122nd ASEE Annual
Conference & Exposition
June 14 - 17, 2015
Seattle, WA
Making Value for Society

Paper ID #11436

Epistemic Network Analysis as a Tool for Engineering Design Assessment

Ms. Golnaz Arastoopour, 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.

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

Naomi C. Chesler is Professor and Vice Chair 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
Zachari Swiecki, University of Madison-Wisconsin

Graduate student in educational psychology, learning sciences area

Epistemic Network Analysis as a Tool for Engineering Design Assessment

Introduction

Historian of education Diane Ravitch [1] argues: "Education means to lead forth, but it is impossible to lead anyone anywhere without knowing where you want to go" (pg. 25). Educational standards developed by instructors and institutions play a critical role in leading students—they define what students should know and what they should be able to do at a certain level. In other words, standards provide a *destination* for where we want students to be at a certain point in their education. However, simply knowing the destination is not enough to help students get there. We have to have a *roadmap* to guide students in their travels and to determine if they have arrived at the destination. The roadmap consists of the curricula students engage in and the assessments used to determine if they have met certain standards. Unfortunately, not all standards set by educational institutions offer a roadmap for curriculum development and assessment.

This is particularly problematic in engineering education, because a critical component of the practice – engineering design – has traditionally been difficult to assess. In this paper, we outline and test an approach to addressing this problem. We examine a set of engineering design education standards and then propose and test a method for developing curricula and assessment that is closely linked to those standards.

ABET Criteria for Student Outcomes

In undergraduate engineering education, ABET, Inc. has developed the most widely adopted engineering learning standards and curricular expectations. Two decades ago, ABET criteria for accreditation were widely criticized by universities. Many believed the guidelines did not prepare students for 21st century practice, claiming the criteria did not focus on student learning outcomes and neglected design education [2], [3]. In response, ABET developed and published the Engineering Criteria 2000 (EC2000), which consist of eight general guidelines for bachelor's level programs. The guidelines include recommendations on a range of programmatic issues, such as: students should be advised during their undergraduate careers and monitored to foster success; programs should have published educational objectives; and faculty members must be qualified and demonstrate abilities to instruct and assess curriculum [4]. Of these broad recommendations, Criterion 3 (Figure 1) directly addresses student outcomes: what students are expected to know and be able to do by graduation. Criterion 3c in particular addresses engineering design abilities.

(a) an ability to apply knowledge of mathematics, science, and engineering	(b) an ability to design and conduct experiments, as well as to analyze and interpret data
(c) an ability to design a system, component, or process to meet desired needs within realistic constraints such as	(d) an ability to function on multidisciplinary teams

economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability	
(e) an ability to identify, formulate, and solve engineering problems	(f) an understanding of professional and ethical responsibility
(g) an ability to communicate effectively	(h) the broad education necessary to understand the impact of engineering solutions in a global,
economic, environmental, and societal context	(i) a recognition of the need for, and an ability to engage in life-long learning
(j) a knowledge of contemporary issues	(k) an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

Figure 1. ABET Student Outcomes Criteria 3(a-k) 2014-2015

Although ABET student outcome criteria identify categories of student competency, the categories are very broad. For example, criteria 3e states “an ability to identify, formulate, and solve engineering problems.” However, there are various aspects to problem identification, formulation, and problem solving that depend on whether the problem is more open-ended or structured. Furthermore, while the criteria provide information about domain-relevant skills and knowledge, two key elements are missing from the criteria. The ABET standards specify neither (1) the sorts of curricula and activities students should be engaged in for each category, nor (2) the types of assessment model that would be effective in measuring competency in each category. In other words, there is no roadmap to guide students toward their destination of meeting these standards or to assist instructors in developing appropriate and effective curricula and assessment models.

Evidence-Centered Design Assessment

Effective educational assessments classify what students say, do, or produce in a particular setting to make broader inferences about their abilities and learning development. Decades of research in the learning sciences suggests that learning is social and situated [5], [6]. Thus, developing curricula requires consideration of the cultural and social settings in which the curricula and assessments are developed and applied [7], but more specifically, it requires consideration of what counts as *evidence* of learning and how that evidence will be *modeled*. An effective assessment design supports assessment claims with student evidence, which Mislevy and colleagues [8], [9] call *Evidence Centered Design* (ECD) assessment. ECD has three key components: Domain Analysis—information about beliefs, abilities, or expertise in a particular domain; Domain Modeling—claims, evidence, or variables that organize claims about proficiency, such as what students’ do, say, or produce in a setting; and Conceptual Assessment—assessment tools, models, and machinery applied to student work that would support claims about student proficiency. This method stresses however that design of an assessment strategy does not necessarily follow a linear process. Rather, design of an assessment strategy requires alignment and realignment of the three components throughout the development process. Figure 2 shows a high-level interpretation of the ECD method.

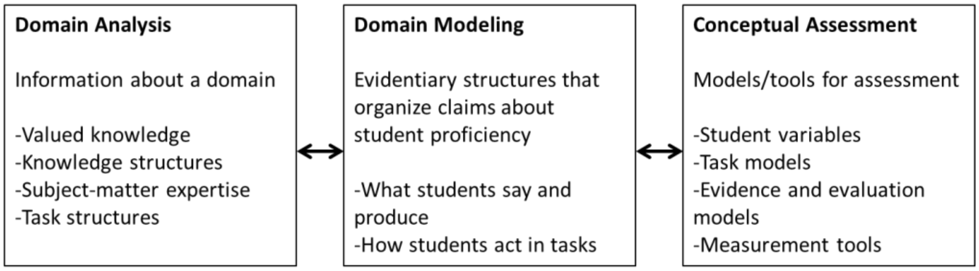


Figure 2. High level interpretation of the Evidence Center Design (ECD) assessment method

As a simple example, consider the ACT College Readiness Exam that high school students take when applying to colleges. The test is based on the College and Career Readiness Standards (CCRS), a set of skills and knowledge elements that students need to master to become ready for college and post-high school careers. One section of the test is based on the CCRS mathematics standards, which includes skills such as identifying a digit's place value and ordering fractions. The test has sixty multiple choice mathematics questions. The assessment consists of determining the number of questions that the student has answered correctly and then rescaling the number so that it ranges between 1 and 36, with 36 being the highest score. The result is the student's ACT mathematics score, which can then be compared to other students' scores or to benchmarks set by colleges [10]. In this example, the domain analysis component is the claim that the student is or is not college-ready in terms of mathematics ability; the domain modeling component is the mathematics questions that the student answers; and the conceptual assessment is the counting and scaling of the student's responses.

In terms of ECD, ABET student criteria address one of the three components. The student outcomes are based on real-world engineering practice, which addresses the domain analysis component. However, the criteria do not detail how evidence of student learning relates to each outcome. In addition, the criteria do not offer suggestions for assessment of the evidence of learning. ABET's intention may have been to provide flexibility by not prescribing forms of acceptable evidence and assessment for faculty and institutions; however, according to ECD, standards, evidence, and assessment are interrelated and inform one another. Mislevy and colleagues [8] claim, "The world presents us with inseparable events, particular people doing specific things in concrete situations. Patterns we conceive in this interplay are the basis of assessment design variables, and our conceptions of them can be regularized (as they are in the ECD structures) but not uncoupled" (p. 8). Therefore, the processes of identifying evidence and developing assessments are not separate from the development of educational standards based on real-world practice—they are dependent on one another.

Using an ECD Model for Engineering Design Education and Assessment

Using the ECD model as an assessment framework, we look at one example of an engineering design educational framework developed from ABET criteria. Then, we identify student evidence of learning and develop assessment tools to align the standards, evidence, and assessment for engineering design thinking.

Standards: Design Attribute Framework

We begin with Safoutin and others' [11] *design attribute framework*, a set of standards that focuses on engineering design abilities and stems from ABET student criteria 3c. Their work consists of a detailed list of standards that transforms the imprecise ABET learning outcomes into information that instructors could use in curriculum and assessment development. The framework provides descriptions of the various stages of the design process and identifies what is required of students at each step. For example, Safoutin identified one component as *needs recognition* and detail several subcomponents, such as identifying needs to be served by the design, evaluating societal needs, evaluating the cost associated with a product, and identifying target customers and markets.

Safoutin and colleagues generated the design attributes from a large number of engineering design process models and from verbal protocol analysis studies, in which students were observed while engaging in a design task. As a result, the design attribute framework is based on engineering practice. In order to establish the forms of student evidentiary structures that align with the practice of engineering, we must understand the environments in which professional engineering practice takes place. More importantly, we must identify the particular and complex ways in which engineers act, know, and think and the ways in which these elements are related.

Evidence: Epistemic Frame Theory and Virtual Internships

We approach professional practice from the work of *epistemic frame theory*—the idea that every practice has a particular way of viewing the world [12], [13]. This theory suggests that professionals rely on domain-specific skills and knowledge to make and justify decisions. They have characteristics that define their identities as members of a practice, as well as a set of values they use to identify important issues and problems in the field. Professionals developed epistemic frames when they make *connections* and understanding links between these skills, knowledge, identities, values, and epistemological elements that are characteristic of the community. Professionals-in-training, i.e., students, exhibit the epistemic frames of professionals when they make the same sorts of connections. For example, an engineering intern might make a design decision to increase the safety factor of a product for the wellbeing of the client based on a completed stress analysis. In this case, the engineer-in-training is justifying the design decision by valuing the safety of the client and executing the skill of conducting and interpreting a stress analysis. She knows which values to consider and what information to gather in order to make a design decision *the way a professional engineer would*. In this view, the *evidence* of student learning can be taken from the notes she has taken in her engineering notebook and the results of her stress analysis tests.

Professionals-in-training develop epistemic frames by participating in a training structure, such as a practicum or an internship, through which they interact with more knowledgeable members of the profession. Some students may participate in training structures that are *simulated* professional practica. For example, we have developed a first-year course in which students role-play as engineering interns and participate in two 8-week-long virtual internship programs [14].

In one virtual internship, *Nephrotex*, students design a filtration membrane for a hemodialysis machine. In a second internship, *RescuShell*, students design an exoskeleton to assist rescue workers. Throughout each internship, students interact with their team members and their mentors via a chat program. Mentors guide them through the activities and occasionally ask them to reflect on their work in a digital engineering notebook. At the end of the course, students present their work to their colleagues and instructor [15]. By participating in the simulated engineering internships, students learn the ways in which engineers at a company connect critical aspects in an engineering design epistemic frame [16], [17]. In virtual internships, the evidence of student learning is captured through digital engineering notebooks and team chat logs.

Assessment: Epistemic Network Analysis

The development of students' epistemic frames in virtual internships can be quantified using epistemic network analysis (ENA) [18], [19]. Because the learning that takes place during a practicum can be characterized by the connections between elements of a professional frame, ENA measures when and how often students make such links during their work. ENA creates a network model in which the nodes of the network represent the key components from a domain. The links between these nodes quantify how often a person has made connections between these elements at some point in time. In this way, ENA models the development over time of a student's epistemic frame, and in turn, quantifies and assesses their ability to think and work like professionals.

Thus, the *assessment* of student learning is done by measuring connections between *standards* (design attributes) identified in *evidence* of student work (students' engineering design notebooks and logs of chat between students).

Research Questions

By framing assessment in terms of ECD, in this study we develop an assessment tool that measure connections between engineering design standards within student work. We first identify evidence of student learning in a virtual internship environment based on a set of engineering design standards. Then, we hypothesize that the appropriate assessment tool in this context would be ENA, which can measure connections in student teams' design discourse and thus assess student teams' design abilities within the context of real-world engineering design practice. To test this hypothesis, we examined first year undergraduate student team data from a virtual internship program, *Nephrotex,* and assessed student team learning using ENA. For the student outcomes, we used Safoutin and colleagues' design attribute framework; for the evidence of learning we used (1) student chat logs from their design teams and (2) the specifications for each team's final device. For the assessment tool we used (1) ENA to model design discourse networks and (2) ran a statistical test to see if ENA revealed differences between students with varying quality of devices.

We ask the following research questions:

1) What types of connections between design attributes are made by student teams that produced **high**-quality designs? And what types of connections between design attributes are made by student teams that produced **low**-quality designs?

2) Is there a significant difference between design discourse networks of students that produced **high**-quality designs and design discourse networks of students that produced **low**-quality designs?

Methods

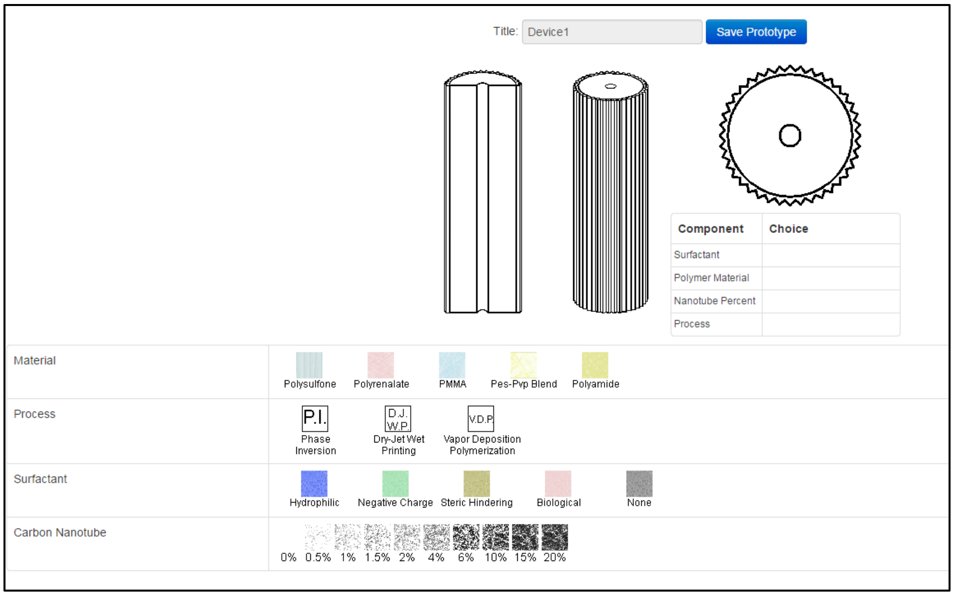
Virtual Internship Description

In *Nephrotex*, students role play as interns at a fictional biomedical engineering design company, where they work on teams to design dialyzers for hemodialysis machines. Research and design activities and team interactions all take place through the web platform that supports the internship. Students begin by logging into the company website, which includes an email and chat interface. Acting as interns, they send and receive emails to and from their supervisor (a non-player character) and use the chat window for instant messaging with other team members and their assigned design advisor.

After conducting background research within the *Nephrotex* website, interns examine company research reports based on actual experimental data with a variety of polymeric materials, chemical surfactants, carbon nanotubes, and manufacturing processes. After collecting and summarizing research data, interns begin the actual design process using the simulated engineering drawing tool (Figure 3). First individually and then in teams, students develop hypotheses based on their research, test these hypotheses in the provided design space, and analyze the results provided.

The design space contains four inputs and five outputs (Figure 4). Interns also become knowledgeable about internal consultants within the company who have a stake in the outcome of their designed prototype. These consultants value different outputs, which are essentially performance criteria. Each of the five internal consultants in *Nephrotex* prioritizes two output parameters and identifies specific threshold values for each output. For example, the clinical engineer would like a high degree of biocompatibility and high flux, and the manufacturing engineer would like a device with high reliability but low cost. The consultants’ concerns are often in conflict with one another (e.g., as flux increases, cost also increases), reflecting the conflicting demands common in professional engineering design projects.

During the second half of the internship, students switch teams and inform their new team members of the research they have conducted thus far in the internship. In the new teams, students test more devices, analyze the second iteration of results, and make a choice for a final prototype. During the final days of the internship, students present their final device design and justify their design decisions to the class and instructor, then complete an exit interview with survey questions about their attitudes towards the engineering profession.



Component	Choice
Surfactant	
Polymer Material	
Nanotube Percent	
Process	

Figure 3. Engineering design drawing simulation where student select input parameters.

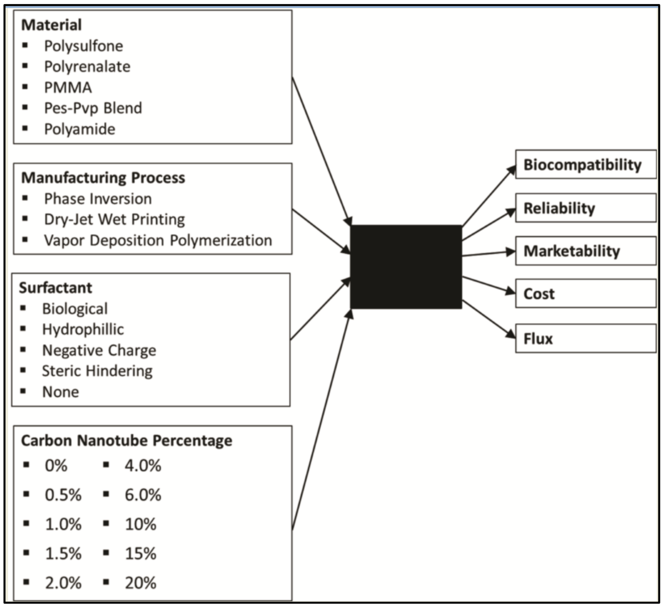


Figure 4. *Nephrotex* design space with four categories of inputs (left) and five categories of outputs (right).

Data Collection and Participants

Participants were first-year undergraduate engineering students at a large public university. These students were enrolled in an introductory engineering course in which students participated in two 8-week long virtual internships.

We collected data from students in *Nephrotex* in two forms (1) chat logs from teams of students from the second half of the simulation in which they made their final design decisions and (2) each team's final design specifications. The data were collected from two instances of *Nephrotex*. Both instances contained five teams of three to five students each, for a total of 10 teams and 46 students overall.

Discourse Analysis

To examine the design processes that students used in *Nephrotex*, we developed a coding scheme based on Safoutin and colleagues' (2000) design attribute framework, which stems from ABET student outcome criterion 3c. Their original coding scheme consisted of fourteen elements: need recognition, problem definition, planning, management, information gathering, idea generation, modeling, feasibility analysis, evaluation, selection/decision, implementation, communication, documentation, and iteration. We selected and modified 7 of the 14 codes that were applicable to *Nephrotex* (Figure 5). We removed need recognition and modeling because students are given the needs statement and the modeling tools within the internship program. We removed idea generation and implementation because students do not create a novel design or a physical prototype—all designs are virtually produced. Finally, we removed iteration and communication because students are required to iterate through two design cycles and to use the chat tool to communicate. In addition, all chat discourse is by its nature used to communicate between team members.

Code	**Description**	**Actions**	**Examples**
1. Problem Definition	Determining design objectives and functional requirements based on needs statement, identifying constraints on the design problem, and establishing criteria for acceptability and desirability of solutions.	• Transform statement of need to statement of design objectives (functional requirements). • Identify or reference constraints on the design problem.	*This material maximizes flux which is very important to the design because it allows patients to have a shorter treatment time.* *yes because some consultants wanted to maximize flux while others wanted to minimize cost*

2. Planning	Development of an initial design strategy, including an overall plan of attack, decomposition of design problem into subtasks, prioritization of subtasks, establishment of timetables and milestones by which progress may be evaluated.	• Develop a design strategy. • Decompose problem into subtasks where appropriate.	*How about everyone describes what the strengths and weaknesses are of the material they worked with based on their previous designs?* *Yes, we should each contribute one prototype containing our material, but we should keep the other variables of each design somewhat constant, so that we can easily compare the results of the different designs.*
3. Management	Guidance of course of action during design and in response to changing conditions.	• Manage time and resources to meet timetable and milestones.	*I think the deliverable is due at 5 so I don't think that would work* *okay, guys.... I think the personal deadline we should set is midnight tonight.*
4. Information Gathering	Gathering information about the design problem, including the need for a solution, user needs and expectations, relevant engineering fundamentals and technology, and feedback from users.	• Gather or reference data to verify the existence of a problem including data on customer perceptions and desires. • Gather or reference relevant engineering fundamentals and technological state-of-the-art.	*The graphs made all of the options easily comparable in a side to side format* *We would need to make sure all toxins can pass through the membrane and anything that needs to stay in the blood does not get filtered out*
5. Feasibility Analysis & Evaluation	Evaluating feasibility of alternatives or proposed solutions by considering	• Evaluate feasibility of multiple alternatives in terms of	*That sounds good but if we wanted a true base model/cheapest we would have to do no surfactant or*

	stated constraints as well as implied constraints such as manufacturability, compatibility, cost, and other criteria. Objectively determining suitability of alternatives or proposed solutions by comparing actual performance to evaluation criteria.	constraints. • Recognize unstated constraints such as manufacturability or assemblability in evaluating designs. • Use evaluation criteria to objectively judge acceptability, desirability of alternatives.	*CNT* *I suppose for the patient, Christopher's may be better, but Scottland's is the only prototype that met all the requirements.* *I think mine would be a good choice too actually because it still meets the internal consultants requests and is at least a little less expensive* *I think Prototype 3 on this last testing batch gave the best results, covering the most aspects, most equally*
6. Selection/Decision	Selection of the most feasible and suitable concept among design alternatives.	• Discern feasible solutions or partial solutions • Use evaluation to select feasible alternative that best satisfies objectives.	*I think that prototype would probably be the best option.* *Okay, so then our prototype would be PESPVP, Dry-jet wet, hydrophilic, 20%CNT* *Then yes let’s use that device* *So we each pick the one that we think will perform the best, and then compare them?* *I think mine would be a good choice too actually because it still meets the internal consultants requests and is at least a little less expensive*
7. Documentation	Production of usable documents of record regarding	• Document decisions and decision criteria.	*We can also include a nice 3 sentence justification.* *Alright I can post my*

	the design process and design state, including decision history and criteria, project plan and progress, intermediate design states, finished product, and use of product.	• Keep a journal or other record of design development. • Create and maintain planning documents and status assessment reports. • Document the finished product or process as appropriate for the discipline according to standard practice.	*notebook after in the shared area* *I have 4 designs and 3 justifications in my notebook* *We created a google document to work in and we divided tasks among the group members*

Figure 5. Design coding scheme based on Safoutin and colleagues' (2009) design attribute framework.

To validate our coding scheme, we used a discourse code developer, The HandCoder, and an automated discourse coder, the AutoCoder [20]. The first coder (Coder One) developed keywords and keyword algorithms for each code in Figure 5 using the HandCoder and a training set of utterances. In this study, an utterance is defined as the text a student types in the chat interface before hitting enter. Once keywords were developed, the AutoCoder automatically coded utterances for instances of those keywords through regular expression matching, assigning a code to utterances with matching keywords. After the automated coding was complete, Coder One and the AutoCoder coded a testing set consisting of 30 to 60 utterances for each code category.

Next, for validation purposes, a second coder (Coder Two), who had no knowledge of the keywords, coded the same utterances as Coder One in the testing set. Inter-rater reliability was measured using Cohen's kappa between the AutoCoder and Coder One and between Coder Two and Coder One (Figure 6). All kappas were above 0.66 (kappas above 0.60 are accepted as substantial [21]). In addition, the HandCoder also computes the likelihood that a coded test set with a given kappa value could have come from that test set. If the p value is less than 0.05, then we can say with 95% confidence that that test set comes from the larger set of utterances. This allows us to validate our coding scheme and generalize our test set to the entire set of utterances. Coders can validate codes with anywhere from 10-50 utterances at a time depending on how frequently the code appears in the data [20].

After measuring inter-rater reliability and computing the statistical likelihoods, we coded our entire set of utterances using the AutoCoder. In total, 2569 utterances were coded using the design coding scheme.

Code	**Kappa between Coder One and AutoCoder**	**Kappa between Coder One and Coder Two**	**Number of Utterances Coded**
Problem Definition	1*	.87*	30
Planning	1*	.66*	30
Management	1*	.93*	30
Information Gathering	.90*	.90*	60
Feasibility Analysis & Evaluation	.83*	.66*	60
Selection/Decision	.89*	.80*	60
Documentation	1*	1*	30

Figure 6. Cohen's kappa between the HandCoder and Coder One (Kappa 1) and Coder One and Coder Two (Kappa 2). * indicates significance at the $\alpha = 0.05$ level.

Next, we analyzed the coded chats of each group from the second half of *Nephrotex* using the ENA tool. We used ENA to measure the development of the connections that students participating in *Nephrotex* made between elements of the design process, as defined by our coding scheme.

ENA is described in greater detail elsewhere [22]–[24], but in short, ENA measures the connections between discourse elements, or codes, by quantifying the co-occurrence of those elements within a defined *stanza*. Stanzas are collections of utterances such that the utterances within a stanza are assumed to be closely related topically. In this study, we defined stanzas in terms of the activities in the internship, such as background research or team design discussions.

More specifically, for any two codes, the strength of their association in a network is computed based on the frequency of their co-occurrence in discourse. For example, the stanza in Figure 7a would be coded for planning and selection/decision, but not for documentation, feasibility & evaluation, management, information gathering, or problem definition. Figure 7b shows this stanza represented as a network where the elements that co-occurred in that stanza are now connected while elements that do not co-occur are not connected. Figure 7c shows this stanza as a symmetric adjacency matrix where the codes are represented both as rows and columns. Elements that co-occurred are represented by a one, and elements that did not co-occur are represented by a zero. Not all codes are included in this representation for visual clarity.

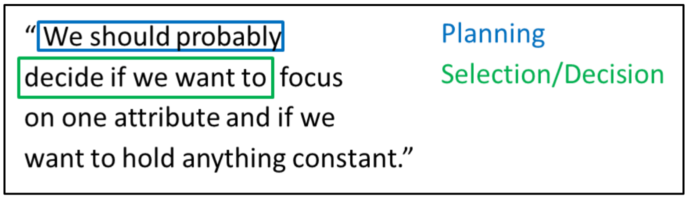

"We should probably decide if we want to focus on one attribute and if we want to hold anything constant."

Figure 7a. Example stanza coded for two design codes

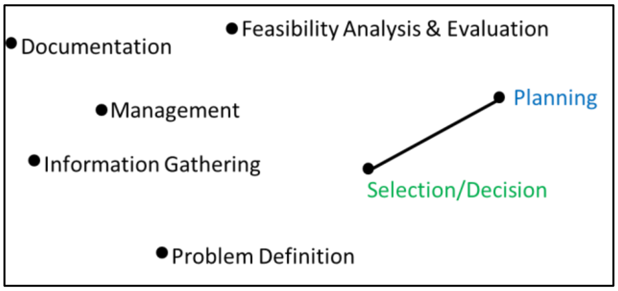


Figure 7b. Example stanza represented as a network

	Planning	**Selection/Decision**	**Documentation**	**...**
Planning	0	1	0	**...**
Selection/Decision	1	0	0	**...**
Documentation	0	0	0	**...**
...	..	...	...	..

Figure 7c. Example stanza represented as an adjacency matrix

ENA constructs an adjacency matrix for every stanza. The adjacency matrices were summed for every team of students and normalized so that groups with *more* discussion in chat would not be weighted more heavily than groups who had less discussion but used the same configuration of connections in their discourse. Finally, the matrices were represented as vectors in a high-dimensional space and a singular value decomposition was conducted to rotate the vectors so as to show the greatest variance among the matrices. This approach is mathematically similar to a principal components analysis. In this rotated space, each team's adjacency matrix was represented as a point in high-dimensional space which roughly corresponded to the network's center of mass. Each dimension in this space can be interpreted by examining the loadings (rotation) matrix, which, again, is similar to the interpretation in a principal components analysis.

We used ENA to create a design discourse network for every team of students, which was an accumulation of the chat discussions among the team members throughout the second half of the virtual internship after students switched teams to design their final devices.

Device Quality Analysis

To investigate the relationship between the teams' design discourse networks and the quality of their final designs, we calculated a quality score for each team's final device. Quality was measured in terms of how many consultant thresholds were met by a team's final device.

We assigned a quality score for each team's final device based on the number of consultant thresholds the device meets. The median quality score was 16.5. Student teams that scored below this value were categorized as low scoring and student teams that scored above were categorized as high scoring (1 = high scoring, 0 = low scoring). Five teams were in the low-quality group and five teams were in the high-quality group.

Significance Testing

Next, we used an independent-samples t-test (assuming unequal variances) to determine if there was a significant difference between the design discourse networks of student teams who produced high-quality devices and the design discourse networks of student teams who product low-quality devices.

Results

RQ1: What types of connections between design attributes are made by teams that generate **high**-quality designs? And what types of connections between design attributes are made by teams that generate **low**-quality designs?

The first two dimensions of ENA results for this study (Figure 8) show that there is some distinction between the groups with low-quality devices (red) and the groups with high-quality devices (blue). In particular, the groups with low-quality devices have lower values on dimension one and the groups with high-quality devices have higher values on dimension one.

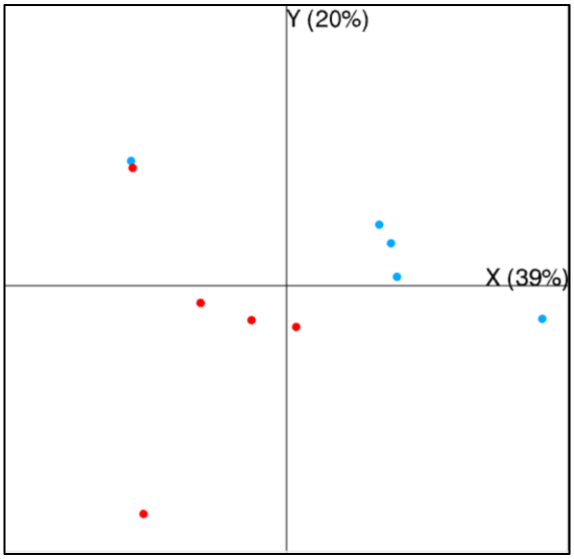


Figure 8. First two dimensions of ENA results for student groups who generate low-quality devices (red) and student groups who generate high-quality devices (blue). The points represent

the centroids of each group’s network. The first dimension (X) accounts for 39% of the variance in the data, and the second dimension (Y) accounts for an additional 20%.

To gain more insight into the differences between student groups that generate low- and high-quality devices, we plotted the mean network connections (Figure 9). The connections distinguishing the low- and high-scoring groups are connections to *management*. That is, the discourse of student teams that generated high-quality devices on average showed more connections between management talk and other elements of engineering design than the discourse of student teams that generated low-quality devices.

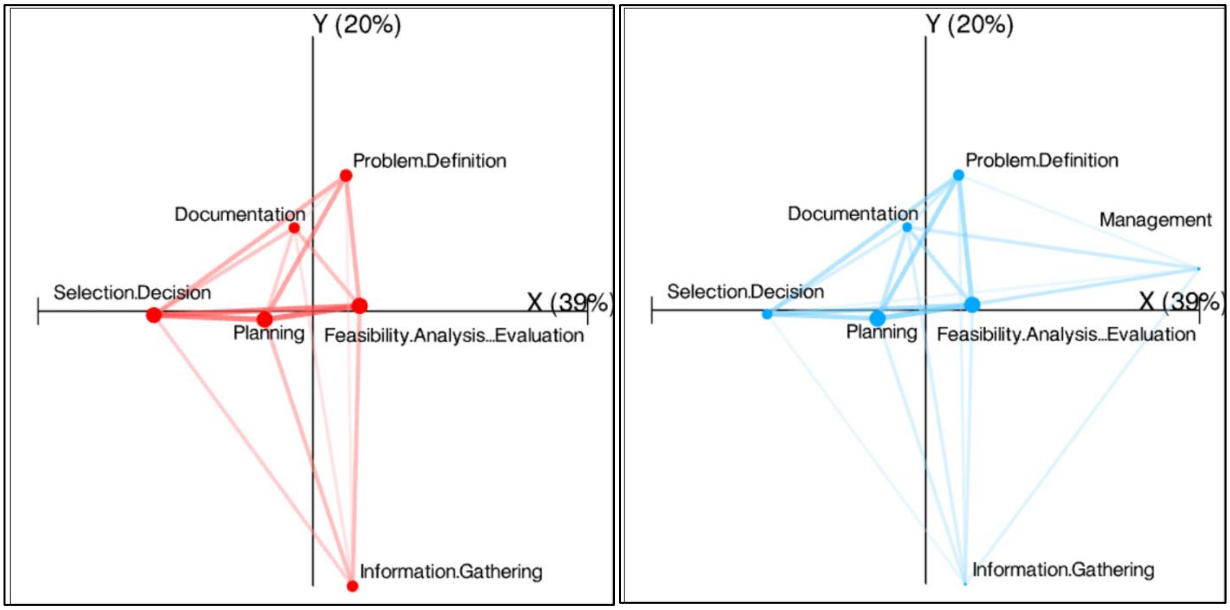


Figure 9. Mean network representations of student teams that generate low-quality devices (red) and teams that generate high-quality devices (blue). Thicker lines indicate stronger and more frequent connections between elements.

For example, students that generated high-quality devices engaged in discourse that involved managing their decision making and planning (Figure 10).

User	Chat Utterance	Design Attribute Code
25	Lets all submit our prototype via email then we can decide on one of us to create the teams batch	Planning and Selection/Decision
36	yeah, yeah.	
36	I don't even think we need until 5 P.M. today	Management
41	Yes, but we need to complete a notebook too	Documentation
36	Yeah, we can complete by like 10:20	Management

Figure 10. The chat discourse of one student team that generated high-quality devices

Because student teams that made more connections with management in their networks are mostly located on the right in Figure 9, we can interpret ENA dimension 1 as an Integrated

Management score. A higher Integrated Management score (i.e., a rightward shift on ENA dimension 1) indicates that a team is making more connections between management and other design attributes.

RQ 2: Is there a significant difference between the design discourse networks of students that produced **high**-quality designs and the design discourse networks of students that produced **low**-quality designs?

There was not a significant difference between design discourse networks on the Integrated Management dimension (ENA dimension 1) for student teams that produced high-quality designs (M=.13, SD=.24) and student teams that produced low-quality designs (M= -.13, SD=.11; t=(5.7)=2.23, p=.07). This suggests that there is not a difference in the Integrated Management dimension between teams that produced high- and low-quality devices. However, the effect size, Cohen's d, was equal to .71, which is interpreted as a large difference between the mean of the two groups.

Discussion

The results above show that ENA can be used to examine student teams' design discourse networks in *Nephrotex*, a virtual internship program for first-year undergraduate engineering students. In addition, we scored the quality of student teams' final devices. Taken together, the discourse networks and the device quality scores reveal that student teams that integrated management with all the design attributes were more likely to produce high-quality devices (these results were not significant at the .05 level, but had a large effect size). Thus, ENA and device quality scoring can be used together as a form of assessing and making claims about student teams' design abilities.

The results above using ENA and product scoring as an assessment method were developed using ABET student criteria as a set of educational standards. However, ABET doesn't provide information on what forms of assessment and student evidence to use for each criterion. In this study, to develop an evaluation tool that assesses ABET criteria, we used ECD as a framework. ECD states that standards, evidence of student work, and assessment are all connected and that assessment development requires alignment and realignment of these three elements. Therefore, we adjusted the design attribute framework (a set of standards based on ABET criteria 3c), to fit with our learning environment, a virtual internship program, which is rooted in epistemic frame theory, the idea that professionals connect ways of knowing, being, and thinking. Then, we identified forms of student work within the virtual internship that aligned with the standards we developed, such as student chat logs. Finally, we employed an assessment method, ENA, which closely aligns with the elements of student work and the theoretical framework that the virtual internship is based on. Therefore, the development of standards, student work, and assessment are all dependent on one another, and the development of one aspect informs the other.

It is in this sense that we can see that not only is learning and the development of standards situational and contextual, but because standards are linked to assessment, assessment methods are also context dependent. This finding suggests that instructors should think about the context of the standards, curricula, and assessment methods that they are employing and consider if they are appropriately aligned to support the claims they are making about student work.

There are of course limitations in this study. Most obviously, the nature of this study means that the conclusions we draw about students' abilities are limited to the particular students in this virtual internship environment. Further, this study uses a small sample size (10 student groups), which may have been contributing to the non-significant results. However, we plan on conducting future studies with larger sample sizes. Finally, this particular study assesses design abilities within student teams and not individual students. However, ENA is a tool that has been used to measure engineering epistemic frame development in individual students in previous studies [25], [26], and in future studies, we plan to use ENA as a tool for assessing individual students' design abilities.

Conclusion

The results here suggest that focusing on how educational standards, curricula, and assessment are interdependent and dependent on the social and cultural setting should be useful for instructors when designing assessments for engineering education. In addition, ENA in particular can be a powerful tool for assessing student design abilities.

Acknowledgements

This work was funded in part by the MacArthur Foundation and by the National Science Foundation through grants DRL-0918409, DRL-0946372, DRL-1247262, DRL-1418288, DUE-0919347, DUE-1225885, EEC-1232656, EEC-1340402, and REC-0347000. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

References

[1] D. Ravitch, *National Standards in American Education: A Citizen's Guide*. Washington, D.C.: Brookings Institution Press, 1996.
[2] L. R. Lattuca, P. T. Terenzini, and J. F. Volkwein, "Engineering Change: A Study of the Impact of EC2000," 2006.
[3] J. W. Prados, "Can ABET Really Make a Difference?," *Int. J. Eng. Educ.*, vol. 20, no. 3, pp. 315–317, 2004.
[4] ABET, "Criteria for Accrediting Engineering." 2014.
[5] A. Bandura, *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
[6] J. Lave and E. Wenger, *Situated learning: Legitimate peripheral participation*. Cambridge, MA: Cambridge University Press, 1991.
[7] R. Mislevy, "Evidence and inference in educational assessment," *CSE Technical Report 414*. National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, 1996.
[8] R. J. Mislevy, L. S. Steinberg, and R. G. Almond, "On the Structure of Educational Assessments," *Meas. Interdiscip. Res. Perspect.*, vol. 1, no. 1, pp. 92–101, 2003.
[9] R. Mislevy and M. M. Riconscente, "Evidence-centered assessment design: Layers, concepts, and terminology," in *Handbook of Test Development*, S. Downing and T. Haladyna, Eds. Mahwah, NJ: Erlbaum, 2006, pp. 61–90.
[10] ACT, "Technical Manual: The ACT," Iowa City, IA, 2014.

[11] M. J. Safoutin, C. J. Atman, R. Adams, T. Rutar, J. C. Kramlich, and J. L. Fridley, “A design attribute framework for course planning and learning assessment,” *IEEE Trans. Educ.*, vol. 43, no. 2, pp. 188–199, 2000.

[12] D. W. Shaffer, “Models of situated action: Computer games and the problem of transfer,” in *Games learning, and society: Learning and meaning in the digital age*, C. A. Steinkuehler, K. D. Squire, and S. A. Barab, Eds. Cambridge, UK: Cambridge University Press, 2012, pp. 403–433.

[13] D. W. Shaffer, *How Computer Games Help Children Learn*. New York: Palgrave, 2007.

[14] N. C. Chesler, A. R. Ruis, W. Collier, Z. Swiecki, G. Arastoopour, and D. W. Shaffer, “A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking,” *J. Biomech. Eng.*, vol. 137, no. 2, pp. 1–8, 2015.

[15] N. C. Chesler, G. Arastoopour, C. M. D’Angelo, E. A. Bagley, and D. W. Shaffer, “Design of professional practice simulator for educating and motivating first-year engineering students.,” *Adv. Eng. Educ.*, 2012.

[16] G. Arastoopour, N. C. Chesler, and D. W. Shaffer, “Epistemic persistence: A simulation-based approach to increasing participation of women in engineering,” *J. Women Minor. Sci. Eng.*, vol. 20, no. 3, pp. 211–234, 2014.

[17] G. N. Svarovsky, “Exploring Complex Engineering Learning Over Time with Epistemic Network Analysis,” *J. Pre-College Eng. Educ. Res.*, vol. 1, no. 2, p. 4, 2011.

[18] D. W. Shaffer, D. Hatfield, G. Svarovsky, P. Nash, A. Nulty, E. A. Bagley, K. Franke, A. Rupp, and R. Mislevy, “Epistemic Network Analysis: A prototype for 21st century assessement of learning,” *Int. J. Learn. Media*, vol. 1, no. 1, pp. 1–21, 2009.

[19] A. Rupp, M. Gustha, R. Mislevy, and D. W. Shaffer, “Evidence-centered design of epistemic games: Measurement principles for complex learning environments,” *J. Technol. Learn. Assess.*, vol. 8, no. 4, 2010.

[20] J. Saucerman and D. W. Shaffer, “The HandCoder: An automated discourse code developer,” 2015.

[21] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

[22] D. W. Shaffer, D. L. Hatfield, G. N. Svarovsky, P. Nash, A. Nulty, E. A. Bagley, K. Franke, A. A. Rupp, R. J. Mislevy, and K. Frank, “Epistemic Network Analysis: A prototype for 21st century assessement of learning,” *Int. J. Learn. Media*, vol. 1, no. 1, pp. 1–21, 2009.

[23] A. A. Rupp, Y. Y. Choi, M. Gushta, R. J. Mislevy, M. C. Theis, P. Nash, D. L. Hatfield, E. A. Bagley, D. W. Shaffer, and G. N. Svarovsky, “Modeling learning progressions in epistemic games with epistemic network analysis: Principles for data analysis and generation,” *Learning Progressions in Science Conference*. Iowa City, IA, 2009.

[24] D. L. Hatfield, *The right kind of telling: An analysis of feedback and learning in a journalism epistemic game*. Madison, WI: University of Wisconsin-Madison, 2011.

[25] G. Arastoopour and D. W. Shaffer, “Measuring Social Identity Development in Epistemic Games,” in *Computer Supported Collaborative Learning Conference Proceedings*, 2013.

[26] G. N. Svarovsky and D. W. Shaffer, “SodaConstructing knowledge through exploratoids,” *J. Res. Sci. Teach.*, vol. 44, no. 1, pp. 133–153, 2007.