6-25-2017

Assessing the Effectiveness of Shah's Innovation Metrics for Measuring Innovative Design within a Virtual Design Space

Abigail Kulhanek

Matthew Markovetz

Golnaz Irgens
University of Wisconsin-Madison

Zachari Swiecki
University of Wisconsin-Madison

Naomi Chesler
University of Wisconsin-Madison

See next page for additional authors

Recommended Citation

Kulhanek, Abigail; Markovetz, Matthew; Irgens, Golnaz; Swiecki, Zachari; Chesler, Naomi; Shaffer, David; and Bodnar, Cheryl, "Assessing the Effectiveness of Shah's Innovation Metrics for Measuring Innovative Design within a Virtual Design Space" (2017). *Henry M. Rowan College of Engineering Faculty Scholarship*. 16.
https://rdw.rowan.edu/engineering_facpub/16

Authors

Abigail Kulhanek, Matthew Markovetz, Golnaz Irgens, Zachari Swiecki, Naomi Chesler, David Shaffer, and Cheryl Bodnar

Assessing the Effectiveness of Shah's Innovation Metrics for Measuring Innovative Design within a Virtual Design Space

Abigail Jane Kulhanek, University of Pittsburgh

Abigail Kulhanek is an undergraduate student studying Chemical Engineering at the University of Pittsburgh.

Dr. Matthew Raymond Markovetz, University of North Carolina, Chapel Hill

Matthew Markovetz is a postdoctoral researcher in the Cystic Fibrosis and Pulmonary Disease Research Center at the University of North Carolina at Chapel Hill. His interest in both engineering education and technical engineering research developed while studying Chemical and Biological Engineering at the University of Colorado at Boulder. Matthew's research in education focuses on methods that assess and increase innovation in product design, and his laboratory research seeks to understand and treat the airway dehydration present in patients with Cystic Fibrosis through mathematical modeling, rheological analysis, and systems engineering principles.

Golnaz Arastoopour Irgens, University of Wisconsin, Madison

Before becoming interested in education, Golnaz studied Mechanical Engineering at the University of Illinois at Urbana-Champaign with a minor in Spanish. While earning her Bachelor's degree in engineering, she worked as a computer science instructor at Campus Middle School for Girls in Urbana, IL. Along with a team of undergraduates, she headlined a project to develop a unique computer science curriculum for middle school students. She then earned her M.A. in mathematics education at Columbia University. Afterwards, she taught in the Chicago Public School system at Orr Academy High School (an AUSL school) for two years. Currently, Golnaz is working with the Epistemic Games Research Group at the University of Wisconsin-Madison where she has led the efforts on engineering virtual internship simulations for high school and first year undergraduate students. Golnaz's current research is focused on how games and simulations increase student engagement in STEM fields, how players learn engineering design in real-world and virtual professional environments, and how to assess engineering design thinking.

Mr. Zachari Lucius Swiecki, University of Wisconsin, Madison

Graduate student in educational psychology, learning sciences area.

Prof. Naomi C. Chesler, University of Wisconsin, Madison

Naomi C. Chesler is Professor of Biomedical Engineering with an affiliate appointment in Educational Psychology. Her research interests include vascular biomechanics, hemodynamics and cardiac function as well as the factors that motivate students to pursue and persist in engineering careers, with a focus on women and under-represented minorities.

Prof. David Williamson Shaffer, University of Wisconsin, Madison

David Williamson Shaffer is the Vilas Distinguished Professor of Learning Sciences at the University of Wisconsin, Madison, in the Department of Educational Psychology, and a Game Scientist at the Wisconsin Center for Education Research. His most recent book is How Computer Games Help Children Learn.

Dr. Cheryl A. Bodnar, Rowan University

Cheryl A. Bodnar, Ph.D., CTDP is an Assistant Professor in the Department of Experiential Engineering Education at Rowan University. Dr. Bodnar's research interests relate to the incorporation of active learning techniques in undergraduate classes as well as integration of innovation and entrepreneurship into the engineering curriculum. In particular, she is interested in the impact that these tools can have on student perception of the classroom environment, motivation and learning outcomes. She obtained her certification as a Training and Development Professional (CTDP) from the Canadian Society for Training and

Development (CSTD) in 2010, providing her with a solid background in instructional design, facilitation and evaluation. She was selected to participate in the National Academy of Engineering (NAE) Frontiers of Engineering Education Symposium in 2013 and awarded the American Society for Engineering Education Educational Research Methods Faculty Apprentice Award in 2014.

Assessing the Effectiveness of Shah's Innovation Metrics for Measuring Innovative Design within a Virtual Design Space

Abstract

Epistemic games, such as the virtual engineering internship Nephrotex, allow students to explore creative ways to approaching engineering problems while providing a novel alternative to the direct transmission method of instruction. Within Nephrotex, students choose a polymer, manufacturing process, surfactant, and percentage of carbon nanotube to create a functioning kidney dialysis membrane prototype. The performance of the membrane is measured using cost, flux, blood cell reactivity, marketability, and reliability thresholds given by stakeholders within the fictitious company. Although Nephrotex has been shown to be a valuable educational tool for modeling the product design process, only limited work has been done to investigate whether it is capable of providing an environment that allows students to generate innovative designs.

The innovation assessment framework of Shah and colleagues employs four metrics of innovation – novelty, variety, quality, and quantity; novelty is further divided into *a priori* and *a posteriori* metrics. This work found that *a priori* and *a posteriori* novelty, variety, and quality were applicable metrics of innovation in the epistemic game environment of Nephrotex. Literature ranges for *a priori* and *a posteriori* novelty scores aligned with those found in this study.

Comparing prior work on Nephrotex that identified innovative student designs based on a proposed literature definition, it was found that the Shah metrics between the innovative and non-innovative groups showed little variation and no statistically significant differences. A t test and a Mann-Whitney U test showed no significant difference between innovative and non-innovative groups with regard to variety or novelty scores; however, these tests did show a significant difference between groups with regards to the quality score. The same results were found when calculating Cohen's Effect Size – *a priori* novelty, *a posteriori* novelty, and variety had a small effect when comparisons were made between the innovative and non-innovative groups while quality had a large effect. The significant difference and large effect in regards to quality however, may be the result of the previous literature definition which employed quality as a measure to define innovation. Results from this study demonstrate that novelty is perhaps the most aligned innovation metric for an epistemic game environment and that both variety and quality can be helpful in understanding the designs generated within these contexts although they may need adjustment based on the application to a constrained design space.

Introduction

Upon completion of their degree, engineers should have the ability to design effective solutions to meet social needs and as such should be able to connect creativity with technical skills when approaching design problems [1]. The design process requires creativity and innovative thought. These qualities cannot be standardized which is why classes that prepare students just to succeed on exams are not the best route to enhance these skills [2]. Innovation is the act or process of introducing a new idea, device, or method that creates value [3]. Creativity is harder to define. Creative products are not universally judged as such by all experts, and creativity is different for different people across domains [4]. A consensual blueprint or formula for producing creative, innovative products does not exist, making creativity a difficult subject to teach students. However, there are aspects of creativity and innovation that can be measured in or taught to students such as attitudes about creativity and the process of product development [4].

Epistemic games are designed to be pedagogical tools for the digital age where players learn to think like professionals by playing a simulated game of their profession [5]. They have the ability to foster innovative thinking while teaching students about the process of product development. Epistemic games allow players to develop expertise by playing as novices training to be experts. Within this environment, novices receive explicit guidance from experts in the field helping them to acquire the knowledge, skills, values, and ways of thinking within their domain. Epistemic games lower the cost of failure by placing action in a simulated world and thus making it possible to learn innovation without the same level of risk that would exist in a work environment, to step into other cultural and intellectual settings in a guided and protected way, and thereby learn to think about issues that matter in a global economy by safely practicing in a virtual world [2].

Nephrotex is an epistemic game where students play as new hires for a fictitious medical device company. In the game, students design a dialyzer filtration membrane for a hemodialysis machine while trying to meet the requirements of the company's internal consultants by using a software platform that allows them to enter in their design parameters, receive feedback on how their design performs and then iterate on this design process [5]. At the start of the internship, students take an entrance interview, create a staff biography page, review internal documents about hemodialysis, filtration membranes and diffusion, and summarize this information in their online engineering notebooks. After conducting background research, interns examine fictionalized company research reports based on experimental data with a variety of polymeric materials, chemical surfactants, carbon nanotube percentages, and manufacturing processes. After collecting and summarizing research data, interns begin the design process. First individually and then in teams, students develop hypotheses based on their research, test these hypotheses in the provided design space, and analyze the provided results. At the end of the internship, students present their work to their colleagues and "supervisors" via a poster presentation. The students' supervisors are their professors and graduate teaching assistants. Nephrotex is a simulated design experience that uses a web-based PHP application and MYSQL database [5]. All activities are web-based, allowing students to access the game from any browser with internet capabilities.

It is important when working with a novel pedagogical tool designed to teach students about product design, such as Nephrotex, to determine whether it is able to support students' development of innovative designs. In a prior study, it was shown that a set of rules could be constructed to discriminate between innovative and non-innovative designs within Nephrotex. Designs that were unique in terms of material selection with quality greater than the section average OR that were the highest scoring non-unique material, but unique in process, design OR achieved a "perfect" quality score of 18 were considered innovative [6]. The study concluded that students with more innovative designs were students who were exposed to external stakeholders in a focus group setting and reported spending more time on management tasks [6]. Although the study by Markovetz *et al.* did provide some initial evidence to support the ability of Nephrotex to allow students to generate innovative designs [6], it was important to determine whether existing approaches to evaluate innovative design from pertinent design literature may also be applicable within the context of the Nephrotex environment.

Nephrotex challenges students with a complex design problem and gives them a space to explore multiple solutions and design hypotheses to determine which solution optimizes the conditions that have been presented to them. It allows students to learn about the product design process while reducing risks and allowing students to fail without detrimental consequences. Most design literature shows different ways of measuring innovative solutions for simple problems, not the complex design problems brought forth within Nephrotex. Effectiveness of ideation can be measured in two ways according to Shah *et al.* – process based where the idea generation process is evaluated or outcome based where the product is evaluated [7].

There have been some attempts to study ideation processes using process based models. Studies from an information processing point of view out of cognitive psychology have examined technological creativity and perception, specifically for design team activities. One example of this approach is when Ullman investigated the task-episode accumulation model (TEAM) of non-routine mechanical design through an analysis of audio and video protocols of five mechanical designers [8]. The components of this model were the design state, proposals, constraints, design operators, and episodes. The design state contains all information about the evolving design; proposals are alternative ways of achieving a goal; constraints are requirements for the design; operators are primitive information processes that modify the design state; and episodes are operators in a meaningful sequence. The TEAM includes ten operators – select, create, simulate, calculate, compare, accept, reject, suspend, patch, and refine – and six types of episodes – assimilate, document, plan, repair, specify, and verify – to create a picture of how mechanical design engineers of varying backgrounds and experience approach and solve design problems [8]. Another attempt to study ideation processes was the work done by Nagy, Ullman, and Dietterich who developed a data representation for collaborative mechanical design implemented in a computerized design history tool (DHT). This technique allows for the recording of design decisions, through collection of final specifications, alternatives considered during the design process, and designers' rationale as part of the database. This data representation was developed based on the Issue Based Information System method and aims to address problems in managing design information for collaborative design projects, including conflicting design requirements, a loss of rationale and design assumptions, and a thin spread of application domain knowledge. The DHT could replace traditional methods of recording mechanical design information – design drawings, plans, and specification sheets – that do not represent the design process, and only record the end results of design decisions [9]. As Shah *et al.* noted in their work, these process based, cognitive theory approaches were derived either ad hoc or through controlled experiments that use simple tasks. The suitability of these models for design problems that are much more complex has never been investigated. This lack of investigation and difficulties met in process based measurements of ideation effectiveness led Shah *et al.* to consider outcome based metrics for their study of engineering design [7]. As such, Shah *et al.* developed a framework to measure ideation effectiveness in simple and complex design situations.

Shah *et al.'s* framework includes metrics that measure the effectiveness of formal idea generation methods. The framework addresses that engineering design must be novel – unusual and unexpected – but also must satisfy some intended function or desired specification – have utility. The framework encompasses the rationale that engineering design is goal oriented because a designer's success is judged by how well his or her design meets desired goals and how well he or she has identified the alternative ways of achieving those goals [7]. The framework developed by Shah *et al.* quantitatively measures innovation in student designs based on four metrics – novelty, variety, quality, and quantity. Shah's measures are outcome based metrics with the premise that an idea generation method is considered effective if its use results in "good" ideas [7]. This framework has been utilized in a number of studies for the determination of innovative capacity within engineering design [10] [11] [12].

This work aimed to determine if designs produced in the simulated environment of Nephrotex were considered innovative based on the existing framework developed by Shah *et al.* This work did not aim to create a framework from scratch rather it was focused on determining if an existing ideation effectiveness framework from the design literature was applicable to an epistemic game environment. This work also builds on a previous study completed by Markovetz *et al.* [6] and discusses whether Shah's ideation effectiveness metrics can be used to distinguish between teams that were identified in the prior study to be innovative and non-innovative.

Research Questions

This work evaluated Shah's metrics for novelty, variety and quality based on students' final designs and their final five prototypes submitted for testing. More specifically, this study asks

- How do engineering students' designs produced from Nephrotex score on Shah's quality, novelty, and variety metrics?
- Is Shah's framework able to distinguish between designs that were identified as innovative and non-innovative based on a literature definition of innovative design in an epistemic game environment?

Methodology

Participants

During the spring 2014 and 2015 semesters, at a large research university, 286 sophomore-level chemical engineering students enrolled in an Introduction to Chemical Product Design course participated in Nephrotex. Overall, a total of approximately 50 student teams of 5-6 students each took part in the epistemic game.

Data Sources and Collection

Throughout the design process, students were asked to document their work in online notebooks. Each week, students completed one to two notebook prompts, detailing information such as background research completed, internal consultants' threshold demands, and prototype testing results. Prototype testing results consisted of students' evaluation of their design prototype(s) relative to desired values by the internal/external consultants for product attributes including marketability, cost, reliability, blood cell reactivity, and flux. In this study, data returned from the testing of the groups' final five prototypes and the final designs selected were collected via the online notebooks and analyzed using Shah *et al.*'s metrics for innovation.

Application of Shah Metrics to Nephrotex

It was determined that the quantity metric was not an applicable measure for this study because all students were required to submit the same number of designs. Therefore, this study adopted Shah's novelty, variety, and quality metrics as a way to analyze student designs. These measurements were identified as applicable based on Shah *et al.*'s discussion of exploring and expanding the design space by the means of the quality, variety, and novelty metrics. This section will provide a brief overview of each of the metrics applied to the Nephrotex environment along with a sample calculation using data obtained from a selected Nephrotex team to illustrate how the calculations were conducted [7].

Novelty

The novelty of a design is the measure of how "unexpected" an idea is compared to other ideas generated by groups working on the design problem. The novelty of an idea may be calculated by the following equation:

$$M_1 = \sum_{j=1}^{m} f_j \sum_{k=1}^{n} S_{1jk} p_k \tag{1}$$

Where M_1 is the overall novelty score for the idea which has m attributes at the nth stage of design. S_{1jk} is the score assigned to each option for the j^{th} attribute considered in the k^{th} stage. A stage weighting factor, p_k, was set equal to 1 because only the embodiment design stage was considered. As defined by Shah and colleagues, there are either conceptual or embodiment stages of design [7]. Conceptual stages require creative, divergent thinking and describe the movement between configuration space and concept space where iterations are made to create a working prototype with desired specifications [13]. Embodiment stages describe the identification of important physical and functional constraints from knowledge of the solution [14]. Within Nephrotex, groups have the option of selecting the polymer, process, surfactant, and carbon nanotube percentage for their final dialysis membrane design, hence m=4 attributes. Only the embodiment phase was considered in this study, making n=1 stage of design. An attribute weighting factor, f_j, of 0.3 was assigned to the polymer, process, and surfactant, and a weighting factor of 0.1 was assigned to the carbon nanotube percentage. These weights were assigned based on the relative importance of each attribute to the overall functionality of a kidney dialysis membrane.

Two approaches were used to calculate novelty. The first method is referred to as *a priori* novelty score where a value of S_1 is established before any designs are submitted. The assigned values for S_1 came from the expected commonality of the polymer, process, surfactant, and carbon nanotube percentage used in the design space based on previous play-through examples of Nephrotex in 2012. The values of S_1 used to calculate the *a priori* novelty score are shown in Table 1.

Table 1: *a Priori* Novelty Sub Score

J	**Attribute**	**Novelty Sub Score S_1**			
		S_1=1	S_1=3	S_1=7	S_1=10
1	**Polymer**	Polyrenalate	PES-PVP	Polyamide	PMMA
					Polysulfone
2	**Process**		Phase Inversion	Vapor Dep. Polym.	Dry-jet Wet Printing
3	**Surfactant**	Biological	Steric Hindrance	Hydrophilic	Negative Charge
4	**CNT %**	0%	2%	10%	1.5%
			4%	20%	

As an example for calculating the *a priori* novelty score, if a group's final design includes the PES-PVP polymer, the Dry-Jet Wet Printing process, the Negative Charge surfactant, and 20% carbon nanotube, their *a priori* **polymer** novelty sub score (S_1) would equal 3 based on Table 1, their **process** novelty sub score would equal 10, and so on. A sample calculation for the group's *a priori* novelty score can be seen below.

$$M_1 = 0.3 * 3 + 0.3 * 10 + 0.3 * 10 + 0.1 * 7 = 7.6$$

The second method for determining novelty is referred to as *a posteriori* novelty score. Unlike the *a priori* novelty score, a calculated S_1 score based on submitted designs is incorporated. The *a posteriori* novelty subscore can be calculated by the following equation:

$$S_1 = \frac{T_{jk}C_{jk}}{T_{jk}} * 10 \tag{2}$$

Where T_{jk} is the total number of ideas used within the design process for the j^{th} attribute at the k^{th} stage. In 2014, 20 groups participated in the product design while 30 groups participated in 2015; therefore T_{jk} is equal to 20 and 30 respectively for each year. C_{jk} denotes the total number of designs in which a particular attribute was used; for example if 12 out of the 20 groups in 2014 used steric hindrance as their surfactant, C_{jk} would be equal to 12. Table 2 gives the values of S_1 used to calculate *a posteriori* novelty score for the 2014 class, while Table 3 shows the values of S_1 for the 2015 class.

Table 2: 2014 *a Posteriori* Novelty Sub Scores

Attribute		C_j	S_{11}
Polymer	PMMA	0	10
	Polyrenalate	0	10
	Polysulfone	1	9.5
	Polyamide	7	6.5
	PES-PVP	12	4
Process	Phase Inversion	5	7.5
	Dry-Jet Wet Printing	5	7.5
	Vapor Dep. Polym.	10	5
Surfactant	Biological	0	10
	Hydrophilic	6	7
	Negative Charge	4	8
	Steric Hindrance	10	5
	None	0	10
% CNT	0	0	10
	0.5	0	10
	1	2	9
	1.5	1	9.5
	2	5	7.5
	4	0	10
	6	0	10
	10	0	10
	15	0	10
	20	12	4

Table 3: 2015 *a Posteriori* Novelty Sub Scores

Attribute		C_j	S_{11}
Polymer	PMMA	2	9.33
	Polyrenalate	3	9
	Polysulfone	1	9.67
	Polyamide	11	6.33
	PES-PVP	13	5.67
Process	Phase Inversion	10	6.67
	Dry-Jet Wet Printing	11	6.33
	Vapor Dep. Polym.	9	7
Surfactant	Biological	1	9.67
	Hydrophilic	7	7.67
	Negative Charge	6	8
	Steric Hindrance	15	5
	None	1	9.67
% CNT	0	0	10
	0.5	0	10
	1	3	9
	1.5	4	8.67
	2	7	7.67
	4	1	9.67
	6	0	10
	10	2	9.33
	15	1	9.67
	20	12	6

As an example for calculating the *a posteriori* novelty score, the group mentioned in the previous example included students from the 2015 class. Therefore, their *a posteriori* polymer novelty sub score (S_{11}) would equal 5.67 based on Table 3, their process S_{11} would equal 6.33, and so on. A sample calculation for the group's *a posteriori* novelty score can be seen below.

$$M_{11} = 0.3 * 5.67 + 0.3 * 6.33 + 0.3 * 8 + 0.1 * 6 = 6.60$$

Variety

The variety score is a measure of how deeply the design space was explored throughout the design process. S_i values, which represent the weighted contribution of an attribute to the product's innovative variety score, were assigned for the different attributes of the design in decreasing order based on a hierarchy of attributes or phases. The first phase of the design that was considered was selection of the polymer (S_1=10) for use in the second phase of the design or the process (S_2=6). The third phase considered was the surfactant (S_3=3) and the fourth considered carbon nanotube percentage (S_4=1) as this was the last portion of the product design considered by students. The weights given assure attributes that more effectively distinguish design performance such as the polymer or process selected provide a greater total variety score than attributes at a more detailed level such as the carbon nanotube percentage where performance is less likely to vary drastically. The variety score can be calculated by the following equation:

$$M_3 = \sum_{k=1}^{4} \frac{S_k * b_k}{n} \tag{3}$$

Where b_k is the number of attributes tested at design stage k. For example, if a group tested three prototypes with designs of [PES-PVP polymer, Phase Inversion process, Steric Hindrance surfactant, 20% CNT], [PES-PVP polymer, Vapor Dep. Polym. process, Hydrophilic surfactant, 15% CNT], and [Polysulfone polymer, Dry-Jet Wet Printing process, Hydrophilic surfactant, 6% CNT], their b_1=2 because they tested two types of polymers, their b_2=3 because they tested three different processes, and so on. A sample calculation of this hypothetical group's variety score can be seen below where n is the total number of ideas. In the hypothetical situation, n=3 because the group tested three prototypes.

$$M_{3-revised} = \frac{10 * 2 + 6 * 3 + 3 * 2 + 1 * 3}{3} = 15.6$$

The variety calculation for this study is a modification of Shah's calculation. In Shah's study, genealogy trees were developed for each group on the same principle that a heavier weight is assigned to attributes that more effectively distinguish design performance. However, instead of b_k representing the number of attributes tested at design stage k, it represents the number of branches. For example, the designs mentioned in the paragraph above would give the genealogy tree shown in Figure 1; because there are only two different types of materials used, resulting in two branches for the materials, b_1=2. There is one branch from the PSF material and two branches from the PESPVP material, resulting in three process branches and therefore b_2=3. The variety score would be calculated as such:

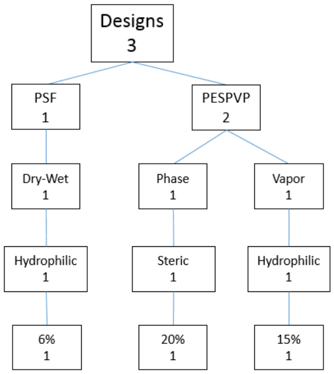

Figure 1: Example group's genealogy tree

$$M_{3-shah} = \frac{10 * 2 + 6 * 3 + 3 * 3 + 1 * 3}{3} = 16.7$$

Quality

The quality score evaluates whether a device in Nephrotex adds "value." This metric was assessed based on the work of Arastoopour and colleagues and takes technical and economic performances, such as marketability, cost, reliability, flux and blood cell reactivity (BCR), into consideration [15]. Four graduated thresholds were placed on each of these criteria, which can be seen in Table 4, and designs were given a point for each threshold met. Thus a "perfect" design that met every threshold in every output category would receive quality score of 20. However, the maximum possible score in Nephrotex was 18 because of constraints placed on the design space.

Table 4: Thresholds for Design Outputs

Rating	**Threshold**	**Marketability**	**Cost ($)**	**Reliability (hrs.)**	**Flux ($m^3/m^2/day$)**	**BCR (ng/mL)**
1	Minimum	250,000	140	1.5	10	90
2	Medium – Low	400,000	120	5	12	75
3	Medium – High	550,000	110	8	13.5	65
4	Maximum	650,000	95	9	17	40

For example, a prototype made from the PES-PVP polymer, Phase Inversion process, Steric Hindrance surfactant, and the 20% CNT was found to have a marketability of 900,000 units, a cost of $120 per unit, a reliability of 9 hours, a flux of 17 m^3/m^2/day, and a blood cell reactivity of 43.333 ng/mL. A sample calculation of the quality score for this prototype can be seen below.

$$M_4 = 4 + 2 + 4 + 4 + 3 = 17$$

Results and Discussion

Using the team scores for each of the Shah metrics, we sought to answer *how do engineering students' designs produced from Nephrotex score on Shah's quality, novelty, and variety metrics?*

Table 5 displays the mean, standard deviation, minimum, maximum, and number of groups for each metric for the combined 2014 and 2015 class sections.

Table 5: Combined Overall Metrics

Shah Metrics	Mean	*stdev*	Min	Max	n
A Priori Novelty	5.82	1.24	3.40	8.40	50
A Posteriori Novelty	6.35	0.85	4.60	8.43	50
Quality	15.6	1.4	13	18	50
Variety	13.5	2.6	6	17	50

Literature ranges from work completed by Shah *et al.* for *a priori* and *a posteriori* novelty scores align with the novelty scores found in this study. The literature range was seen to be 3.0 to 8.05 for *a priori* and 3.28 to 8.46 for *a posteriori* [7] while the range in this study was found to be 3.4 to 8.4 and 4.6 to 8.433 for *a priori* and *a posteriori* respectively. The literature illustrates the application of the novelty score with designs from a student design competition where seven designs were considered "unusual" and therefore novel. These designs received *a priori* scores of 6.25, 6.25, 4.40, 8.05, 4.40, 4.80, and 8.05, and *a posteriori* scores of 6.45, 6.45, 4.46, 8.46, 4.46, 5.16, and 8.46. In the example design competition, there are designs that received higher novelty scores than the ones listed above that were not considered novel [7]. Since the calculated novelty scores in this study match the novelty scores in the literature, this metric seems to be applicable to the Nephrotex environment but a low novelty score doesn't necessarily mean that the design lacks novel components, based on the scoring of the example design competition provided within Shah's work [7]. According to Shah's work, to determine novelty of a design, all designs must be analyzed together and any "unusual" component, number of components; way of design, etc. will make the design novel and therefore innovative. For this reason, a design can have a low novelty score but still contain novel components dependent upon the weighting of the attributes. In Shah *et al.*'s example design competition, students were challenged to build a device made from a fixed set of materials and powered by a fixed volume of pressurized air. The key functions and characteristics of the designs were identified as propulsion/thrust method, medium of travel, motion of device, and number of parts of device. The thrust and medium attribute were weighted 0.35 whereas the motion attribute was weighted 0.2 and a weight of 0.1 was given to the number of parts attribute. A design can have a "usual" propulsion method and travel through an "expected" medium but have an "unusual" motion and number of parts, causing the design to have a low novelty score but novel components. Therefore, it is possible that designs with innovative components may have been produced from students in the Nephrotex environment but not identified as such because of a low novelty score.

The variety score calculated in this study was similar to that Shah *et al.* discusses in their work; however, this study altered the metric slightly. The number of branches in Shah's equations was substituted with the number of attributes tested. An instance was found where students kept the surfactant, manufacturing process, and carbon nanotube percentage constant while changing only the polymer. This team received a perfect variety score because students were required to submit five prototypes for testing, there were five different polymers to choose, and the polymer attribute is at the top of the genealogy tree, meaning the polymer attribute had more of a weight in calculating the variety score than other attributes. This team's genealogy tree is shown in Figure 2 and the variety score would be calculated as such:

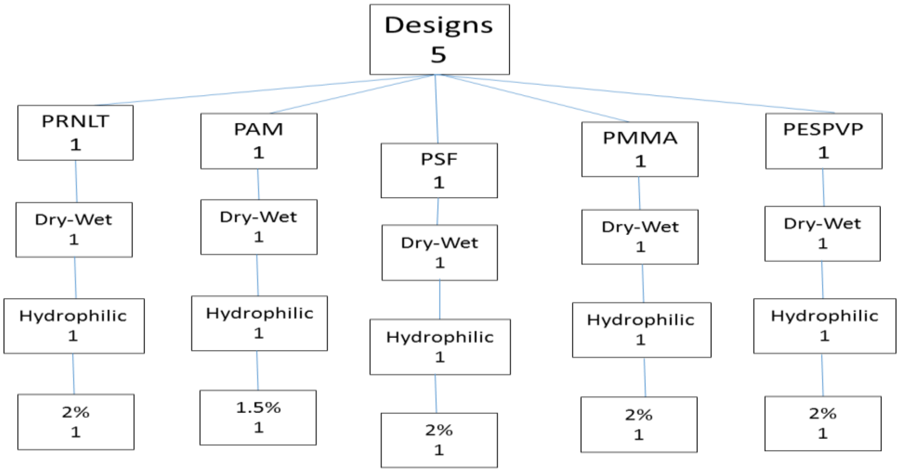

Figure 2: Genealogy tree for an idea set with perfect variety score

$$M_{3-shah} = \frac{10*5+6*5+3*5+1*5}{5} = 20$$

Shah's variety score is considered "perfect" despite the minimal variety within the designs tested. This was the motivation in calculating the variety score as described in the Data Assessment subsection where all aspects of the design impacted the variety score. An example calculation for the genealogy tree shown in Figure 2 using the modified approach to variety is shown below.

$$M_{3-revised} = \frac{10*5+6*1+3*1+1*2}{5} = 12.2$$

Brent Nelson and Jeannette Yen also found flaws in Shah's variety metric and sought to correct for them by creating a new variety metric that addressed these flaws [16]. The flaws discovered by Nelson and Yen include lower scores for higher variety in design ideas and normalizing a group score. The former flaw addressed the double-counting of ideas and suggests counting the number of differentiations in design principles rather than counting the number of branches in each level – the number of differentiations is always one less than the number of branches. For example, for the genealogy tree seen in Figure 2, Nelson and Yen would calculate the variety score to be 10 as they do not continue to count designs within the genealogy tree once they contain only a single design.

$$M_{3-Nelson/Yen} = \frac{10*5+6*0+3*0+1*0}{5} = 10$$

Since each box contains only one design in the top level, nothing is considered in lower levels. Nelson and Yen reworked Shah's equation to encompass their suggestions and developed the following equation where all variables are identical to those addressed in the Methodology section with one added variable d_l which is the number of differentiations at node l:

$$M_{3-Nelson/Yen} = \sum_{j=1}^{m} f_j\left(S_1(b_1-1) + \sum_{k=2}^{4} S_k \sum_{l=1}^{b_k-1} \frac{d_l}{(N-1)}\right) \tag{4}$$

The latter variety flaw addressed the error in giving an average variety score to a set of designs because variety can only be calculated for a set of multiple designs and therefore only applies to the set itself. From fixing this flaw, Nelson and Yen eliminated the need for the quantity metric by incorporating the number of designs in the variety calculation, as our study also had to do. To address this flaw, Nelson and Yen eliminated (N-1) in the above reworked equation [16]. Combining the modifications made to the variety score in this study and the study conducted by Nelson and Yen may allow for the variety matric to better map to the constrained design environment of Nephrotex than the original variety metric designed by Shah *et al.*

The quality metric Shah *et al.* discusses does not give design outcomes in which to measure the quality metric but states evaluation procedures tend to be domain specific. How the prototype performed in marketability, cost, reliability, flux, and blood cell reactivity was the basis for calculating the quality score in this study. These design outputs were developed by Arastoopour and colleagues and consider a prototype's technical and economic performance [15].

The second research question this study sought to answer was *is Shah's framework able to distinguish between designs that were identified as innovative and non-innovative based on a literature definition of innovative design in an epistemic game environment?*

Table 6 displays Shah's metrics for combined teams from 2014 and 2015 identified as innovative within Nephrotex compared against teams that were identified as non-innovative based on the previous work by Markovetz *et al* [6]. Based on this comparison, it was found that the Shah metrics between innovative and non-innovative groups showed little variation and no statistically significant differences [6]. As seen in Table 6, a t test and a Mann-Whitney U test showed no significant difference between innovative and non-innovative groups with regard to variety and novelty scores; however, these tests did show a significant difference between groups with regard to the quality score. The same results were found when calculating Cohen's Effect Size, which again can be seen in Table 6. Novelty and variety scores have a small effect for innovation while quality was found to have a large effect.

Table 6: Combined Metrics for Innovative vs. Non-innovative Designs within Nephrotex

	Teams with Innovative Designs			Teams without Innovative Designs			t test	Mann-Whitney U test	Cohen's Effect Size
Shah Metrics	Mean	*stdev*	n	Mean	*stdev*	n	*p*		*d*
Priori Novelty	5.9	1.117	7	5.809	1.27	43	0.86	0.848	0.0761
Posteriori Novelty	6.59	0.593	7	6.314	0.888	43	0.432	0.364	0.3657
Quality	16.714	1.113	7	15.442	1.333	43	0.021	0.029	1.0361
Variety	13.343	1.843	7	13.53	2.764	43	0.864	0.603	0.0796

A significant difference between innovative and non-innovative teams and a large effect was seen in the quality metric. However, two of the classifications used to categorize an innovative design within Nephrotex in the study by Markovetz *et al.* [6] was "a perfect quality score of 18" and "… design with a quality greater than the section average." Therefore this requirement for innovative design may likely contribute to the significance seen between innovative and non-innovative teams and the large effect the quality metric had for innovative designs in the comparison performed in Table 6. Furthermore, a quality metric was predefined within the Nephrotex environment based upon constraints provided to the students through the internal consultants, indicating that this could have influenced how students approached their design process.

Based on the results obtained, it may be hypothesized that Shah's metrics as a whole may be incorrect for determining innovative designs within a constrained design space. This hypothesis is supported by (1) the need to eliminate Shah's quantity measure from the study since students were limited to the number of designs they were able to submit; (2) the variety metric required adjustment to be able to be applied within the Nephrotex context; and (3) the quality metric was predefined within Nephrotex because of the epistemic game's digital nature. However, it is important to point out that Shah *et al.* [7] noted that not all metrics may apply to each specific design situation and that these metrics don't consolidate well into an overall effectiveness measure. Novelty, as already defined by Shah *et al.*, may be the only metric that can map to a constrained design space without adjustment because of the alignment of literature ranges and this study's calculated ranges.

Study Limitations

This study and its subsequent analysis was all performed on a population of sophomore chemical engineering students which would limit its generalizability to other student populations. However, this research still provides the basis for determining how innovation metrics can be applied to constrained design spaces and provides suggestions for future research that may lead to a better understanding of design in an epistemic game environment.

Conclusion

Epistemic games, such as the virtual engineering internship Nephrotex, allow students to explore creative ways to approach engineering problems while providing a novel alternative to the direct transmission method of instruction. Nephrotex fosters an understanding of the product design process and mitigates the risks associated with product design by allowing students to create designs without the risk of failure that would occur in the real world. When applying Shah's innovation metrics to Nephrotex, it was observed that the entirety of the metrics may not be applicable to the constrained design space. Literature ranges for the *a priori* and *a posteriori* novelty scores overlap the novelty ranges found in this study, showing that this metric is likely most applicable to the Nephrotex environment out of all the innovation metrics discussed by Shah. Shah's variety score needed to be modified in this study to eliminate a perfect variety score being awarded to teams with very little variety in their designs although after modification was able to provide a better mapping to students' design process within the constrained environment. The quality metric was calculated identically in this study as in Shah's study. Shah's quality metric evaluates the design ideas and not the testing of a design after it has been built; therefore is well suited for a virtual internship where tangible models cannot be made;

however, this metric was predefined within the Nephrotex environment. Finally, the quantity measure was not applicable to this study because all students needed to submit the same number of designs. Overall, it was observed that novelty was the most applicable of the innovation effectiveness metrics discussed by Shah *et al.* however with modifications both variety and quality could be applied within the constrained design space.

This study further concluded that the Shah metrics may not be correct for distinguishing between innovative and non-innovative designs in a constrained design space. Quality was the only metric found to show a significant difference between innovative and non-innovative designs and have a large effect for innovative designs. However, based on the literature definition used, quality was a major factor when considering if a design was innovative – designs with “quality greater than section average” and “a perfect quality score of 18” were considered innovative. Variety scores were not considered in the literature definition.

Though Nephrotex is a constrained design space, it is an excellent introductory experience for students to learn aspects of engineering design. It allows students to make design mistakes without the associated real world consequences, provides students with example materials, surfactants, and processes to use in the design of a kidney dialysis membrane, and shows students the necessity of research and performance outcomes within the realm of engineering design. For more experienced design students, the Nephrotex design space could be expanded to allow students to research the design problem and identify materials, surfactants, and processes that could possibly be a solution to the posed design problem. Afterwards, students could design their solution, and develop ways to measure performance outcomes and potentially build these solutions in real life to determine if they perform the way Nephrotex predicts.

Acknowledgements

The authors thank Andrew Glowacki, Timothy Knab, Andrew Kozbial, Jason Lee, Sharlee Mahoney, Mitchell Groenenboom, Hao Chi, and Melissa Lash from the University of Pittsburgh for their role as virtual design advisors for the classes studied herein. The work performed was supported by a University of Pittsburgh Innovation in Education grant as well as an NSF Research Initiation Grant in Engineering Education (RIGEE EEC#1340426), for which the authors are also grateful.

References

[1] Dym, C. L., A. M. Agogino, O. Eris, D. D. Frey & L. J. Leifer. “Engineering Design Thinking, Teaching, and Learning.” *Journal of Engineering Education* 94.1 (2005): 103-120.
[2] Shaffer, D. W. “Epistemic Games to Improve Professional Skills and Values.” *Organization for Economic Co-operation and Development.* 24 Oct. 2006.
[3] “Innovation” Def. 1. *Merriam Webster Online*, Merriam Webster, n.d.

[4] Smith, S. M., David R. Gerkens, Jami J. Shah & Noe Vargas-Hernandez. “Empirical Studies of Creative Cognition in Idea Generation.” *Creativity and Innovation in Organizational Teams.* Ed. Leigh Thompson, Ed. Hoon Seok Choi. Manwah: Lawrence Erlbaum Associates, 2006. 3-20.
[5] Arastoopour, G. “Nephrotex: Measuring First-Year Students’ Ways of Professional Engineering Thinking in a Virtual Internship.” *ASEE 2012 Annual Conference, June 10-13th, 2012.* San Antonio, Texas. 2012.
[6] Markovetz, M. R., R. M. Clark, G. Arastoopour, Z. L. Swiecki, D. W. Shaffer, N. C. Chesler & C.A. Bodnar. “Innovative Design within the Context of Virtual Internships: How Can it be Defined and How is it Related to the Student Design Process.” *ASEE 2016 Annual Conference, June 26-29th, 2016.* New Orleans, Louisiana. 2016.
[7] Shah, J. J., S. M. Smith & N. Vargas-Hernandez. “Metrics for Measuring Ideation Effectiveness.” *Design Studies* 24.2 (2003): 111-134.
[8] Ullman, D. G., T. G. Dietterich & L. A. Stauffer. “A Model of the Mechanical Design Process Based on Empirical Data.” *Academic Press Limited.* 2 Aug. 1988. 33-52.
[9] Nagy, R. L., D. G. Ullman, & T. G. Dietterich. “A Data Representation for Collaborative Mechanical Design.” *Research in Engineering Design* 3.4 (1992): 233-242.
[10] Chusilp, P., and Y. Jin. “Impact of Mental Iteration on Concept Generation.” *Journal of Mechanical Design* 128.1 (2006): 14-25.
[11] Perttula, M., and P. Sipila. “The Idea Exposure Paradigm in Design Idea Generation.” *Journal of Engineering Design* 18.1 (2007): 93-102.
[12] Linsey, J. S., E. F. Clauss, Tolga Kurtoglu, & Arthur B. Markman. “An Experimental Study of Group Ideation Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods.” *Journal of Mechanical Design* 133.3 (2011): N.p.
[13] Austin, S. A., J. Steele, S. Macmillan, & R. Spence. “Mapping the Conceptual Design Activity of Interdisciplinary Teams.” *Design Studies* 22.3 (2001): 211-232.
[14] Chakrabarti, A., T. P. Bligh, & T. Holden. “Towards a Decision-Support Framework for the Embodiment Phase of Mechanical Design.” *Artificial Intelligence in Engineering.* Elsevier Science Publishers Ltd., 1992.
[15] Arastoopour G., W. Collier, N.C. Chesler, J. Linderoth, & D.W. Shaffer. “Measuring the Complexity of Simulated Engineering Design Problems.” *ASEE 2015 Annual Conference, June 14-17th, 2015.* Seattle, Washington. 2015.
[16] Nelson, B. A., J. O. Wilson, D. Rosen, & J. Yen. “Refined Metrics for Measuring Ideation Effectiveness.” *Design Studies* 30.6 (2009): 737-743.