

# Developing a Task Model for Evidence-Centered Assessment Design in a Speech Signal Processing Course

*Shonda Bernadin<sup>1</sup> and Tejal Udhan<sup>2</sup>*

**Abstract** – In this paper, the evidence-centered design (ECD) framework based on Bayesian inference networks is explored to develop structured assessment tasks that accurately measure student proficiency in a *Speech Signal Processing* elective course. Tasks are an important part of assessment design because they provide evidence for constructive feedback, informed decisions and predictive results of student proficiency. Many assessment tasks in engineering courses are not designed in such a way as to adequately identify and address the foundational gaps in student learning and pre-requisite knowledge. ECD provides an assessment framework that is structured and methodical and can help to uncover knowledge gaps in student learning. The results of this research will show the effective design of task assessments that will expose learning gaps and help to improve instructional practice, student proficiency and engagement in an undergraduate electrical engineering course.

**Keywords:** evidence-centered design, assessment methods, Bayesian inference networks

## INTRODUCTION

In this research the term *assessment* is defined as a process that documents and evaluates the knowledge, skills, attitudes and beliefs (KSABs) of learners in measurable and meaningful ways. Educational assessments are key components of all education systems since they significantly contribute to the progression of a student's learning experience. They also inform policymakers on how to evaluate the efficacy of education programs; help administrators decide how to allocate resources; contribute to curriculum design and improvements; and determine a student's mastery of skills [1-3]. Many assessments in engineering courses are not designed in such a way as to adequately identify and address the foundational gaps in student learning and their pre-requisite knowledge. Thus, students are not fully prepared to make completely sound contributions in the field of interest due to their limited knowledge in certain content areas. The focus of this research is on the effective design of course assessments (i.e. tasks) in a combined section of an Electrical and Computer Engineering (ECE) undergraduate/graduate elective course, namely, *Speech Signal Processing*. This targeted course was chosen because it is a new course offering in the ECE curriculum and lends itself quite well to an innovative course design due to its strong application of prerequisite knowledge. Since this course is new, it is unclear how many students would compose an average sample, however, based on other combined sections of ECE elective courses in the program the average student enrollment would typically fall between 15 and 20 students. The goal of this research is to formulate a task model that can be used in future course offerings, and possibly used in other courses. The tasks are designed with the anticipation that they will reveal knowledge gaps in student learning in efforts to improve instructional practice, student proficiency and engagement in the course.

Generally-speaking, there are several types of assessments that can be used to measure KSABs and they are well-documented in literature [1-8]. The most common types are briefly described here for completeness. According to [1], the most common types of assessments include **diagnostic**, which assesses a student's strengths, weaknesses,

---

<sup>1</sup>Florida A&M University-Florida State University-College of Engineering, 2525 Pottsdamer Street, Tallahassee, FL 32310, bernadin@eng.fsu.edu

<sup>2</sup> Florida A&M University-Florida State University-College of Engineering, 2525 Pottsdamer Street, Tallahassee, FL 32310, tu13b@my.fsu.edu

knowledge, and skills a priori instruction, **formative**, which assesses a student's performance during the instructional period and **summative**, which measures a student's performance post-instructional period. In the targeted electrical engineering course, summative assessment is most useful because it measures integrated knowledge at the end of the semester using an evaluation of a capstone project.

Designing an appropriate summative assessment item that clearly and effectively evaluates the achievement of student proficiency is a major task. Most assessment tasks start with a claim about students' knowledge, skills, attitudes, beliefs, performance, proficiency, engagement, etc. For example in the *Speech Signal Processing* course, the typical claim is that by the end of the semester students will (1) know the theory and practical skills to design speech processing algorithms, and (2) be able to design and implement a variety of speech applications. The claims are direct examples of goals that have been highlighted for the course. Implicit to these assumptions is that students have enough prerequisite knowledge and understanding to achieve the goals outlined in the course. However in practice and through informal and formal observations, it remains clear that many students are not adequately prepared to successfully achieve the course goals. Additionally after careful examination of the course goals, they do not clearly and explicitly indicate the underlying assumptions about a student's knowledge, skills and abilities. By explicitly articulating these assumptions, assessment tasks can be designed that are effective at identifying underlying gaps in proficiency and enabling better instructional methods. In assessment design, different techniques can be used that will breakdown implicit assumptions and force explicit evaluation of a claim. This process will help identify learning gaps and highlight bottlenecks in student proficiency in a formal structured way through designing a logical flow of observed data [4].

There are many assessment design methods and theories that have been used to design assessments, including the popular classical test theory and item response theory (IRT) [6]. One common theme among most educational assessments according to [6], is that inferences about students KSABs are made based on assumptions and/or observations. One advantage to using inference-by-observation is that observations can be formalized through assessment tasks which provide *evidence* for assessment claims or arguments. This concept is the premise of Evidence-centered assessment design (ECD). ECD is a framework for designing structured problem-based, critical thinking assessments using evidentiary argumentation and Bayesian inference networks[4-6]. It was chosen for this project because it defines a logical path that corroborates assessment arguments and breaks down implicit assumptions to reveal underlying gaps in student learning. The five-layer process includes a task model that is designed for assessments. In general, task modeling is the creation and development of activities, examinations, essays and other items that influence performance variables. In this paper, the development of a task model in the context of ECD framework will be presented.

The next section describes the Evidence-Centered Assessment Design framework in more detail and includes the implementation of these layered applications to the *Speech Signal Processing* course. A discussion of Bayesian inference networks is given along with its implementation in the elective course. Finally, future work and conclusions sections are given.

## EVIDENCE-CENTERED ASSESSMENT DESIGN FRAMEWORK

The Evidence-Centered Design (ECD) method was developed by Educational Testing Services (ETS) with the researchers in [5] leading the efforts. ECD is an assessment design technique that is based on evidentiary arguments that support student capabilities in a subject content. It provides a structured framework using Bayesian inference network fragments that logically articulates assessment reasoning and claims that determine the successful achievement of a given result. There are five layers in the ECD approach: *domain analysis, domain modeling, conceptual assessment framework, assessment implementation, and assessment delivery*.

In this work, the principles of evidence-centered assessment design (ECD) are applied to a *Speech Signal Processing* course in efforts to improve assessment, instructional effectiveness and ultimately, student proficiency in this course. Layers 1, 2, and 3 are investigated in this paper. Layers 4 and 5 are beyond the scope of this particular research study and are not described here. The interested reader is referred to [6] for more details involving these two layers. The following sections summarize each layer as described by authors in [4].

**Domain Analysis (Layer 1)**

Domain analysis is the first layer of ECD in course assessment. This is the information-gathering layer in which content and contextual information about the subject domain of interest is collected. Several components are explored in domain analysis including content knowledge, concepts and terminology, tools, situational knowledge (which includes knowledge about actions, behaviors and appropriate responses by students in practical situations), and cognitive analysis (which indicates how people use their knowledge to gain understanding). Implicit to the components defined is the interpretation of the nature of knowledge and how it is acquired and used in their environment. This element of assessment represents a psychological perspective of knowledge because it defines an inherent self-learning process that determines the validity and value of knowledge with respect to student capabilities [4]. In general, the information from domain analysis identifies categories of information which become inputs to the next layer in ECD process. The basic categories of domain of interest are

- i. Valued work involves surveying the professionals in the field to understand the knowledge and behaviors that are used in the domain.
- ii. Task features are the aspects of the work that are recurring and salient in the domain of interest.
- iii. Representational forms of information are the symbols or parameters that are associated with the domain including graphical forms of data.
- iv. Performance outcomes define indicators that determine when an appropriate level of knowledge is reached.
- v. Valued knowledge is the content that is needed in the domain of interest, which can be gathered from several sources including textbook and curriculum materials.
- vi. Knowledge structure and relationships identifies patterns or connections among valued knowledge that may provide insight to how knowledge is developed in learners.
- vii. Knowledge-task relationships indicate how task features interact with knowledge differences in learners.

The implementation of how layer 1 is applied to the *Speech Signal Processing* course is given in table 1.

**Table 1-Domain Analysis for Speech Signal Processing Course**

	<b>CATEGORY</b>	<b>CASE STUDY EXAMPLE</b>
i)	Valued Work	What behaviors do speech researchers use in conducting their research?
ii)	Task Features	What important, dominant tasks or skills do speech researchers use or implement on a recurring basis?
iii)	Representational Forms	What are the basic units, symbols or graphical notations that are used in speech processing research?
iv)	Performance Outcomes	When a student effectively communicates in a manner similar to that of the professional then an appropriate level of knowledge has been attained.
v)	Valued Knowledge	What content knowledge in signal processing is needed in the speech processing domain?
vi)	Knowledge Structure and Relationships	Are there themes or patterns within speech processing content knowledge that are useful for developing professional skills?
vii)	Knowledge Task Relationships	Do students use task features efficiently in professional speech analysis research settings regardless of level of content knowledge attained?

**Domain Modeling (Layer 2)**

Domain modeling is the second layer in ECD process. It articulates the elements that are needed in assessment and defines a clear logical structure for the assessment argument. Typically, the domain modeling layer considers an assessment design pattern to help build a structure for the assessment argument or claim. Toulmin’s argument structure as presented by authors in [4-6] provides an insightful viewpoint for developing the assessment argument. Consider an adaptation of Toulmin’s basic structure for argument [9] as shown in Figure 1. As described by [4] and summarized here: *the data (D) leads to (i.e. infers) a claim (C) based on some observation or warrant (W) through some type of a priori knowledge/theory or backing (B). Also, C can be validated by some alternate justification (A) based on observations from rebuttal evidence (R).* Common verbage used to state the argument based on dataflow is given in quotations next to arrows.

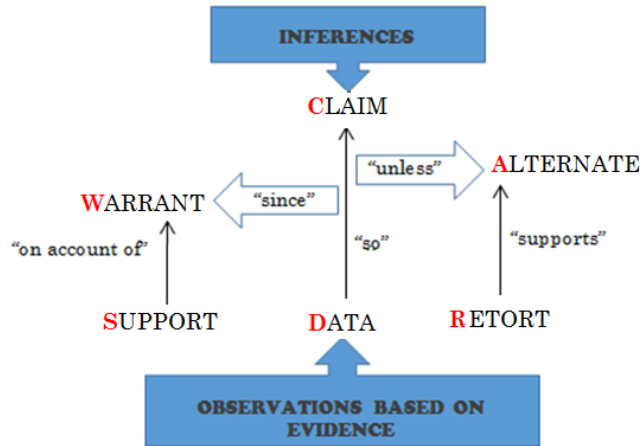


Figure 1-Adaptation of Toulmin's [9] Basic Structure for Argumentation- *the data (D) leads to (i.e. infers) a claim (C) based on some observation or warrant (W) through some type of a priori knowledge/theory or support (S). Also, C can be validated by some alternate justification (A) based on observations from a retort or rebuttal evidence (R).*

For example, a simple assessment argument for the *Speech Signal Processing* course can be made using Toulmin’s argument structure as shown in Figure 2. The assessment argument may read similar to the following:

*“Student X clearly articulates a well-defined problem statement and thoroughly investigated successful problem solution to the capstone speech analysis problem; **so** Student X gained the appropriate and relevant knowledge and participation-level in the speech signal processing domain; **since** Student X exhibited the appropriate content knowledge, research abilities and professional skills that are needed in the speech signal processing profession.”*

In domain modeling a formal detailed design pattern should also be constructed such that attributes of the assessment tasks and their associated values within the context of the assessment argument are mapped to corresponding assessment items. There are several different design patterns that can be used such as “Model Elaboration”, “Analyzing Data Quality”, and “Self-monitoring” which are all examples from the PADI-Principled Assessment Design for Inquiry project [7,8]. A formalized design pattern is not the focus of this research and thus is not included in this paper.

**Conceptual Assessment Framework-CAF (Layer 3)**

The conceptual assessment framework (CAF) is the third layer of the ECD model. Similar to layer 2, it highlights the details of the assessment argument by operationalizing claims using observable variables. CAF uses symbolic modeling and statistical theory to define claims and to map out a data path that justifies the claims. This operationalization will reveal the technical details required for the next ECD layer- *assessment implementation (layer 4)*. To help operationalize the assessment structure, the CAF uses three models: *Student Model*, *Evidence Model* and *Task Model*. Each model will be briefly discussed here and as before, a more complete description of the CAF design can be found in references [4-6].

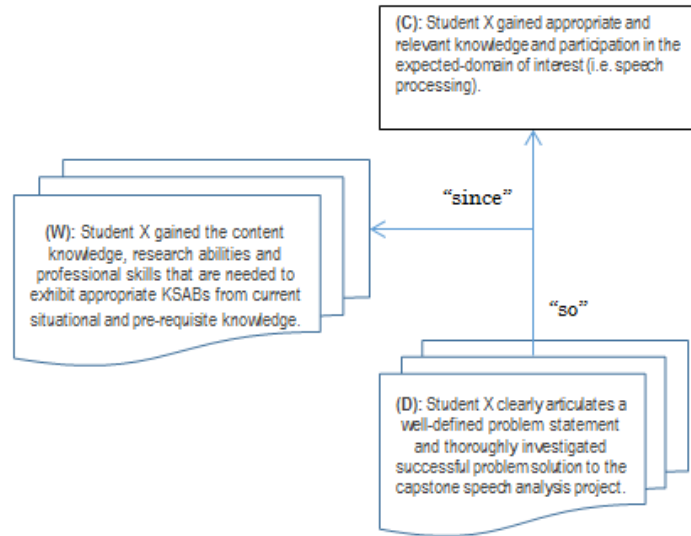


Figure 2-Example of basic structure of assessment argument for *Speech Signal Processing* course.

- i. **Student Model** – The student model will link student performance of tasks to the claims that were made about student proficiency using variable notation. There can be one variable  $\{\theta\}$  or many variables  $\{\theta_1, \theta_2, \theta_3, \dots\}$  to represent claims on student proficiency. The student model answers the question, “*What is being measured?*”. Probability distributions can be used applied to these variables to evaluate what is known about student performance at a given point in the learning process [10].
- ii. **Evidence Model** –The evidence model answers the question, “*How is it going to be measured?*”. It contains an evaluation component and a measurement component to corroborate claims made on student proficiency. The evaluation component determines the criteria for how the values of observable variables are determined. An example of an evaluation component is a scoring rubric or an answer key. The measurement component quantifies the degree to which an observable variable contributes to the claim or assessment argument. It is typically characterized as a weighting measurement with some directional component identified. There are many formulations of measurement models based on probabilistic methods [11-14]; however, one of the most common implementations is using weighted or unweighted scores over items in the scoring rubric.
- iii. **Task Model** – The task model answers the question, “*Where is it going to be measured?*”. In the task model, the actual activities, sometimes called work products, are developed in the context of a learning environment that encompasses the components needed to complete the assessment tasks. Work products can include multiple-choice items, short essay questions, examinations, course projects, etc. In this model the placement of the assessment task is important to the outcome of the observation since it represents a point at which knowledge should be gained. As mentioned previously, the common types of assessment items include assessment at the beginning of the learning period (diagnostic); during the learning period (formative) and at the end of the learning period (summative). An evaluation of a combination of these assessments will give a more comprehensive perspective of student performance and can expose knowledge gaps more clearly and concisely.

This paper highlights the initial development of a Bayesian Inference Network, i.e. Bayes Net, for student proficiency in the *Speech Signal Processing* Course, as described in the next section.

### DEVELOPMENT OF BAYES NET FOR THE TASK MODEL

In the development of course level assessments for the *Speech Signal Processing* course, the general pattern of tasks for assessing student performance initiates with course goals and student-level outcomes. The following course goals for the *Speech Signal Processing* course have been identified as symbols, *G1* and *G2*.

*G1* = {provides theoretical knowledge and practice for designing speech algorithms

2014 ASEE Southeast Section Conference

$G2 = \{enables\ students\ to\ design\ and\ implement\ a\ variety\ of\ speech\ applications\}$

The student-level outcomes have similar symbolic descriptions:  $S1, S2, S3$  and  $S4$

$S1 = \{Increase\ basic\ scientific\ and\ engineering\ knowledge\ of\ speech\ production\ and\ perception\}$

$S2 = \{Implement\ different\ representations\ of\ the\ speech\ signal\}$

$S3 = \{Design\ speech\ algorithms\ that\ estimate\ basic\ speech\ parameters\}$

$S4 = \{Design\ and\ implement\ a\ practical\ speech\ application\}$

In a typical approach to assessment design, the assessment map in Figure 3 may be developed.

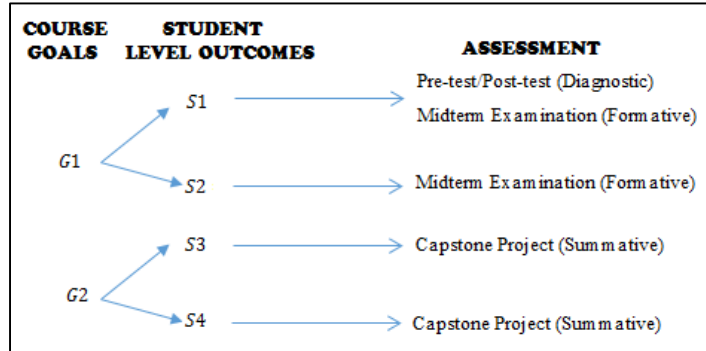


Figure 3-Typical Assessment Map for Speech Signal Processing

Assessment items are extensions of course activities which directly map to student-level outcomes, which in turn, are fed back into the course goals to determine level of achievement. For example, a diagnostic assessment (e.g. pre-test and post-test) can be used to determine if *students have increased their basic scientific and engineering knowledge in speech perception and production (S1)*. Implicit to this outcome; however, is that students' should have proficient understanding of signal processing concepts and acoustic theory; of which both of these topics encompass a broad spectrum of knowledge. Furthermore, summative assessment (i.e. the capstone project) assumes that students have a strong foundation in content knowledge (signal processing and acoustic theory), professional skills (written and oral communication skills) and research abilities (develop a problem statement and problem solution based on the observations). These three categories are not explicitly defined in the course goals or outcomes, but they represent observable variables that are imperative to the level of student proficiency and engagement in this course. To group them into the broad category defined by  $G2$  is a great disservice to the understanding of the underlying knowledge gaps and inhibits effective instruction and performance. The advantage of ECD is that it breaks down these implicit assumptions and *forces* explicit evaluation of observed variables (i.e. student proficiency).

In order to apply the ECD task model the following common notation is defined. The overall proficiency variable is given as  $\theta$  which is represented as a function of key foundational variables:  $\theta_A, \theta_B$  and  $\theta_C$ :

$$\theta = \{\theta_A, \theta_B, \theta_C\}$$

where  $\theta_A$  is the *content knowledge* variable,  $\theta_B$  is the *research abilities* variable and  $\theta_C$  is the *professional skills* variable. The overall claim ( $\theta$ ) is that "by the end of the semester a student will have an appropriate level of content knowledge, research abilities and professional skills to successfully articulate a well-defined problem statement and implement a suitable problem solution for a capstone design project". The claim on content knowledge ( $\theta_A$ ) is that "by the end of the semester a student will have an appropriate level of knowledge of signal processing concepts (including continuous-time and discrete-time signals, sampling, linear-time invariant systems, spectral analysis, finite-impulse response (FIR)/infinite-impulse response (IIR) filtering, frequency response, harmonic analysis, resonance) and have an appropriate level of knowledge of acoustic theory (including pressure waves, loudness, intensity, pitch, decibels (dB), Sound Pressure Level (SPL) Scale, resonators, phonemes, phonetic representation of words) to be able to articulate a well-defined problem statement".

The claim on research abilities ( $\theta_B$ ) is that “by the end of the semester a student will have gained an appropriate level of situational knowledge to be able to select a research project in the domain of interest and be able to solve the problem using an engineering design process (basic steps include problem definition, background research, determine optimal solution, implement solution, test and redesign)”. The claim on professional skills ( $\theta_C$ ) is that “by the end of the semester, a student will have exhibited professionalism appropriate to the targeted domain through effective communication and collaboration. Using these definitions, a fragmented Bayesian Inference Network can be developed to form a logical path towards assessment argument validation based on observed data or evidence.

Bayesian inference networks (Bayes Nets) are graphical models that can be used to make inferences about observations in data [9]. They are based on the well-known Bayes’ Rules for probabilistic inference using conditional probabilities. For example, consider the probability of variable  $Y$  given an observed variable  $X$ ,  $P(Y|X)$ , is

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

where  $P(X|Y)$  is the probability of observed variable  $X$  given  $Y$ ,  $P(Y)$  is the prior probability of  $Y$  before  $X$  is observed and  $P(X)$  is the probability of  $X$ . If variable  $Y$  represents a claim about student proficiency ( $\theta$ ) and variable  $X$  represents the observed behavior or performance on a task item, then the conditional probability becomes

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

In this probabilistic model,  $P(\theta)$  and  $P(X)$  will be estimated from historical data from prerequisite courses. A Bayes Net can be developed that maps the claims about student proficiency ( $\theta$ ) to the observed behavior ( $X$ ) as shown in Figure 4. Notice that the overall claim of proficiency ( $\theta$ ) maps into three individual claims about student performance ( $\theta_A, \theta_B$  and  $\theta_C$ ) which in turn, maps into subsequent claims ( $\{\theta_{A1} \dots \theta_{An}\}, \{\theta_{B1} \dots \theta_{Bn}\}$  and  $\{\theta_{C1} \dots \theta_{Cn}\}$ ). The observed performance data  $\{X_{A1} \dots X_{An}\}, \{X_{B1} \dots X_{Bn}\}$  and  $\{X_{C1} \dots X_{Cn}\}$  is gathered from activities such as midterm examinations, homework assignments, group assessments, capstone projects and is used to inform the decision criteria. Based on the resulting decision, a corresponding action is triggered.

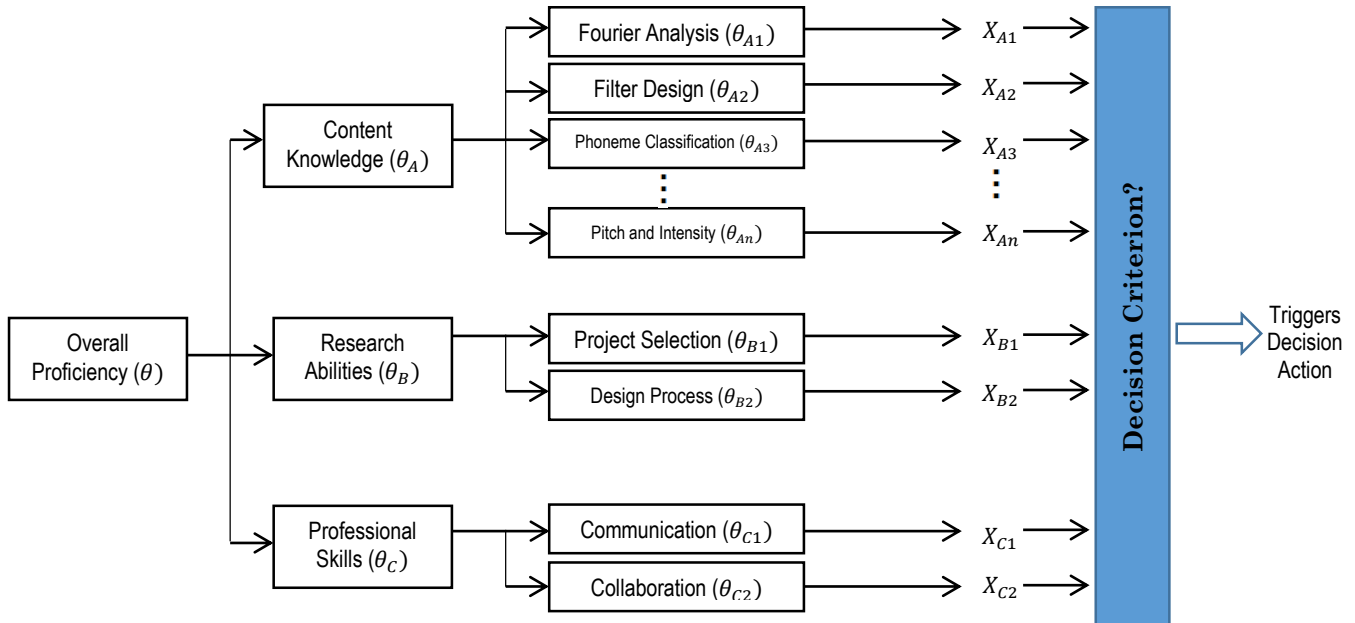


Figure 4-Bayes Net fragment for Student Proficiency variable in *Speech Signal Processing*

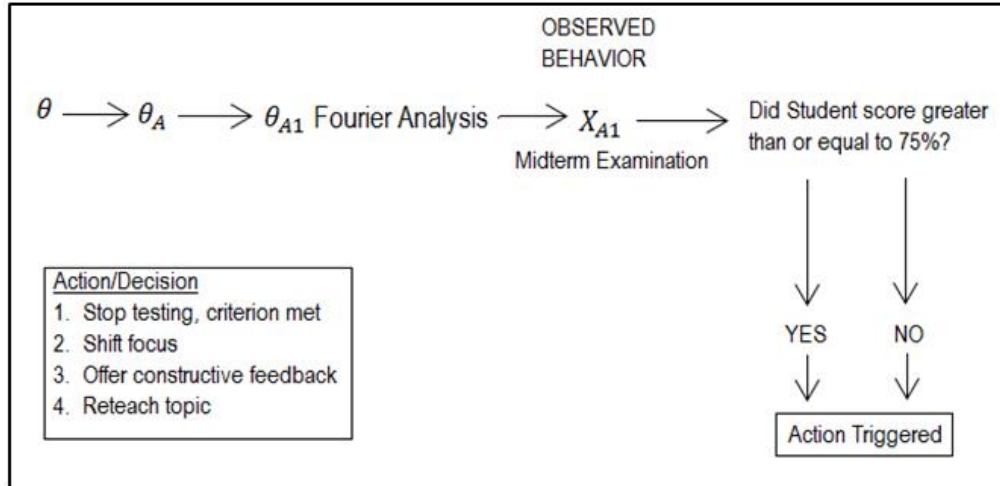


Figure 5-Bayes Net datapath of one assessment claim and decision action triggered

An example of the decision criteria and corresponding action is highlighted for one datapath in figure 5 to illustrate the concept. For example, let's consider the student proficiency in Fourier analysis which is claimed as  $\theta_{A1}$ . The evidence is observed in student performance on a midterm examination. Based on the student's performance on the Fourier analysis problem, an action is triggered and a decision is made. The inset in Figure 5 shows typical actions that may result as a consequence of the performance.

A fragmented Bayes Net task model has been constructed from the goals outlined in the *Speech Signal Processing* course. The underlying assumptions about student proficiency were exposed to highlight potential knowledge gaps that may occur as a result of observed data.

### FUTURE WORK

This paper highlights the design of a task model using Bayes inference networks. The next step in this research involves implementation of the task model to gather actual data and analyze the results. The task model is just one part of one layer in the ECD framework. Other future work will explore several aspects of the overall assessment design including the investigation of assessment implementation and assessment delivery, ECD layers 4 and 5 respectively; the exploration of assessment design patterns in domain modeling and a detailed analysis of evidence modeling using probability-based measurement models. This research provides a rich platform for investigating knowledge gaps in students' learning foundations using probability-based analytical methods.

### CONCLUSIONS

In this paper, the evidence-centered design (ECD) framework based on Bayesian inference networks was presented and used to design a task model that defines a data path that logically and accurately measures student proficiency in a *Speech Signal Processing* elective course. Since task assessments can be used to gather information and justify arguments about student proficiency, the work presented in this paper helps to define a probability-based analytical framework to identify and expose the level of implicit knowledge that is required for student proficiency in this course. Future work describes aspects of ECD that can be further expanded to develop a complete ECD model for measuring student proficiency.



REFERENCES

- [1] Lepi, Katie, *The 6 types of Assessments (and how they are changing)*, July 22, (2013), Edudemic:Conneting Education and Technology, from <http://www.edudemic.com/the-6-types-of-assessments-and-how-theyre-changing/>
- [2] Mislevy, Robert J. "Linking Educational Assessments: Concepts, Issues, Methods, and Prospects." (1992).
- [3] Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). *On the structure of educational assessments*. Measurement: Interdisciplinary Research and Perspectives, 1, 3–67.
- [4] Mislevy, Robert J., and Michelle M. Riconscente. "Evidence-centered assessment design: Layers, structures, and terminology." *Menlo Park, CA: SRI International* (2005).
- [5] Mislevy, Robert J., Linda S. Steinberg, and Russell G. Almond. *On the roles of task model variables in assessment design*. National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation, Graduate School of Education & Information Studies, University of California, Los Angeles, (1999)
- [6] Mislevy, Robert J., Russell G. Almond, and Janice F. Lukas. "A brief introduction to evidence-centered design." *RESEARCH REPORT-EDUCATIONAL TESTING SERVICE PRINCETON RR 16* (2003).
- [7] Riconscente, Michelle M., Robert J. Mislevy, and Lawrence Hamel. "An introduction to PADI task templates." *PADI Technical Report 3* (2005).
- [8] Mislevy, R. "Overview of the principled assessment design for inquiry (PADI) project." *annual meeting of the American Educational Research Association, Chicago, Illinois. 2003.*
- [9] Toulmin, Stephen E. *The uses of argument*. Cambridge University Press, (1958/2003).
- [10] Mislevy, Robert J., Russell G. Almond, Duanli Yan, and Linda S. Steinberg. "Bayes nets in educational assessment: Where the numbers come from." In *Proceedings of the fifteenth conference on uncertainty in artificial intelligence*, pp. 437-446. Morgan Kaufmann Publishers Inc., (1999).
- [11] Heckerman, David, and John S. Breese. "Causal independence for probability assessment and inference using Bayesian networks." *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 26.6 (1996): 826-831.
- [12] Schum, David A. *The evidential foundations of probabilistic reasoning*. Northwestern Univ Press, (1994).
- [13] Keppens, Jeroen. "Towards qualitative approaches to Bayesian evidential reasoning." In *Proceedings of the 11th international conference on Artificial intelligence and law*, pp. 17-25. ACM, (2007).
- [14] Mislevy, Robert J. "Evidence and inference in educational assessment." *Psychometrika* 59, no. 4 (1994): 439-483.

## **2014 ASEE Southeast Section Conference**

### **Shonda L. Bernadin**

Shonda L. Bernadin is an Associate Professor in the Electrical and Computer Engineering Department the Florida A&M University-Florida State University College of Engineering. Dr. Bernadin received her B.S. degree in Electrical Engineering from Florida A&M University in 1997, her M.S. degree in Electrical and Computer Engineering from University of Florida in 1999, and her Ph.D. degree in Electrical Engineering from Florida State University in 2003. She is currently the founding director of the Speech Processing and Data Analysis Laboratory (SPADAL). Her research interests include speech analysis and pattern recognition, feature extraction, data mining, instructional design and engineering education.

### **Tejal Udhan**

Tejal Udhan graduated from Marathawada Institute of Technology (Aurangabad, India) with a B.E. in Electronics and Communication. Ms. Udhan is currently pursuing a doctoral degree in Electrical Engineering at Florida State University (FSU) and is a research assistant in the Speech Processing and Data Analysis Laboratory. Her research interests include speech processing systems, data mining and pattern recognition techniques.