121st ASEE Annual Conference & Exposition
Indianapolis, IN
June 15-18, 2014
Paper ID #9168

The Development of a Rubric to Evaluate and Promote Students' Integration of Stakeholder Considerations into the Engineering Design Process

Dr. Alexandra Emelina Coso, Georgia Institute of Technology

Alexandra Coso is a Postdoctoral Fellow at Georgia Tech's Center for the Enhancement of Teaching and Learning. She recently completed her Ph.D. in Aerospace Engineering at Georgia Tech. Prior to her time at Georgia Tech, she received her B.S. in Aerospace Engineering from MIT and her M.S. in Systems Engineering from the University of Virginia. Her research interests include the integration of stakeholders into the engineering design process, development and evaluation of interdisciplinary engineering courses and programs, mixed methods research designs, and graduate student experiences in engineering programs.

Dr. Amy Pritchett, Georgia Institute of Technology

The Development of a Rubric to Evaluate and Promote Students' Integration of Stakeholder Considerations into the Engineering Design Process

Abstract

Approaches exist for assessing student performance in some activities during the design of complex systems such as aircraft. However, resources are limited for assessing students' abilities to consider design from a broad perspective and to account for a design's impact on its stakeholders. The purpose of this paper is to introduce a rubric to assess how students perceive and integrate stakeholders into the design of a complex system. Following a description of the rubric and its development, this paper describes results from the initial application and evaluation of the rubric by a panel of faculty, graduate students, and research scientists, as they used the rubric to assess aircraft design projects. This initial evaluation demonstrated the strengths of the rubric (particularly with regards to validity) and how the reliability of the ratings among raters was sensitive to the few, discrete scores that the initial version of the rubric allowed; potential improvements to improve reliability are noted. The paper concludes with discussion of how the rubric can be used by design instructors as both a formative assessment tool to identify and describe to students key aspects of the design process necessary to account for stakeholder considerations, and as a summative assessment tool.

Introduction

With the current and future challenges facing society in infrastructure, renewable energy development, and medicine, and a general trend for designs to increase in complexity, there is a movement within the engineering design education community to introduce design approaches and experiences that place a higher value on the needs and limitations of stakeholders[1-5]. At the undergraduate level, these approaches and experiences may allow students to interact with stakeholders throughout the design process and develop a design solution that, in many cases, can be delivered to the client for future use[2-5]. Through these experiences, students can gain valuable insight into how designers can incorporate stakeholder considerations. However, in the design of complex systems, such as aircraft, the extent to which students have the opportunity to experience the entire design process and interact regularly with stakeholders is limited. For example, many aerospace design courses, due to time, financial, and equipment constraints, may establish student projects that focus on the conceptual design process, but do not include design activities like prototyping and building. Thus, students may not experience the design stage that provides for direct assessment of the design's ability to benefit stakeholders. In addition, resources are limited for instructors to assess students' abilities to consider design from a broad perspective and to account for the impacts of the design on its stakeholders.

Thus, this paper describes the development of a rubric to assess how students perceive and integrate stakeholders into the design of a complex system. The rubric is grounded within the engineering design, human factors, and human-centered design literature and is based on conditions that can be observed from student design projects and presentations. The paper begins with a discussion of related frameworks and rubrics along with the specific studies that informed the development of this rubric. Then, the Stakeholders in Design rubric is introduced, followed by the process for initially evaluating the validity, reliability, and overall usability of the rubric. These results were examined through the initial application of the rubric by a panel of faculty, graduate students, and research scientists who used the rubric to assess design projects from an aircraft design course. This initial application demonstrated strengths in the rubric's design, although the overall reliability of the rubric was affected as scores varied little for some teams and significantly for other teams. The paper then concludes with a discussion of future work to further improve the reliability of the rubric and to support design instructors' use of the rubric as both a formative assessment tool to identify and describe to students key aspects of the design process necessary to account for stakeholder considerations, and as a summative assessment tool.

Background on Rubrics and Rubric Development

Rubrics provide a systematic methodology for judging the quality of student work based on criteria for different performance measures. A rubric is described in the educational literature as a "simple assessment tool that describes levels of performance on a particular task and is used to assess outcomes in a variety of performance-based contexts from kindergarten through college (K-16) education"[6] (p. 131). These assessment tools are used across many disciplines to assess student reports and papers[7,8], students' team skills[9], oral presentations[7,10], and large-scale student projects[10-13]. In addition, many of these rubrics are specifically focused on design-related work and students' understanding of the design process[10, 12-14].

Beyond summative assessment tools for instructors, rubrics can also provide students with a system for peer and self-assessments[6, 8, 9, 14]. If distributed at the start of an assignment, students can use the rubrics to guide them as they complete the assignment. The rubrics, in this case, are used to clarify instructor expectations or as a form of pedagogical transparency[6, 8, 9, 11, 12]. Instructors can also view rubrics as mechanisms for evaluating the effectiveness of in-class activities and assignments[6, 8, 12]. As one researcher explains, "rubrics represent not only scoring tools but also, more important, instructional illuminators"[15] (p. 75).

Rubric Development

The development of a rubric follows an iterative process, as described in the education literature[6, 7, 9, 12, 15]. The process begins by defining the purpose and objectives of the particular assessment, which can be assessed through observations of students or examinations of student deliverables. Once the purpose and objectives have been defined, it is necessary to develop scoring criteria that address each objective. In the case of a rubric for evaluating stakeholder integration in a team design project, the scoring criteria should also align with the existing and relevant human-centered design, human factors, and engineering design frameworks. When considering which scoring criteria are appropriate, Popham reminds rubric developers "each of these criteria is eminently teachable" (1997, p. 75). With the scoring criteria defined, each criterion needs to be decomposed to clearly identify the qualities that describe the top and lowest levels of performance and an appropriate number of scoring levels. At this point, the developers should discuss the rubric length, as a balance is needed between the level of detail in the rubric and the time required for the assessment of each project. This balance is especially critical in large design courses, where project reports can range from 50 to 100 pages and the number of student teams can be substantial.

The next step is to select a scoring strategy, analytic or holistic, for the rubric. Analytic rubrics can be viewed as similar to a checklist [7] or a top-down methodology for assessment [13]. As such, each criterion is assigned a separate score by the rater, which may or may not be combined into a final score. Typically, these rubrics are used for task-specific evaluation and to help instructors and students isolate areas for improvement [6, 13, 14]. Some researchers view analytic rubrics as more objective methods of assessment [6, 12, 13]. Previous research has demonstrated that in open-ended cases such as the assessment of design tasks or projects, analytic rubrics can provide high levels of reliability while maintaining the validity of the rubric [6].

Holistic rubrics, on the other hand, assess multiple criteria within a single score. Research suggests the use of holistic scoring when evaluating criteria with significant overlap [7] or with scores requiring broader judgments of the quality of the work [6, 7]. In general, this scoring strategy restricts the ability to quickly isolate areas for feedback and improvement [14], but it can be less time consuming for instructors as compared with analytic rubrics [14]. One study described holistic rubrics as “bottom-up” approaches to scoring, based on identified groupings of previous students’ responses [13]. One possible shortcoming of holistic rubrics is rater bias, in part due to the necessity to make a broad judgment about a large-scale outcome, which can negatively impact reliability [6, 12].

Once the scoring strategy is determined, the final step is to consider how the overall score will be calculated. For instance, will the scores be summed or averaged or will specific weightings be associated with the different criteria? In the case of Watson and colleagues (2013), as an example, the researchers examined potential points for each criterion (based on the criterion’s applicability to the design project) and compared those points with the evidence of student incorporation of that criterion in their project.

Context for Rubric Development

The research site for this study was an 80 student senior aircraft design course within the aerospace engineering department at a large public, research institution focused largely on engineering. The aircraft design sequence is comprised of two courses, one in the fall and one in the spring. The purpose of the two senior design courses, as defined in the syllabus, is to give students experience with a conceptual design methodology that integrates methods for vehicle sizing, configuration selection and layout determination, propulsion system design, vehicle performance analysis, and cost analysis. During the fall semester, the course instructors present lectures and the students complete mini-projects, which introduce students to the design process and methods such as weight sizing and constraint sizing. Additionally, the students attended three lab sessions dedicated to understanding the role of stakeholders in the design of an aircraft and approaches for integrating stakeholder considerations into the design process. In the spring semester, students separate into teams to develop a solution to a Request-for-Proposals (RFP) selected by the course instructors. The instructors utilized the RFP developed by the American Institute of Aeronautics and Astronautics (AIAA) as the specification for the final projects and as a summative assessment tool. In general, the RFP implicitly incorporated stakeholders with respect to the noise and emissions requirements and explicitly incorporated stakeholders through an optional passenger time trade study.

Developing a Stakeholder in Design Rubric

To evaluate how student teams take into account stakeholder considerations within the design process and to promote student understanding of stakeholder considerations in future design courses, specific objectives were defined. The objectives of this rubric needed to be general enough that the resulting rubric could be easily applied to other engineering disciplines, but also specific enough to capture the intricacies of aerospace vehicle design. The resulting objectives define that the rubric should be able to:

1. Evaluate how students account for stakeholder considerations during the design process as represented within their design project deliverables;
2. Compare and contrast how students integrate stakeholder considerations within design decisions at different stages of the design process as represented within their design project deliverables;
3. Articulate observable outcomes in a manner that encourages students to apply and document good design processes; and
4. Be flexible, such that it can be easily applied to the conceptual design process of a variety of design projects.

To define the appropriate scoring criteria for these objectives, relevant literature in human-centered design and engineering design education was examined to uncover frameworks that describe how students may consider stakeholder considerations during the design process. A phenomenographic study of students' experiences designing for others mapped how students experience human-centered design (HCD) into two dimensions[16] (see Figure 1). The first dimension is based on how the students experience the design process and integration, ranging from a non-existent process to an empathic process[16]. The second dimension describes how students understand the user: this could describe a lack of appreciation of the user, for example, or a desire to involve the user in the design process[16]. The outcome space illustrates the breadth of the two dimensions and the intersections where students' experiences were found to inhabit[16]. Consider a student in spot #4, for example: this student sees design as an integrated and iterative process, views the user and the stakeholders as having particular needs, and examines those needs throughout the design process. The complexity of the experience increases as one moves from the lower left corner of the space to the upper right corner.

Two factors were found to affect students' placement in the outcome space: previous design experience and "internally motivated or externally motivated" user understanding [16] (p. 146). The results indicate that certain threshold concepts may exist that inhibit students from fully committing to human-centered design approaches, including lack of design skills and lack of appreciation of the user [16].

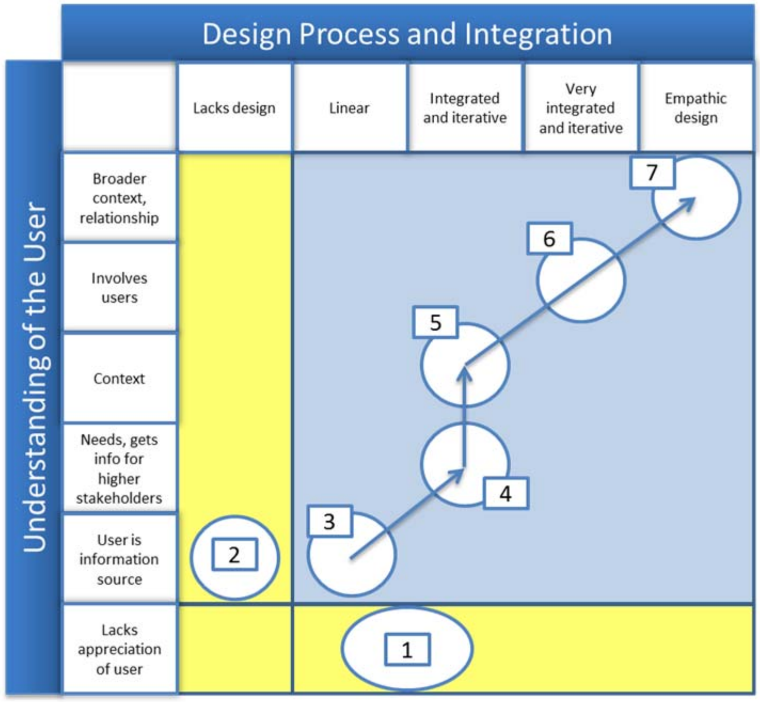

Figure 1: Students' Ways of Experiencing the Human-Centered Design Outcome Space with Seven Common Categories [16]

Researchers have examined ways to operationalize this outcome space into a rubric or learning taxonomy[2, 17, 18]. One group of researchers utilized the outcome space to evaluate students' responses to a discipline-neutral, in-class assessment aimed at assessing students' understanding of HCD[17]. The group was successfully able to categorize students' responses based on the seven categories in the outcome space labeled in Figure 1[17]. The assessments used in that study were smaller-scale, user-centered design problems, rather than the complex system design problems being examined with this work. Still, their success suggests that the two dimensions, *Understanding the User* and *Design Process and Integration,* can serve as a starting point for the Stakeholder in Design scoring criteria.

Since the objectives of the Stakeholder in Design rubric are explicitly focused on the integration of stakeholder considerations, as opposed to how students involved users or got information from users, the first scoring criterion will be called *Stakeholder Integration* and its scale and descriptors are adapted from the *Understanding the User* scale. Based on the finding that design skills may be a threshold concept for students, the second scoring criterion focuses on students' *Design Understanding* and is adapted from the *Design Process and Integration Scale* as well as other frameworks for design understanding. In addition, due to the lack of literature on the relationship between design understanding and stakeholder integration within complex system design, the scoring criteria for each of the two dimensions are kept separate within this rubric. Future work will include a comparison of the relationships found through the use of the rubric and the relationship defined within the outcome space previously discussed.

Scoring Scale and Descriptions

In considering the appropriate scoring scale and descriptions based on the criteria and the defined objectives, we examined the literature on human-centered design, human factors, and engineering design to capture the characteristics of approaches and designs that integrate stakeholder considerations successfully. Additionally, we explored how stakeholders can be considered at different phases of the design process.

From the human-centered design (HCD) literature, Maguire (2001) describes key principles of HCD, which include (1) the active involvement of users, (2) clear understanding and specification of the user and task requirements, and (3) iteration of design solutions. These principles are critical in completing the processes and achieving the standards defined in ISO 13407 (a standard on HCD)[19]. The process for performing HCD, as defined in this ISO, includes the following steps:

(1) ***Planning*** **–** Successful HCD brings together the stakeholders of the project to determine how HCD can contribute to the overall goals of the project and how it will be integrated into the overall design process.
(2) ***Understanding and Specifying the Context of Use*** **–** This stage includes identifying the stakeholders and the context: What are the required objectives and tasks associated with the design? To gather sufficient data, designers may observe users in their work environment, perform task analysis, administer surveys, or run focus groups.
(3) ***Specifying Requirements*** **–** This stage identifies what expectations or requirements the design must meet. To define the requirements, designers complete a more comprehensive and sophisticated stakeholder analysis (defining the main roles, goals, and responsibilities of the different stakeholders).
(4) ***Creating Design Solutions (and Prototypes)*** **–** This stage includes the iterative development of low- to high-fidelity prototypes along with the final design solution.
(5) ***Evaluating the Designs*** **–** This stage focuses on the user-based testing, whether usability or ergonomic or other, to ensure the design meets the needs of the user and supports user performance and safety[19,20].

Within the human factors literature, Chua and Feigh (2011) reviewed methods for incorporating human factors considerations into the different stages of the system design process. For the requirements and problem scoping stage of the design process, they recommended establishing agreement on top-level requirements and expectations between the customer and the design team, through an analysis of the context in which the system may operate once designed[21]. The techniques used at this stage are similar to those used by the HCD designers in step 2, as previously mentioned, such as surveys, interviews, and focus groups[21]. Chua and Feigh explained that "the earlier the design team develops an appreciation for user requirements, the more effective final design will be"[21] (p. 2). Their second stage is concept generation, where designers should consider what design trade-offs are necessary due to the stakeholder considerations[21]. Finally, in the preliminary design stage, designers should evaluate candidate designs to verify that they meet stakeholder-related requirements[21]. During this evaluation, additional trade-offs may need to be examined and considerations such as training and personnel implications of the design need to be explored[21].

Within the aerospace engineering design literature, previous studies have examined stakeholder integration within aerospace engineering practice and design education[22,23]. From these studies, stakeholders are considered in some fashion within each stage of the conceptual design process. Yet, overall, stakeholder integration is dependent on the initial requirements and whether the considerations are quantifiable[22,23]. In addition, beyond the stages of design described previously, some design textbooks include implicit discussions of stakeholder considerations in other sub-design stages, such as the technology integration stage.

The rubric is broken down into three parts. The objectives of the Stakeholder in Design rubric are explicitly focused on the integration of stakeholder considerations, as opposed to how students involved users or got information from users. Thus, the first scoring criterion is called *Stakeholder Integration* and its scale and descriptors are adapted from the *Understanding the User* scale. Based on the finding that design skills may be a threshold concept for students, the second scoring criterion focuses on students' *Design Understanding* and is adapted from the *Design Process and Integration Scale* as well as other frameworks for design understanding.

Based on these studies and those described above, the *Stakeholder Integration* scale, as previously mentioned, (see Table 1) was adapted from the Ways of Experiencing Human-Centered framework described earlier[16]. Not all of the categories within the original framework were included, reflecting the attributes of complex system design in the projects of interest here. For instance, the high levels of performance on the Stakeholders in Design rubric adapt some of the constructs of the higher levels on the *Understanding the User* scale to encourage students to understand and leverage stakeholder considerations without necessarily requiring that they have interactions with stakeholders. In addition, the level related to *Context* was removed to avoid confounding student understanding of stakeholder considerations with student understanding of other contextual considerations. Overall, the scoring scale was created to meet the first and fourth objectives of the rubric, i.e., allowing for the evaluation of how students consider stakeholders and incorporating flexibility. The scale is neither discipline-specific nor project-specific.

Table 1: Stakeholder Integration Scale

Score	Stakeholder Integration Scale	Description
0	*Lacks appreciation for stakeholders*	Lacks integration of contextual and stakeholder considerations.
1	*Considers stakeholders implicitly*	Includes stakeholder-related considerations, but does not explicitly discuss the human aspect of the consideration. (e.g., noise and people living near the airport, cost, and the customer, etc.)
2	*Incorporates stakeholder considerations at isolated points*	Identifies stakeholder considerations at isolated points in the design process, but overall these considerations are not the basis for design decisions and are not addressed consistently through design
3	*Integrates stakeholder considerations consistently throughout the design process*	Exhibits a commitment to incorporating stakeholder considerations throughout the design process. Design decisions clearly account for their potential impact on stakeholders and related research or communication with experts/stakeholders.
4	*Leverages the multiple perspectives of stakeholders to pursue a more innovative, competitive design*	Demonstrates how multiple perspectives (and at times, competing requirements) were integrated to develop their final solution. Overall solution can be marketed as a design driven by stakeholder, context, and performance requirements.

To develop a scale for design understanding, the *Design Process and Integration* scale, described previously, was adapted using terminology from Bloom's Taxonomy [24] and constructs from the Informed Design Teaching and Learning Matrix [25] (see Table 2). More specifically, the Application scores (1 and 2) capture when students apply the design process as presented within their design curriculum. The higher Abstraction scores (3 and 4) are synonymous with the "Creating" construct in Bloom's taxonomy [24]. These scores describe when students abstract the elements of the design process to develop a more innovative solution. This abstraction allows students to form a new design process from the elements of the design process presented in their course to match the particular context of the problem. To differentiate among the levels, the beginning designer and informed designer descriptions from the Informed Design Teaching and Learning Matrix were utilized. For example, the Abstraction II level captures the *Managed and Iterative Designing* behavior used by the informed designers [25]. As with the *Stakeholder Integration* scale, the *Design Understanding* scale is flexible, such that it can be used for a variety of projects.

Table 2: Design Understanding Scale

Score	Design Understanding Scale	Description
0	*Design Process Knowledge is Absent*	Lacks a basic understanding of the design process, as exhibited by missed steps or an incorrect application of design process knowledge to the problem
1	*Application I*: Design Process is Linear	Reproduces the design process as presented within the course, but does not recognize when findings later in the design process warrant revision of earlier steps
2	*Application II*: Design Process is Integrated and Iterative	Applies the design process as presented within the course and demonstrates how the final design concept improved iteratively via feedback and additional analysis
3	*Abstraction I*: Very Integrated and Iterative	Exhibits deep understanding of the design process. Abstracts and augments the principles of the design process to reach an innovative solution to the problem through an iterative process (e.g., may develop a new tool for modeling or analysis).
4	*Abstraction II*: Reflective Designer	Exhibits a deep understanding of design. Design process is very iterative and reflective. Frequently re-evaluates ideas relative to new knowledge. Notes any limitations in the design process or modeling tools and the impact of those limitations on the final design concept.

To meet the second and third objective and to provide more information about how students considered stakeholders, the rubric includes a section that addresses stakeholder integration within each design stage (see Appendix). This section of the rubric acknowledges that students may consider stakeholders differently during the different stages of the design process. As such, the rubric divides the design process into three stages (i.e., Requirements/Problem Definition, Concept Generation/Development, and Technology Integration), which can be modified depending upon the design project, and a fourth category for the overall design.

Each design stage is examined more closely using questions about the incorporation of stakeholder considerations at that stage,

1. Did the student(s) state an intention to incorporate stakeholder concerns at this phase?
2. Did the student(s) apply a design process at this stage that could include stakeholder concerns?
3. Was the student(s) successful in integrating stakeholder concerns?

These questions examine the intent of students to incorporate stakeholder considerations relative to whether the students applied design methods that could support their intentions. The connection between statements of intent and subsequent behavior has been examined within the literature and provided a foundation for examining the possible connection between students' stated intentions and their success in integrating stakeholder concerns [26]. In addition, these questions support the examination of the design project report or presentation holistically by allowing the instructor or student to look at each stage of design first and then at the whole design solution.

Scoring Strategy

This final rubric is holistic with some analytic components. The scores for the *Stakeholder Integration* (see Table 1) and *Design Understanding* (see Table 2) scales require the instructors or students to make broad, holistic judgments about the student teams' overall performance on the project. As with the Ways of Experiencing Human-Centered design framework, the *Stakeholder Integration* and *Design Understanding* scores should be viewed in pairs, reflecting the two dimensions of the framework. This reflection can help identify whether students' ultimate integration of stakeholders may have been more driven by their intention to integrate stakeholders or by their design understanding. The design phase questions serve as the analytic component of the rubric (see Appendix), permitting an evaluation of specific design stages and allowing for the determination of possible areas of improvement. The scoring varies by question. For Questions #1 and #2, the student team earns 0 points for a "No" response or 1 point for a "Yes" response. For Question #3, the scoring is as follows: 0 points – No, the student(s) was not at all successful, 1 point – Yes, but in a superficial manner, and 2 points – Yes, and in integral manner. Further, each space for scoring on the rubric itself allows the instructor or student to provide any specific evidence that served as the basis for a given score.

The decision to pursue a more holistic scoring strategy was due in part to the necessity to balance rubric length with the time required for the assessment of a single project. As previously stated, project reports for this context range from 50 to 100 pages and instructors may have several to assess each semester. In addition, this rubric can be used alongside the project specifications (i.e., RFP) in the assessment of student reports and presentations. Thus, the rubric was restricted to one-page in length.

Initial Evaluation of Rubric

The Stakeholder in Design rubric was developed to assess how students integrate stakeholder considerations into the design of a complex system. To evaluate the ability for the rubric to measure what it is intended to measure and to produce consistent results, we examined the validity, clarity, and reliability of the rubric.

Validity

Literature on rubric development specifies three types of validity that should be examined. The first, content validity**,** defines "the extent to which a student's responses to a given assessment instrument reflects that student's knowledge of the content area that is of interest"[27] (p. 2). Content validity is most commonly explored using Subject-Matter Experts (SMEs)[6,12,28]. Construct validity considers the relationship between what is being evaluated by the rubric and the criteria being used to evaluate it[28]: Does the rubric measure what it is supposed to measure? Finally, criterion validity defines the predictability of the measurements with current or future performance[27,28]. For instance, can performance on this report, as defined by the rubric, be generalized to future performance in industry or future design projects?

Clarity and Reliability

During the rubric development process, it is necessary to consider how to promote consistency among raters with scoring criteria descriptions and overall rubric instructions[27,28]. This consistency is examined using measures for inter-rater reliability, intra-rater reliability, and clarity[27]. Inter-rater reliability measures the variability among the raters, while intra-rater reliability measures variability among a single rater[27]. There are many statistical measurements that can be used to examine inter-rater reliability, from consensus agreements to consistency estimates[6]. The decision of which measurement to use is dependent on the number of raters and ratings as well as whether each rater rated all of the sample or only a partial sample. Threats to intra-rater reliability include rater fatigue or rater bias (e.g., if the rater is knowledgeable that they are rating someone who may fail a class if they do poorly on this assessment)[27]. Statistical measures exist for intra-rater reliability, but again, are dependent on the research design[6].

Beyond statistical measures for examining reliability, the clarity of a rubric can also be evaluated to understand the reliability of the scoring criteria. This evaluation includes asking questions to raters, such as “are these scoring categories well defined?”, and “are the differences between the score categories clear?”[27]. Finally, when considering the implementation of the rubric, researchers have defined methods for improving consistency among the raters, including

- The use of anchor papers, which provide an example how the rubric is used[27];
- The use of a sample set of responses for raters to evaluate, which provides information about discrepancies among the raters[27]; and
- The use of sample responses that represent the top level of performance[12].

Methods

During the rubric development process, specific measures were taken to improve the validity of the rubric itself. Specifically, the process incorporated the use of specific objectives to guide the selection of the scoring criteria, scale and the descriptions for construct validity. Each level of the scoring criteria was created to align with relevant literature that considered how stakeholders could be integrated into a design both in a higher education setting and in an industrial setting to support construct and criterion validity.

To further examine the validity, clarity, and reliability of the rubric, a group of seven subject-matter experts (SMEs) were recruited to assess student performance on their senior design reports using the rubric and to evaluate the rubric from an instructor perspective. The group represents researchers and engineering educators in varying sub-disciplines of aerospace engineering, including conceptual aircraft design and cognitive engineering.

Each of the SMEs was provided a rubric packet with an explanation of the rubric objectives, the scales, and the scoring method. In addition, the packet included a rubric design questionnaire (adapted from the work of Moskal and Leyden, 2000, and Stevens and Levi, 2005). The questionnaire examines the clarity of the scoring criteria, the descriptions, the scale, and the overall rubric, along with content validity, construct validity, and criterion validity[27,29] (see the sample questions in Table 3 and the complete questionnaire in the Appendix).

Table 3: Sample Questions from the Rubric Development Questionnaire[27,29]

	Rubric Evaluation Questions
Evaluation Criteria	Are the evaluation criteria distinctly different from each other?
Overall Rubric	Does it balance the necessary level of detail with the time required for the assessment of a single project?
Content-Related Validity	Does this rubric evaluate how students consider stakeholders in the design process?
Construct-Related Validity	Are all the important components of the integration of stakeholders into the design process evaluated in the rubric?

Prior to utilizing the rubric, each SME was required to receive one-on-one training on the rubric. This training was provided to support consistency and overall understanding of the rubric among the SMEs[12,28]. The SMEs were introduced to the rubric and the individual scales and were encouraged not to consider the rating process as a "grading" process. In other words, a well-done design project does not necessarily mean the team must receive a "4" on the *Stakeholder Integration* scale or even the *Design Understanding* scale. The same was true for a poor design project. Additionally, it is possible for a team that performs poorly in one scale to perform well in another.

Following training, the SMEs were asked to read and assess a series of senior aircraft design capstone projects from the site for this research study. Within the course, ten teams of students, ranging in size from 6 to 9 students, developed a conceptual design for an aircraft based on an RFP developed by the AIAA. Each team was required to submit a report at the end of the semester, documenting their design solution and their approach for developing that solution. The reports were used for this rubric evaluation with approval by the Institutional Review Board.

Due to the size of each design project (approximately 90 to 100 pages), six of the seven SMEs were assigned four of the reports, while the seventh SME read all ten reports. This approach for assigning the reports allowed for all of the reports to be rated by at least three SMEs. To examine the variability among the raters more closely, one of the reports was read by all of the SMEs. Finally, each report assignment was organized such that the SMEs did not read the same projects in the same order. The distribution of reports is illustrated in Table 4.

Table 4: Assignment of Reports to SMEs [NOTE: All identifying information was removed from each report prior its distribution]

Team #	SME 1	SME 2	SME 3	SME 4	SME 5	SME 6	SME 7
A	1	1			1		
B	1		1			1	
C	1			1			1
D	1	1			1		
E	1		1				1
F	1			1		1	
G	1	1	1	1	1	1	1
H	1	1					1
I	1		1		1		
J	1			1		1	

Based on the small number of reports read by each SME, traditional statistical measures for reliability (e.g., Cohen's kappa, Krippendorff's alpha, interclass correlation coefficient) could not be applied in this evaluation. A more in-depth review of the rubric's reliability using a larger sample of reports per SME will be completed in a future study. At this stage, the variability of among the SMEs is examined using the Rubric Development Questionnaire questions on scoring criteria, descriptive statistics, and a qualitative examination of SME's comments about each team.

Results

Validity and Clarity

Overall, six of the seven SMEs believed the evaluation criteria were clear and distinctly different from each other without addressing any extraneous content. Yet, within the *Stakeholder Integration by Design Stage* section, it appeared that four of the SMEs did not agree with the use of three design stages and the fourth "overall design" category. In some cases, these stages were seen as not aligned with the project requirements, which had not emphasized the problem definition stage and which had required students to perform other analyses not included explicitly in these design stages.

In regards to the descriptions, all of the SMEs indicated that the descriptions matched the evaluation criteria and were clear and different from one another. All of the SMEs felt that the rubric could be understood by external audiences and balanced the necessary level of detail with the time required for the assessment of a single project. In addition, they all commented that they could see themselves using it in their classroom. Still, some of the SMEs believed they misinterpreted the "Did the student(s) apply a design process at this stage that could include stakeholder concerns?" question, which may have impacted their scoring of the projects.

In terms of content validity, the SMEs described that the rubric successfully evaluates how students consider stakeholders in the design project and could be used for different projects within and outside of aerospace engineering. However, a couple of the SMEs raised the following question: *How can the rubric account for teams that considered one stakeholder thoroughly versus one that considered a lot of stakeholders superficially?* Finally, one SME noted that the *Design Understanding* scale could be more thorough, while another explained how it may be an unnecessary scale if instructors utilize their own rubric for evaluating students' design understanding.

For the questions about construct validity, the SMEs reported that all the important components of the integration of stakeholders into the design process were evaluated in the rubric. Nevertheless, as defined previously, some of those components could be divided differently or expanded upon slightly. In the discussions of criterion validity, five of the seven SMEs believed that students' performance on this report could be generalized to their future performance. However, in general, the SMEs did not believe the generalization would necessarily be accurate at the individual level or for all of the scores. For instance, one SME explained how an excellent score would indicate students' ability to incorporate stakeholder requirements within an iterative and integrated design process, but a poor score could be attributed to other influences besides students' ability to incorporate stakeholders.

Finally, two SMEs discussed the impact of the project requirements (i.e., RFP) on students' *Stakeholder Integration* scores. While some teams may have acknowledged stakeholders as important, they may not have integrated stakeholder considerations because they were adhering to the pre-defined project requirements. In addition, differing viewpoints among the SMEs emerged on what it means to consider stakeholders during the project evaluations. Even though some teams incorporated stakeholder considerations inherently in the design (e.g., cabin sizing, cost, noise), some SMEs did not view this as evidence these student teams had valued stakeholders, but instead these teams had adopted some measures suggested by the RFP or instructors that happened to relate to stakeholders.

Reliability

The resulting scores for each of the scales are presented in Table 5 and Table 6. The SME's scores varied little for some teams and significantly across the scores given other teams. With both scales, only 40% of the ratings per team were within one performance level.

One possible reason behind these results could be the SME's scoring strategy and the few ratings options. When examining the scoring criteria scale, two of the SMEs responded in the questionnaire that they would have preferred the option for half points, while one of the SMEs scored the projects using half points. A review of the SMEs' comments on the team rating sheets confirmed that allowing for half points might have improved the overall reliability. Five of the seven SMEs explicitly discussed indecision between scoring levels on their rating sheets.

Table 5: Stakeholder Integration Scores by SME

Team #	SME 1	SME 2	SME 3	SME 4	SME 5	SME 6	SME 7	Range
A	0	2			1			2
B	2		2			2		0
C	3			2			3	1
D	0	2			2			2
E	0		0				0	0
F	0			2		2		2
G	1	2	2	0	1	1	2	2
H	1.5	2					2	.5
I	2		0		3			3
J	0.5			3		2		2.5

Table 6: Design Understanding Scores by SME

Team #	SME 1	SME 2	SME 3	SME 4	SME 5	SME 6	SME 7	Range
A	3	2			2			1
B	1.5		1			1		0.5
C	3			1			3	2
D	2.5	2			2			.5
E	3		2				1	2
F	1			3		1		2
G	1.5	1	1	0	3	1	1	3
H	2	2					1	1
I	2		1		3			2
J	2.5			4		2		1.5

The scores within the *Stakeholder Integration by Design Stage* reflected the same variability among SME scores as the two scales (see Table 7). With these results, many of the SMEs disagreed completely as to each team's intention to incorporate stakeholder considerations and overall success in incorporating those considerations. Similarly to the two scales, three SMEs requested a wider range for scoring the *Stakeholder Integration by Design Stage* section (for instance, 0 to 4 points).

Table 7: Example Ratings from the Stakeholder Integration by Design Stage

Team J	SME #1	SME #4	SME #6		Team D	SME #1	SME #2	SME #5
Requirements/ Problem Definition					**Requirements/ Problem Definition**			
Intention	0	1	1		*Intention*	1	1	1
Application	0.5	1	0		*Application*	1	1	0
Success	0	2	1		*Success*	1	2	1
Concept Generation/ Development					**Concept Generation/ Development**			
Intention	1	1	0		*Intention*	0	1	1
Application	0	1	0		*Application*	0	1	0
Success	0.5	2	1		*Success*	0	0	1
Technology Integration					**Technology Integration**			
Intention	0	1	0		*Intention*	0	1	1
Application	0	1	0		*Application*	1	1	0
Success	0	1	0		*Success*	0	1	1
Overall Design					**Overall Design**			
Intention	0.5	1	0		*Intention*	0	1	1
Application	1	1	0		*Application*	0.5	1	0
Success	1	2	1		*Success*	0	1	1

Discussion of Rubric Implementation & Future Work

This initial evaluation of the Stakeholder in Design rubric as a summative assessment tool highlights both benefits and possible areas of improvement within the descriptions and scoring strategy. This section examines these areas of improvement further and considers how this rubric could be implemented within an engineering design learning environment as both a formative and a summative assessment tool.

To improve the inter-rater reliability of the rubric, the literature suggests considering how the variation in the data may relate to the descriptions and scoring strategy[28]. In this case, the SMEs' varied definitions of what constituted implicit stakeholder integration could have negatively affected the consistency of the scores. For instance, all of the projects were required to include an evaluation of aircraft emissions and noise, which are implicit stakeholder considerations. Some of the SMEs perceived this requirement as students implicitly considering stakeholders; other SMEs perceived that the students didn't consider stakeholders in this case unless they considered stakeholders beyond the scope of the RFP or explicitly recognized that these requirements mapped to specific stakeholders. As a future intervention, questions could be added to the rubric typifying how (ideally) students should consider stakeholders at each stage of the design process.

Another option for improving inter-rater reliability is to include anchor papers with the rubric and to discuss those anchor papers during rubric training. For instance, these papers would help instructors distinguish between the actions that define category 3 and category 4 in the *Stakeholder Integration* scale or isolate the actions that demonstrate successful integration of stakeholder considerations. With this initial sample of design projects, the anchor papers could be executive summaries from these design projects that highlight high and low performing teams.

The intent of the rubric implementation as a summative assessment tool is to have the rubric supplement any current rubrics or scoring schemes addressing other aspects of design. The holistic nature of both scales allows instructors to step through the Stakeholder in Design rubric quickly.

While this initial evaluation only applied the rubric as a summative assessment tool, the rubric can and should be used as a formative assessment tool as well. For example, students can be provided with a copy of the rubric at the same time as the project's RFP to emphasize the expectation that integrating stakeholders is an important part of the design process. The use of the rubric as a formative assessment tool would also serve to improve the validity of the rubric, as the rubric would help define the task requirements (or in this case, design project requirements) and thus address situations where the task requirements relative to stakeholder integration might not be clear to the students. For the design projects examined in this study, the design project requirements included stakeholder considerations implicitly, which likely deterred some students from considering stakeholders more explicitly during the design process.

The successful integration of stakeholder considerations into a complex system design requires an appropriate balance between performance and stakeholder considerations to avoid overdesigning for any given consideration. The holistic nature of the rubric permits its use during oral design reviews throughout the semester. This repeated use of the rubric would allow instructors to support students' understanding of how to balance these critical design considerations throughout the design process. Overall, the use of the rubric could strengthen students' perception of the importance of considering stakeholder considerations and, as such, could result in a broader perspective by students on the role of stakeholder considerations in complex system design.

Conclusion

To support the integration of stakeholder considerations into complex system designs, a rubric was developed, grounded within the engineering design, human factors, and human-centered design literature, to serve as both a formative and summative assessment tool by instructors and students. Following its development, the Stakeholder in Design rubric was evaluated by Subject-Matter Experts as they used the rubric to assess the incorporation of stakeholder considerations in a series of senior aircraft design project reports. This initial evaluation demonstrated strengths of the rubric (particularly with regards to validity), although the reliability of the ratings among raters was sensitive to the few, discrete scores that the initial version of the rubric allowed. Additional efforts over the coming year, based on a larger sample of design projects and an updated rubric, will support the improvement of the reliability and usability of the rubric for future use as a tool for the engineering design learning environment to encourage students to consider the impact of complex system design decisions on stakeholders.

Acknowledgments

The authors would like to thank all of the Subject-Matter Experts who volunteered their time to support this research and the reviewers for their insightful recommendations. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-0644493.

References

1. F. Marbouti and H. A. Diefes-Dux, "First-year students' understanding of direct user in open-ended problem solving activities," in *Frontiers in Education Conference*. Oklahoma City, OK: IEEE, October 2013, pp. 320–326.
2. S. Jordan and M. Lande, "Practicing needs-based, human-centered design for electrical engineering project course innovation," in *ASEE Annual Conference and Exposition*, San Antonio, TX, June 2012.
3. L. Oehlberg and A. M. Agogino, "Undergraduate conceptions of the engineering design process: Assessing the impact of a human-centered design course," in *ASEE Annual Conference and Exposition*, Vancouver, BC, June 2011.
4. C. B. Zoltowski, W. Oakes, and S. Chenoweth, "Teaching human-centered design with service learning," in *ASEE Annual Conference and Exposition*, Louisville, KY, June 2010.
5. C. Titus, C. B. Zoltowski, and W. C. Oakes, "Designing in a social context: Situating design in a human-centered, social world," in *ASEE Annual Conference and Exposition*, Vancouver, BC, June 2011.

6. A. Jonsson and G. Svingby, “The use of scoring rubrics: reliability, validity and educational consequences,” *Educational Research Review*, vol. 2, no. 2, pp. 130–144, Jan. 2007. [Online]. Available: http://-linkinghub.elsevier.com/retrieve/pii/S1747938X07000188
7. B. M. Moskal, “Scoring rubric: what, when, and how?” *Practical Assessment, Research, & Evaluation*, vol. 7, no. 3, pp. 1–7, 2000. [Online]. Available: http://PAREonline.net/getvn.asp?v=7&n=3
8. R. S. Kellogg, J. A. Mann, and A. Dieterich, “Developing and using rubrics to evaluate subjective engineering laboratory and design reports,” in *American Society for Engineering Education Annual Conference and Exposition*, Albuquerque, NM, June 2001.
9. C. Plumb and D. Sobek, “Measuring student ability to work on multi-disciplinary teams: Building and testing a rubric,” in *American Society for Engineering Education Annual Conference and Exposition*, Honolulu, HI, June 2007.
10. D. Jones and A. Tadros, “Successful use of rubrics to assess student performance in capstone projects,” in *American Society for Engineering Education Annual Conference and Exposition*, Louisville, KY, June 2010.
11. A. Chong and L. Romkey, “Evolving a rubric for use in assessing engineering graduate attributes in a student senior research,” in *American Society for Engineering Education Annual Conference and Exposition*, San Antonio, TX, June 2012.
12. M. Watson, C. Noyes, and M. Rodgers, “Student perceptions of sustainability education in civil and environmental engineering at the Georgia Institute of Technology,” *Journal of Professional Issues in Engineering Education & Practice*, vol. 139, no. 3, pp. 235 – 243, 2013.
13. R. Bailey, Z. Szabo, and D. Sabers, “Assessing student learning about engineering design in project-based courses,” in *American Society for Engineering Education Annual Conference and Exposition*, Salt Lake City, UT, June 2004.
14. A. Kline, E. Tsang, B. Aller, J. Asumadu, J. Morgan, S. Beyerlein, and D. Davis, “Creating and using a performance measure for the engineering design process,” in *American Society for Engineering Education Annual Conference and Exposition*, Nashville, TN, June 2003.
15. W. J. Popham, “What’s wrong - and what’s right - with rubrics,” *Educational Leadership*, vol. 55, pp. 72–75, October 1997.
16. C. B. Zoltowski, “Students’ ways of experiencing human-centered design,” Ph.D. dissertation, Purdue University, May 2010.
17. R. B. Melton, M. E. Cardella, W. C. Oakes, and C. B. Zoltowski, “Development of a design task to assess students’ understanding of human-centered design,” *42nd ASEE/IEEE Frontiers in Education Conference*, pp. 1–6, Oct. 2012. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6462469
18. R. B. Melton, C. B. Zoltowski, M. E. Cardella, and W. C. Oakes, “Work in progress - development of a design task to assess students’ understanding of human-centered design,” in *40th ASEE/IEEE Frontiers in Education Conference*, Washington, D.C., October 2010.
19. M. Maguire, “Methods to support human-centered design,” *International Journal of Human-Computer Studies*, vol. 55, no. 4, pp. 587–634, Oct. 2001. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/-S1071581901905038
20. G. Vos and N. Ezer. (2009, Dec) Introduction to human factors and the human centered design process. [PowerPoint slides and Video]. NASA’s Human Systems Engineering and Development Division. [Online]. Available: http://www.nasa.gov/centers/johnson/slsd/about/divisions/hefd/project/hsi-videos-hcd.html
21. Z. K. Chua and K. M. Feigh, “Integrating human factors principles into systems engineering,” in *IEEE: Digital Avionics and Systems Conference*, Seattle, WA, October 2011.
22. A. Coso and A. Pritchett, “The integration of stakeholder requirements within aerospace engineering design education,” in *American Society for Engineering Education Annual Conference and Exposition*, Atlanta, GA, June 2013.
23. ——, “Incorporating stakeholder considerations in the aircraft design process: A focus on aircraft design education,” in *AIAA Scitech: 52nd Aerospace Sciences Meeting*, National Harbor, MD, Jan 2014, pp. 1–9.
24. M. Forehand. (2012, October) Bloom’s taxonomy. University of Georgia - Emerging Perspectives on Learning, Teaching, and Technology. [Online]. Available: http://projects.coe.uga.edu/epltt/-index.php?title=Bloom%27s_Taxonomy
25. D. P. Crismond and R. S. Adams, “The Informed Design Teaching and Learning Matrix,” *Journal of Engineering Education*, vol. 101, no. 4, pp. 738–797, Oct 2012.
26. I. Ajzen, “The theory of planned behavior,” *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179–211, 1991.

27. B. M. Moskal and J. A. Leydens, “Scoring rubric development: Validity and reliability,” *Practical Assessment, Research, & Evaluation*, vol. 7, no. 10, 2000. [Online]. Available: http://PAREonline.net/getvn.asp?v=7&n=10
28. C. Plumb and D. Sobek, “Measuring student ability to work on multi-disciplinary teams: A method for determining validity and reliability of a rubric,” in *American Society for Engineering Education Annual Conference and Exposition*, Pittsburgh, PA, June 2008.
29. D. D. Stevens and A. Levi, *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning*. Sterling, VA: Stylus Publishing, LLC, 2005.

Appendix – Stakeholders in Design Rubric

Instructions: As you read through the project, please score each project by considering how stakeholders are integrated into each phase of the design process. Provide any specific evidence which served as the basis for your score.

	Stakeholder Integration by Design Phase			
	Requirements/ Problem Definition	**Concept Generation/ Development**	**Technology Integration**	**Overall Design**
Did the student(s) state an intention to incorporate stakeholder concerns at this phase? *[Yes – 1pt, No – 0pts]*				
Did the student(s) apply a design process at this stage that could include stakeholder concerns? *[Yes – 1pt, No – 0pts]*				
Was the student(s) successful in integrating stakeholder concerns? *[Yes, in an integral manner – 2pts, Yes, but in a superficial manner – 1pt., No, the student(s) was not successful – 0pts.]*				

Considering the students' work as a whole, use (1) the design understanding scale to rate how the team applied or abstracted the engineering design process and (2) the stakeholder integration scale to score how the team perceived and integrated stakeholders in the design of this complex system. The scales are defined and examples are provided on page two. Also please provide comments or evidence from the project to support the reasoning behind your score.

Design Understanding Score (0 to 4pts) ____________________________	**Stakeholder Integration Score (0 to 4pts)** ____________________________
Comments:	***Comments:***

Rubric Evaluation

	Rubric Evaluation Questions	Yes	No	Comments
Evaluation Criteria	Are the evaluation criteria clear?	__	__	
	Are the evaluation criteria distinctly different from each other?	__	__	
	Do the evaluation criteria address any extraneous content?	__	__	
	Do the evaluation criteria address all the aspects of the intended content?	__	__	
	Is there any content addressed in the task that should be evaluated through the rubric, but is not?	__	__	
Descriptions	Do the descriptions match the evaluation criteria?	__	__	
	Are the descriptions clear and different from each other?	__	__	
	Is there a clear basis for assigning the points for each evaluation criteria?	__	__	
Scale	Do the descriptors under each level truly represent that level of performance?	__	__	
	Does the rubric have a reasonable number of levels for the age of the student and the complexity of the assignment?	__	__	
	Would two independent raters arrive at the same score for a given response based on the scoring rubric?	__	__	
Overall Rubric	Does the rubric clearly connect to the outcomes that it is designed to measure?	__	__	
	Can the rubric be understood by external audiences (avoids jargon and technical language)?	__	__	
	Is the rubric of appropriate length?	__	__	
	Does it balance the necessary level of detail with the time required for the assessment of a single project?	__	__	
	Is this a rubric you could see yourself using in your classroom?	__	__	
Content – Related Validity	Does the rubric evaluate how students consider stakeholders in the design process?	__	__	
	Could this rubric be used for groups outside of AE?	__	__	
	Or for different project descriptions within AE?	__	__	
Construct – Related Validity	Are the elements of the responses being evaluated appropriate indicators of students' abilities to consider the stakeholder in design?	__	__	
	Are all the important components of the integration of stakeholders into the design process evaluated in the rubric?	__	__	
	Are any of the evaluation criteria irrelevant to the construct of interest?	__	__	
Criterion – Related Validity	Can the students' performance on this report be generalized to their future performance as new hires and in their careers?	__	__	
	Are all the important components of the students' future performance evaluated in the rubric?	__	__	
	Are there any components of the students' future or related performance that are not reflected in the scoring criteria?	__	__	