental and societal impact.
- Work on teams to design a complex process.

Insert 1: Specific Course Learning Objectives.

be rated. After submission of the first project proposal draft a half-page set of strengths and weaknesses were given to each team to guide improvement on an improved project proposal due two weeks later. Assignments continued in this vein with a recursive continuous improvement strategy.

The next set of assignments focused on using the CTR in rating an entire prior project report (in this case a design of Willy Wonka's chocolate waterfall), in guiding team projects, in rating each other's projects, and in instructor rating and feedback on assignments. To promote adherence to the CTR-centered approach, a graduated scale was given in the syllabus with progressively higher ratings needed on each assignment as the semester progressed. General statements were made on the criterion for each letter grade, e.g., an "A" would be earned when "given an open-ended problem, students demonstrate the ability to define the problem, determine the underlying principles that apply, make reasonable assumptions, come up with a solution and then discuss the implications of that solution." To gain familiarity with the CTR and understand project expectations, in-class scaffolding was provided by the instructor through another guided think-pair-share on aspects of the "Willy Wonka" project, following which individual students performed and turned in their own rigorous analyses and CTR ratings. Feedback was given on student ratings, again using the CTR to rate how well they did in rating the case study project. A mid-semester team project was submitted and rated by the instructor and TAs based on the CTR. Specific feedback including both strengths and improvements was provided on each metric in the CTR for improving the next and final project submission. In lieu of a final exam, each team was given the task of rating another team's final project report, again using the CTR. To differentiate effort among team members, and motivate individuals to participate in the team effort, each student rated all team members, including themselves, on their level of participation in the rating process – this served to provide a multiplier to the team score to provide a method for assessing contributions resulting in a differentiated grade for each student on a team.

Credit for the various assignments was given as follows. The project was worth a total of 60% and was a

group assignment with a graduated increasing percentage weighting correlated to the difficulty of the assignment and expected level of team development within the semester. Project Proposal I was worth 2%, Proposal II 8%, initial project report 10%, and final project report and presentation 40%. The first "case study" or rating of a prior project was worth 10% and was an individual assignment. Standard individual homework assignments made up 20%, and the final "case study", where the student groups evaluated another groups' design made up the final 10% of the course grade.

ETS Convention in Rubric Ratings

We follow the convention laid out by Educational Testing Services (ETS, 2010) when using the rubric. In this convention typically each paper is rated by at least two people. Ideally raters would sit down as a group to rate all of the papers, with everyone rating one paper together at the beginning and at 1-hour intervals until the end. The purpose of this being to establish a norm, or common understanding of the rubric. Periodic checks throughout the rating session provide an indication of how well the norm is being maintained. These norming sessions involve everyone rating a paper individually, followed by a discussion intended to reconcile differences in the ratings. In the standard ETS convention a difference is any rating in which the raters have assigned an average score that is more than 1 point from another rater's, or which crosses the line between developing and competency since this is a key demarcation in achievement of success in an assignment.

Ratings by academics

A group, which will be referred to as academics, did most of the ratings in the study. This group consisted of one professor and three graduate students. One of the graduate students had spent some time as a lecturer, and the other two both have some time in industry; one of these two had also spent the last ten years as the laboratory supervisor for the chemical engineering department. This group spent a full semester utilizing the rubric, norming, and discussing their results. They rated six separate assignments throughout the semester; one assignment was individual, the rest were group. This group began the semester with extensive norming sessions wherein they came to an agreement on how to apply the rubric criteria. No changes were made to the rubric as part of this process.

For reasons of practicality there were some slight deviations from the ETS convention. Rather than having the ratings occur when all were present and having periodic norming sessions, we would norm on one paper, then divide the rest between the raters, bearing in mind that each paper must be rated by at least 2 raters. A subset of papers was still rated by everyone in order to check whether a new norming session was needed. An important reason

for this deviation from convention was that we needed to provide a fast turnaround since the rating was being used as a primary grade assigning criteria for the course. Nonetheless a meeting was still held at the end of each rating period to provide time for the group to discuss any differences that arose during the rating process. Again, and in response to student feedback after rating the first group assignment, we deviated from convention by norming to a difference of 0.5 rather than 1. For the purpose of assigning grades, this extra degree of agreement made sense. The norming process itself consisted of looking at the results from each rater and discussing any ratings that varied by more than 0.5 from the average. To resolve the differences we would examine the scores in each category to determine where the difference arose, and then discuss our reasoning for the score. After discussion, each rater was given an opportunity to alter their score. Usually a small shift was all that was needed to reconcile the scores.

Ratings by students

The second case study provides a data point for students using the rubric to rate the same set of assignments the instructors were rating. At this point, the very end of the semester, the students had been working with the rubric throughout the semester and were familiar with it. This was done in lieu of a final exam and had a significant impact on the student course grades. It is possible that this influenced students to be extra critical of each other's work. Also, since each group rated only one other group's paper, between group IRR statistics are not possible.

Ratings by employers

With the assistance of the college development office, we recruited a group of eight alumni to rate the students' final design presentation and to provide feedback, using the questions shown in Table 1, on the applicability of both the assignment and the rubric to an industrial career. These alumni ranged in experience from 20 to 50 years and represented industries including chemicals, petroleum, pharmaceuticals, environmental cleanup, aluminum and pulp and paper. Four of the alumni had last held executive level positions, and the remaining four held management positions. These alumni represent a highly experienced group who should be accustomed to evaluating designs and proposals. Their insights here should provide a comparison between industrial and academic expectations for this type of assignment.

The papers were distributed so that each paper was rated by at least two alumnus. This, however, required two rounds of recruiting due to alumni either dropping out of the study or not responding. In spite of the extra recruiting, we were only able to collect one numerical rating for one of the papers, which were somewhat corroborated by the written comments of a second rater. Since this was a geographically dispersed group, we made no attempt to arrange a norming session. The ratings were collected using an online survey system.

Now review the assignment prompt and evaluation criteria					
How accurately do the skills described above ¹ reflect the skills and knowledge required of professionals in your employee?					
Extremely Inaccurate	Mostly Inaccurate	Moderately Inaccurate	Moderately Accurate	Mostly Accurate	Extremely Accurate
How frequently do the criteria in the rubric (in concert with the project) reflect the skills and knowledge required of professionals in your employee?					
Never	Very Rarely	Rarely	Occasionally	Very Frequently	Always
Please explain your responses. Reference specific rubric dimensions and criteria when possible.					
How important are the skills described in the previous rubric for success in the professional world?					
Unimportant	Of Little Importance	Moderately Important	Important	Very Important	Extremely Important
From your professional vantage: To what extent does this assignment provide the student an opportunity to prepare for work in industry?					
Greatly Constrains the Opportunity	Mostly Constrains the Opportunity	Slightly Constrains the Opportunity	Slightly Affords the Opportunity	Mostly Affords the Opportunity	Greatly Affords the Opportunity
From your professional vantage, would you hire people who have produced the kind of project you have reviewed? Please indicate the level at which you hire.					
Not Ready for Employment	Internship with Close Supervision	Internship	Entry level	Assistant Managers	Associate Managers
What would you do to improve this assignment? Why?					
<small>¹ The term "above" here is in reference to the rubric. The alumni would have just finished rating a paper using the rubric at this point.</small>					

Table 1: Survey given to the alumni raters regarding the rubric and assignment. These results were given a numerical value ranging from 1 to 6 in intervals of 0.5.

Statistical Methods – Reliability

The previous discussion of what constitutes a “difference” leads directly to one of the key statistics for this type of rubric use, Inter-Rater Reliability (IRR). Quite simply, this is the ratio of the number of comparisons of the aver-

age score that agree within 1 divided by the total possible number of comparisons. For a given paper, each rater can compare scores with each other rater. In the case of four raters, this gives six possible comparisons. Mathematically the total number of possible comparisons is as follows:

$$\#comparisons = \sum_{i=1}^{NUMBER\ OF\ RATERS} i - 1.$$

By convention, an IRR of 70% or greater is satisfactory. The IRR can be calculated for ratings both within group and between groups. By grouping the ratings we can see whether we had reliability between the different groups of raters. For example, if we group the alumni and academic ratings the resultant IRR includes the reliability between these two groups. IRRs within and across groups highlight differences in opinion on the relative strength of an argument being made in a student paper. Where IRRs are in agreement it may imply strong agreement on meaning of the rubric criteria or at least that students have met the criteria, albeit that different raters may have seen the criteria met by differing statements in a report. There is a possible source of bias in this due to the larger sample size for the academic raters. It is likely that this reduced the influence of other groups’ ratings when calculating a between group IRR.

Qualitative Methods – Credible to all stakeholders

To examine whether our rubric is part of a credible assessment, we need to determine if what we are measuring matters and makes sense to the stakeholders in our students’ education. In this regard our study was exploratory. Along with rating a small number of papers, alumni were given a brief survey requesting their feedback on both the rubric and the project assignment. The purpose of this survey was to assess the content relevance of the assignment and rubric to an industrial career. This was appended to the survey page for collecting the alumni’s ratings. This survey was not rigorously developed or administered, and is meant to provide direction for further inquiry rather than providing definitive answers. Table 1 shows the survey questions and response choices.

As another exploratory companion to the alumni survey, we held a brief informal discussion with a trio of students who were involved with another aspect of this research project. During this discussion we probed their perceptions of the rubric and the assignment.

Results

Overview

The first, and easiest to assess, set of data to examine is the scores from the second project submission. This may be seen graphically in Figure 1. There is a visually apparent trend toward the students rating more harshly and alumni rating more easily compared to the instructors. However, as can be seen from the error bars, the differences are not statistically significant, possibly due to the small sample size (n=8). At this small sample size, a difference would need to be quite large in order to be significant. As none of the differences between any of these groups even approach one standard deviation, no further statistical testing was required to determine that the differences

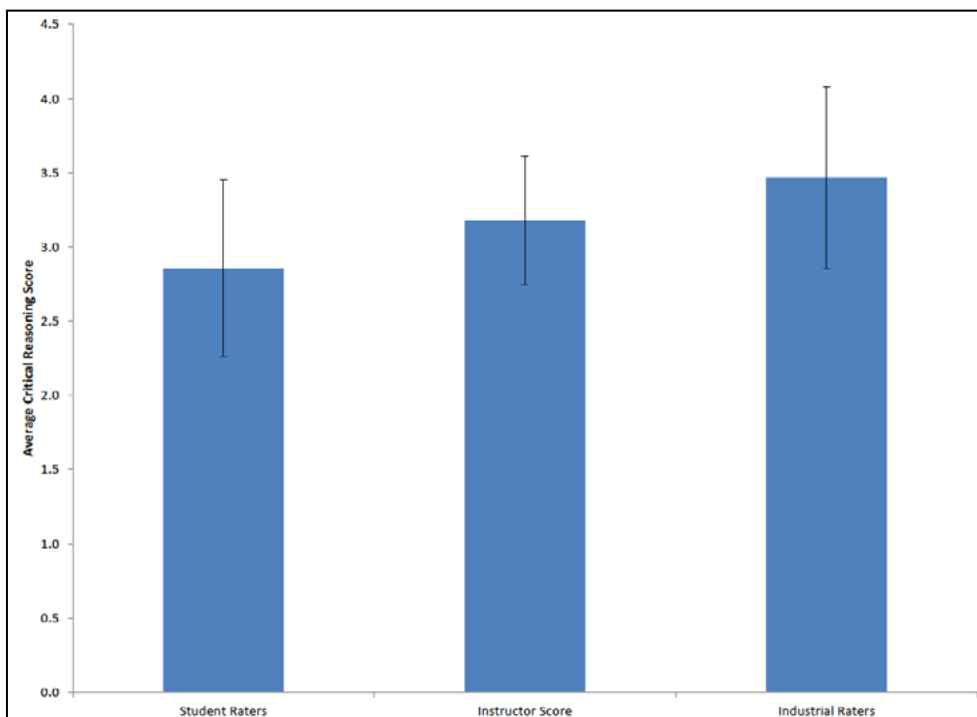


Figure 1: Average ratings by each group of raters, sorted from highest to lowest, with 1 standard deviation error bars.

	Proposal 1	Proposal 2	"Case Study" 1		Project 1			Project 2		"Case Study" 2	
			Initial	Rec.	Initial	Rec.	Alumni	Initial	Rec.		
IRR	94%	86%	77%	95%	100%	100%	100%	36%	100%	100%	

Table 2: Within group inter-rater reliabilities, over the semester. Unless otherwise noted, IRRs are for the academic group. Where possible both initial and reconciled IRRs are given.

	IRR: Faculty, Alumni, & Students	IRR: Faculty & Alumni	IRR: Faculty & Students	IRR: Alumni & Students
Average	75%	70%	97%	56%

Table 3: Inter-rater reliabilities between groups for the second project assignment.

Rubric Scale	IRR
Problem Identification	94%
Fluid Mechanics Principles	80%
Heat Transfer Principles	100%
Fluid Mechanics Assumptions	89%
Heat Transfer Assumptions	94%
Other Assumptions	72%
Equipment Specification	89%
Solution Quality	89%
Organization & Communication	94%

Table 4: Inter-rater reliabilities for each individual rubric scale. The results shown are the reconciled results for the academic raters and project 2.

	Average	Descriptor of Average	Range	Consistency	Abstain
Skills and Knowledge Accurately Described	4.3	Mostly Accurate	3.5 – 5.0	25%	3
Skills and Knowledge Frequently Used	4.2	Occasionally	3.0 – 5.0	50%	3
Skill Importance	4.2	Important	3.0 – 5.0	0%	2
Assignment help prepare students for industry	4.5	Mostly Affords the Opportunity	2.5 – 5.5	20%	2
What Level of Hire	3.1	Internship / Entry Level	2.0 – 4.0	NA	2

Table 5: Alumni feedback on the rubric. Possible answers ranged from 1 to 6. In this table, consistency refers to the percentage of times alumni rated the same way more than once.

could be explained by chance. When phrased as grades the differences are more striking with alumni giving five scores of A- or better (ranging from A+ to C, averaging A-), academics giving only two As (ranging from A to D+, averaging B), and students giving only one A- (ranging from A- to D-, averaging B-).

Reliability Results

The IRRs within each group may be found in Table 2 which, as shown, range from 86–100% for reconciled values with improvement from the 77% for initial use with the "Case Study" to 100% for initial ratings of the Project 2 and final "Case Study". An exception is the alumni IRR of 36%. Where the data are available the initial IRR is included. All of the IRRs were calculated using the ETS standard of differences of 1.0 or less being not significant, even though a more stringent criterion of 0.5 was used

for establishing the actual scores to provide the students with greater perceived accuracy and granularity for grading. Similar tabulated values of IRRs for the varied pairs of faculty, alumni and student grouped data sets may be found in Table 3. The IRRs for reconciled project 2 scores by individual rubric scale are included in Table 4 ranging from 72–100% for the nine categories in the rubric.

To test the stability of the results a single paper was selected at random after a significant time (after an extensive revision and resubmission process) and rated by one member of the academic group. The paper had received an average rating of 3.4 initially while the rater's original rating was 3.6. The new rating was 2.8. Since the difference is less than 1, the score had not changed significantly by the ETS convention, indicating that, once a rater is properly trained in the rubric's use, their ratings will likely not appreciably change over time. However, if

it is desired to maintain agreement to within 0.5, as was important in this class for establishing grading criteria, it may be important for raters to re-norm if significant time has elapsed before another assignment is rated. However, over the course of a semester we were able to maintain this higher, within 0.5, level of accuracy.

Credibility of Stakeholders Results

Alumni answers to the survey questions are summarized in Table 5. In addition to the numerical summary, average and range, there is a pair of extra columns that reflect a misunderstanding with regard to the survey. Though the survey was specific to the rubric rather than the paper being rated, in several instances, an individual rater gave different scores for questions that should have had consistent answers regardless of the student paper being rated, since the questions were focused on the rubric. The consistency column is added to show the percentage of times where the raters gave the same score across different papers, and the abstention reflects that some questions were randomly skipped.

When asked to give their reasoning some of the alumni responded with comments that were clearly directed at the reports.

"Report is well organized and clearly discussed. Engineers usually have trouble writing and explaining with the written word."

Others indicated that the rubric was a good starting point for this type of assignment.

"The rubric elements plus economic analysis would also be required."

"The rubric covers the essentials in the assessment of a ChE's [Chemical Engineer's] design and does an excellent job of (assessing) their needed capabilities to start. . . . recommend a specific addition to the rubric covering the specific writing skills and not hidden in the communication assessment."

Finally we asked what they would do to improve the assignment. Responses largely dealt with issues that are, currently at least, outside of the scope of our course.

"Addition of economic analysis and formal report writing. These are essential for opportunity for advancement."

"Pictures and drawings help. A course in . . . PFD [process flow diagram] software would help their ability to communicate intent and technical thinking. This could be a simple part of ChE 101 orientation course."

Two of the suggestions would be implementable in our course. First adding an abstract or executive summary to the report:

"Organization of their thinking . . . start by adding an abstract . . ."

This would also address many of the comments that came up earlier in the ratings of the each report, and help support the organization and communication dimension of the rubric. Lack of an executive summary or detailed flow

diagram was a very common complaint in the alumni comments. The other suggestion gets back to the other portion of the stated goal for this class, being able to analyze an existing design.

"... many times work in industry involves investigation of processes that are functioning below optimum. The challenge in these cases is identification of the mechanical or process issue causing energy or raw material waste."

We had attempted to meet this goal by having the students evaluate a past report at the onset of the course, as well as each other's reports at the conclusion of the course. This portion of the course was outside of the scope of this survey, but it is encouraging to see that the alumni have included it.

Discussion with Students: Rubric

We held a small, informal discussion with three students, who were involved with other aspects of this research, to elicit their viewpoint regarding the rubric and the assignments. One student mostly dominated the discussion. The manner in which we approached the course made the linkage to industry explicit. They lacked the perspective to know whether the project really did simulate the activities of an engineer in industry though.

"We had no clue, cause (sic) that was our first project. That was our first ChemE [chemical engineering] project really."

One student claimed that using the rubric also left the students without a sense of who their audience was, and made the reports more difficult to write; though it also provided motivation to make the reports more professional in appearance.

Discussion

Overview

We had expected to see the alumni give the harshest ratings and students the easiest, the data in figure 1 shows the exact opposite trend though. When queried about this, the individual from the development office reported that it is exactly the trend he would have expected. His hypothesis leads to one possible explanation for this. Due to using the development office to recruit participants, we had a group which consisted of senior level to retired people who were perceived as being interested in and willing to help the department and students. Whether a different result would be obtained if we were working with younger engineers remains to be seen. Another possible explanation could be that the student ratings were done as a final exam for the course. The depth of the comments and analysis given as feedback by the student groups indicate that they took the task very seriously. Informal discussions with the students reinforce this. The students stated that they did not want to be penalized for missing anything.

Reliability Discussion

Unsurprisingly, the results in Table 2 show that we achieved extremely high reliabilities for a group of raters who had the opportunity to work together multiple times over

the course of a semester, the academic raters; a group that had no opportunity to work together at all, the alumni, had very low reliabilities. When taking the average IRR for a combined group of raters, all of the combinations that included the academic raters were at least 70%, which represents adequate reliability. In contrast, the combined reliability of the students and alumni was low (average 56%), see Table 3. It is likely that this is an artifact of the sample size. There were more comparisons made by academic raters than by either of the other groups. This will skew the results towards the reliability between just the academic raters. We would also expect and hope that academics represent a bridge of sorts between the industrial and student viewpoints. That the academic raters are reliable both with the student and the alumni raters could be taken as evidence that this is the case. Another possible interpretation of these results is that they indicate that academics and students share a similar perspective about the meaning of grades in school, whereas alumni do not share this perspective with either students or academics. This possibility raises some potentially interesting questions about the perceived level of readiness of new graduates and the amount of learning that takes place during the first few post-college years of an engineering career.

The results in Table 4 do show the 100% IRR on the average found for academic raters on project 2 does mask some differences. However, all of these IRRs are above the minimum criteria of 70%. As an interesting aside, all but one of the differences noted had a discrepancy between competent (4.0) and developing (3.5), while within 0.5 cross the competency anchor of 4.0. The one remaining difference was a difference of 1, and would not have counted if not for the student-requested change from ETS convention to a tighter 0.5 scale for better differentiation of grades. The IRR breakdown reveals the rubric may be used effectively as is, though the ideal for what students should graduate capable of doing is less well developed when it comes to "Other Assumptions" such as safety and environmental concerns. This may be taken into account in future refinements of the rubric.

Credibility Discussion

Though the results from the alumni were exploratory and the survey implementation resulted in complications, the results are roughly indicative that we have succeeded in examining a construct that is important to an industrial career. Based solely on the average responses in Table 5, it appears that the rubric measures a set of skills which are important to an industrial career and that the assignment provides a good opportunity for students to practice these skills. However, taking the abstentions and the consistency into account, we are not justified in making a strong claim. Based on the range of scores though, we can say that indications are that the rubric probably measures a set of important skills and the assignment seems to provide a good opportunity to practice the measured skills. The written comments, especially those regarding the rubric as a good

starting point for design, support this claim. It is clear that in follow up work, questions regarding the quality of the rubric and assignment need to be given separately after alumni have had a chance to rate some papers.

In some ways it is more difficult to assess the broader meaning behind the question of "at what level would you hire. . . ." The average result of 3.1 is on the cusp between internship and entry level, 3.0 being the highest score for internship and 3.5 being the lowest score for entry level. One would expect students to easily be ready for an internship by the end of their junior year, in this respect the students are meeting our expectations and in some cases exceeding them, specifically four of the eight groups rated at entry level. Based on this, the assignment appears to provide students a good opportunity to showcase their skills. This feedback also lends support to the indications that the assignment and rubric help build skills necessary for an industrial career.

Discussion with the students, albeit in very small numbers, illustrated that the students harbor some misconceptions regarding what their future careers will be like. The discussion about the intended audience for their reports highlights this. The students had the impression that the initial audience for their reports would be the final audience. The reality in the workplace, either industrial or academic, is that you have no control over where your writings end up. Ironically, even though the students complained about this, they also credited it with helping them produce a more professional report. While this portion of the work was highly informal, it serves to remind us that students need more than just technical preparation.

Coupling these interview comments with the reliability results yields a potentially disturbing insight. If students claim to be basically clueless and have no basis for the projects they did, but they share a common perspective with academics, what does this say about academics? Are we also essentially clueless, or are we just failing at introducing our students to project work prior to their senior year?

Further Implications

While the work presented in this paper is an initial effort to validate a critical think rubric in the chemical engineering context, we have demonstrated success in terms of inclusion of specific categories for application in fluid mechanics and heat transfer, achieved an outstanding level of agreement between trained faculty raters to the extent that in implementations of the rubric at the final report stages there was 100% IRR without the need of norming, that students trained in the assessment approach provide results that are of lower overall ratings than the faculty, though still in excellent agreement with a 97% IRR, and that while faculty, alumni and students are within an acceptable 75% IRR, that improvements that can be made are obvious in that industrial raters will benefit from a norming process at least between industrial members and perhaps with faculty members as well.

Conclusions

The rubric we have developed shows indications of being a valid tool for measuring a student's critical thinking with regard to engineering design and problem solving in a manner that is consistent with the needs of industrial employers. Unfortunately this statement must be tempered with the knowledge that our efforts to validate this instrument in the broader context of industrial needs is only beginning and we have only established a preliminary validity in this regard. Aside from this exploratory conclusion, the data for the academic raters clearly show that providing the raters an opportunity to get together and discuss the rubric and their ratings very significantly enhances reliability in rubric measurements, as starkly evidenced by the 100% IRR among academic raters compared to the 36% IRR among alumni.

A brief analysis of combined reliabilities indicates that academics probably occupy an in-between perspective between that of students and alumni. When combined with the student comments, which indicate an inaccurate view of the workplace, this indicates an opportunity for faculty to address student misconceptions regarding what the workplace will be like. Based on the reliability results however, it is also possible that student and academic perspectives may be closer together than academic and alumni perspectives. These results are ambiguous due to differences in sample size and the very high reliability among academic raters potentially masking the results.

It is clear that much more needs to be done. To begin with we need to sample a broader cross-section of alumni. Selecting two papers from the set used in this study and soliciting ratings from a large group of alumni, ranging from five to 50 years in experience, would determine whether there are effects based on experience level. The survey requesting feedback specifically on the rubric and assignment needs to be given separate from the rating collection. Both this survey and either a formal interview or survey with students need to be formally developed in order to draw stronger conclusions from the data.

Acknowledgements

The authors acknowledge support from the US National Science Foundation CCLI Grants DUE-0618872, 1023121, 143267, the Norcliffe Foundation, and a World Bank STEPB Grant through collaboration with the chemical engineering Department at Ahmadu Bello University, Nigeria, all on advancing and assessing use of new pedagogies fostered by use of Desktop Learning Modules. The authors appreciate valuable comments and guidance from members of the Washington State University Engineering Education Research Center, namely Profs. Shane Brown, Denny Davis, and Richard Zollars, and Dr. Ashley Ater-Kranov. We further acknowledge assistance from TAs Baba Abdul and Ashfaq Ansery as well as from WSU Engineering Development Office Director Don Shearer.

References

- ABET. (2003). Criteria for accrediting engineering programs. Retrieved from <http://www.abet.org/images/Criteria/E001%2004-05%20EAC%20Criteria%2011-20-03.pdf>
- Adams, R., Evangelou, D., English, L., Figueiredo, A. D. D., Mousoulides, N., Pawley, A. L., . . . Wilson, D. M. (2011). Multiple perspectives on engaging future engineers. *Journal of Engineering Education, 100*(1), 48-88.
- Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science, 29*(1), 1-32.
- Baskin, P. (2008). Boeing to rank colleges by measuring graduates' job success. *The Chronicle of Higher Education, 55*, A1.
- Bransford, J. D., Brown, A. L., Cocking, R. R., Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington DC: National Academy Press.
- Brown, G. R. (2004). Getting real in the academy. *About Campus, 9*(4), 10-18.
- Chatterji, M. (2003). *Designing and using tools for educational assessment*. Boston, MA: Pearson Education.
- Damron, R. L., & High, K. A. (2008). *Innovation in linking and thinking: Critical thinking and writing skills of first-year engineering students in a learning community*. Paper presented at the ASEE/IEEE Frontiers in Education Conference, Saratoga Springs, NY.
- Dickson, B. R., & Grant, C. D. (2007). *Engineering management skills: The present and the future for technical graduates*. Paper presented at Offshore Europe, Aberdeen, Scotland.
- ETS. (2010). Retrieved from <http://www.ets.org/portal/site/ets/menuitem.3a88fea28f42ada7c6ce5a10c3921509/?vgnextoid=85b65784623f4010VgnVCM1000022f95190RCRD>
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction: Research findings and recommendations*. Retrieved from Newark, DE: <http://eric.ed.gov/?id=ED315423>
- Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic, 20*(1), 66-84.
- Felder, R. M., Brent, R., & Prince, M. J. (2011). Engineering instructional development: Programs, best practices, and recommendations. *Journal of Engineering Education, 100*(1), 89-122.
- Golter, P. B., Abdul, B., Thiessen, D., Brown, G., & Van Wie, B. J. (2010). *Shifting gears: Moving away from the controlled experimental model while improving rigor in engineering education research*. Paper presented at the American Society for Engineering Education Annual Conference, Louisville, KY.
- Golter, P. B., Jiang, Y., Thiessen, D., Van Wie, B. J., Brown, G. R., & Yurt, N. (2008). *Developing, implementing and assessing desktop scale engineering laboratory apparatus*. Paper presented at the Global Colloquium for Engineering Education, Cape Town, South Africa.
- Golter, P. B., Van Wie, B. J., & Brown, G. R. (2007). *Comparing student experiences and growth in a cooperative, hands-on, active, problem-based learning environment to an active, problem-based environment*. Paper presented at the American Society for Engineering Education Annual Conference, Honolulu, HI.
- King, P. M., Wood, P. K., & Mines, R. A. (1990). Critical thinking among college and graduate students. *The Review of Higher Education, 13*(2), 167-186.
- Litzinger, T. A., Lattuca, L. R., Hadgraft, R. G., & Newstetter, W. C. (2011). Engineering education and the development of expertise. *Journal of Engineering Education, 100*(1), 123-150.
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation, 7*(10), 1-6.
- Paul, R., & Elder, L. (2005). *A guide for educators to critical thinking competency standards: Standards, principles, performance indicators, and outcomes with a critical thinking master rubric*. Dillon, CA: Foundation for Critical Thinking.
- Peter D. Hart Research Associates. (2008). *How should colleges assess and improve student learning? Employers' views on the accountability challenge*. Retrieved from http://www.aacu.org/sites/default/files/files/LEAP/2008_Business_Leader_Poll.pdf
- Terenzini, P. T., Springer, L., Pascarella, E. T., & Nora, A. (1995). Influences affecting the development of students' critical thinking skills. *Research in Higher Education, 36*(1), 23 - 39.
- Thiessen, D., Brown, G., Yurt, N., Abdul, B., Golter, P., & Van Wie, B. (2009, 2009). *Aligning assessment tools with course subject and goals*.
- Toulmin, S. (1979). Can science and ethics be reconnected. *Hastings Center Report, 9*(3), 27-34.
- Varma, A. (2003). Future directions in ChE education: A new path to glory. *Chemical Engineering Education, 37*, 284-289.

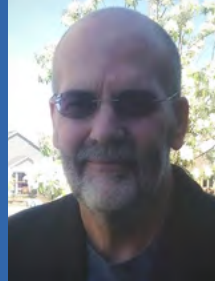
Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.

Wiggins, G., & McTighe, J. (2005). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Paul Golter did his B.S. at the University of Idaho, his M.S. and Ph.D. along with postdoctoral research at Washington State University, and held an adjunct lecturer position at the University of Idaho. His research has been focused on developing and assessing “Desktop Learning Modules”. He is currently a mechanical engineering lecturer at Ohio University.



Gary Brown is a Senior Fellow with the Association of American Colleges and Universities (AAC&U) and Academic Director of the international ePortfolio organization, AAEEBL. With AAC&U, Dr. Brown recently served as the assessment lead exploring the potential of Lumina’s Degree Qualifications Profile (DQP) as a way to effectively mediate transfer. The initiative engaged 13 states including two and four-year institutions. At AAEEBL, Dr. Brown co-directs the research initiatives and co—chairs the Vendor Advisory Council. Dr. Brown serves on advisory boards for Quality Matters and Campus Technology. His teams have received six best research awards and two best publication awards.



Bernard Van Wie did his B.S., M.S. and Ph.D. along with postdoctoral research and visiting lecturer positions at the University of Oklahoma. He has been on the Washington State University (WSU) Voiland School of Chemical Engineering and Bioengineering faculty for 33 years, 19 years of which have been focused on developing and assessing “Desktop Learning Modules”. He won WSU’s 2009 Marian Smith and 2016 inaugural Teaching Innovation Awards. His technical research is in the areas of biomedical engineering and biomass conversion.



Appendix 1 – Critical Thinking Rubric

Dimension	0	1	2	3	4	5	6	Score
1. Problem / Question	Does not identify a specific question or necessary and appropriate system or systems. The question or system, if identified, is confused or simplistic. Little evidence of understanding of salient constraints (e.g., cost of water, available space, etc.). Little evidence that goals or potential/useful outcomes identified and understood.		Identifies a somewhat focused question/goal that is interesting but not particularly challenging or is simplistic, tends to ignore essential constraints. The problem/goal is vaguely defined and characterized. Consideration of goals or outcomes is vague or suggests important omissions of key considerations.		Identifies a focused, unique, original question/goal that is challenging and well defined. Thoroughly understands constraints. The question/goal is thoroughly defined and characterized.			
2. Fluid Mechanics Principals / Equations How chosen/appropriateness Depth of Use / Integration Completeness	Analysis of fluid mechanics principles is absent, incomplete or inappropriate relative to the presenting problem—or insufficiently related to the challenge the project entails. There is little Interpretation of fluid mechanics principles, or there may simply be a restatement of inherited facts.		Analysis of fluid mechanics principles is essentially correct; perhaps some is off tangent or barely related to the presenting problem, but generally in the vicinity of the challenge the project entails Interpretation is adequate and clear, though perhaps not fully integrated with other sources or perspectives. Barely extends, if at all, beyond routine exploration.		Analysis of fluid mechanics principles is thorough and correct. Interpretation is well integrated with other chemical engineering principles, sources and professional perspectives.			
3. Heat Transfer Principals / Equations How chosen/appropriateness Depth of Use / Integration Completeness	Analysis of heat transfer principles is absent, incomplete or inappropriate relative to the presenting problem—or insufficiently related to the challenge the project entails. There is little Interpretation of heat transfer principles, or there may simply be a restatement of inherited facts.		Analysis of heat transfer principles is essentially correct; perhaps some is off tangent or barely related to the presenting problem, but generally in the vicinity of the challenge the project entails Interpretation is adequate and clear, though perhaps not fully integrated with other sources or perspectives. Barely extends, if at all, beyond routine exploration.		Analysis of heat transfer principles is thorough and correct. Interpretation is and well integrated with other chemical engineering principles, sources and professional perspectives.			
4. Fluid Mechanics Assumptions	There is little or no synthesis from literature; or what is presented is incoherent or patched together without explanation or demonstration of underlying fluid mechanics assumptions. Assumptions tend to be confused or perhaps contradictory. Key aspects of the challenge are neglected.		Application of material discerned from literature or outside reading is adequate in scope and accuracy, though perhaps slightly confused at times or partially inaccurate. Sometimes questions and attempts to support the validity of assumptions.		Application of material discerned from literature as well as from outside reading is complete and accurate. Questions and supports the validity of assumptions.			
5. Heat Transfer Assumptions	There is little or no synthesis from literature; or what is presented is incoherent or patched together without explanation or demonstration of underlying heat transfer assumptions. Assumptions tend to be confused or perhaps contradictory. Key aspects of the challenge are neglected.		Application of material discerned from literature or outside reading is adequate in scope and accuracy, though perhaps slightly confused at times or partially inaccurate. Sometimes questions and attempts to support the validity of assumptions.		Application of material discerned from literature as well as from outside reading is complete and accurate. Questions and supports the validity of assumptions.			
6. Other Assumptions	There is little or no synthesis from literature; or what is presented is incoherent or patched together without explanation or demonstration of underlying heat transfer assumptions. Assumptions tend to be confused or perhaps contradictory. Key aspects of the challenge are neglected.		Application of material discerned from literature or outside reading is adequate in scope and accuracy, though perhaps slightly confused at times or partially inaccurate. Sometimes questions and attempts to support the validity of assumptions.		Application of material discerned from literature as well as from outside reading is complete and accurate Questions and supports the validity of assumptions.			
7. Equipment Specification	Improper or superficial equipment selection. No attempt at sizing, evaluating alternatives, etc. Equipment selection demonstrates little depth, attention to or understanding of professional conventions. Simple list of equipment.		Proper equipment selection, but limited in terms of evaluation of alternatives, sizing, etc. Equipment selection demonstrates some use of professional convention (rules of thumb). Some systems level thinking.		Equipment selected and optimally sized after a thorough evaluation of alternatives. Thorough use and understanding of appropriateness of professional convention. Integrated systems thinking.			
8. Solution Quality	Solutions are missing or inaccurate / unreasonable. The implications of the solutions are absent.		Solutions are reasonable, though perhaps incomplete or limited. Solutions relate to the design problem and arise from the analysis presented, though there may be gaps or redundancies. May include speculation about implications -- mostly plausible, but not necessarily reasonable useful or creative.		Solutions are accurate, appropriate, thorough, and clearly linked to design problem. Solutions and recommendations are qualified, balanced and can be extended to other situations.			
9. Organization and Communication	Presentation of problem, analysis, solution and interpretation seems haphazard, inconsistent, or misleading; one or more elements may be missing or confused. Organization of ideas / multiple errors obscure meaning, and may misinform or mislead audience. Many parts seem difficult for the audience to follow. Communication style, written and/or oral, is not appropriate to this discipline, or is confusing. Does not use language of the discipline, or uses it incorrectly. Frequent errors may obscure ideas. Media and format are poor choice for content; some materials confusing or distracting, or served as filler. Communicates little or no engagement or ownership of the work.		There is a discernable progression from problem to analysis, solution and interpretation, linked to the problem and solution. Presentation is adequate for intended audience, though there may be occasional gaps, errors, or inconsistencies which require effort from audience in order to understand. Communications style, both written and oral, is appropriate to this discipline; though not at a professional level, communication is adequate. Incorporates some language of the discipline, though imperfectly. Some errors may distract audience. Appropriate media and format used for content. Most materials clear and pertinent. Communicates some engagement and ownership of the work.		Progression from problem to analysis, solution and interpretation is concise, creative, and clearly ties problem to solution and implications. Needs and interests of intended audience effectively inform presentation's approach and organizations. Audience seems well able to follow the presentation. Communications style, both written and oral, is appropriate to this discipline, and is polished, professional, and virtually error free. Uses language of the discipline fluidly and effectively. Appropriate media and format used for content; all materials clarified, with pertinent or high interest information. Clearly communicates engagement and ownership of the work.			

Appendix 2 – C2J2T Project Proposal

Due to the nature of Christmas, elves find themselves incredibly busy throughout the working year. Because of this, the plants should be running at optimum efficiency in all aspects. Presently, the production of tea and cocoa has a few inefficient aspects. The beverages are prepared in a separate building from the factory creating a gap in work whenever the elves require something to drink. It also requires them to trek through the polar winds to acquire a hot drink, a counterproductive process.

To fix this, we propose a means of transporting the drinks directly from the place of production to the factory itself. We will use insulated tubing located above the ground. The process will begin with heating water to roughly 136 degrees Fahrenheit. To expend minimal energy, we will first use several heat exchangers utilizing the effluent gas from the coal furnace located in the factory. This gas will be sufficiently hot to increase the

water temperature a fair amount without requiring any additional energy on our part. To further heat the water to the desired temperature, we believe the reindeer can be useful. By having the reindeer power a treadmill or a turbine, we can simultaneously generate energy for heating and allow them to continue exercising on days other than Christmas.

When the water is heated to an appropriate temperature, we will split it into two separate streams. One stream will pass through a packed-bed reactor filled with peppermint tea leaves which will infuse the passing water. The separate stream for hot cocoa will go into a continuous stir reactor where it will be combined with milk, cocoa, and sugar. Both streams will then pass into the factory itself, where dispensers will be located.

The design can be modified to suit any number of elves, but initially we plan to produce 10 cups of hot choc-

olate and tea per elf in the factory daily. We can modify the flow of water using pumps located before and after the splitter to suit change in preference. All pipes will be of appropriate gauge and size to match the pressure and temperature of the liquid, and appropriately insulated to weather the cold. The original design will be made slightly larger than required, in case there is a future increase in elves.

Overall, the design should accomplish the task of providing refreshments to the elves. Costs have been minimized by using the reindeer and coal furnaces already located in the plant. Due to an increase in elf efficiency, production will go up and more children will receive presents. We can use the excess coal that was left ungifted to further power the furnace, which will further improve the cost efficiency of our design.