Proceedings of the ASME 2020 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2020
August 16-19, 2020, St. Louis, MO, USA

IDETC2020-22613

QUANTIFYING INDIVIDUALS' THEORY-BASED KNOWLEDGE USING PROBABILISTIC CAUSAL GRAPHS: A BAYESIAN HIERARCHICAL APPROACH

Atharva Hans
School of Mechanical Engineering
Purdue University
West Lafayette, IN

Ashish M. Chaudhari
School of Mechanical Engineering
Purdue University
West Lafayette, IN

Ilias Bilionis
Assistant Professor
School of Mechanical Engineering
Purdue University
West Lafayette, IN

Jitesh H. Panchal
Associate Professor
School of Mechanical Engineering
Purdue University
West Lafayette, IN

ABSTRACT

Extracting an individual's knowledge structure is a challenging task as it requires formalization of many concepts and their interrelationships. While there has been significant research on how to represent knowledge to support computational design tasks, there is limited understanding of the knowledge structures of human designers. This understanding is necessary for comprehension of cognitive tasks such as decision making and reasoning, and for improving educational programs. In this paper, we focus on quantifying theory-based causal knowledge, which is a specific type of knowledge held by human designers. We develop a probabilistic graph-based model for representing individuals' concept-specific causal knowledge for a given theory. We propose a methodology based on probabilistic directed acyclic graphs (DAGs) that uses logistic likelihood function for calculating the probability of a correct response. The approach involves a set of questions for gathering responses from 205 engineering students, and a hierarchical Bayesian approach for inferring individuals' DAGs from the observed responses. We compare the proposed model to a baseline three-parameter logistic (3PL) model from the item response theory. The results suggest that the graph-based logistic model can estimate individual students' knowledge graphs. Comparisons with the 3PL model indicate that knowledge assessment is more accurate when quantifying knowledge at the level of causal relations than quantifying it using a scalar ability parameter. The proposed model allows identification of parts of the curriculum that a student struggles with and parts they have already mastered which is essential for remediation.

1 Introduction

Scientific knowledge is central to engineering design. Designers use the knowledge of scientific theories such as mechanics of materials, thermodynamics, and controls to make decisions about concepts, materials, and manufacturing processes. Designers use scientific knowledge to generalize from an experiment to the real world applications. Scientific knowledge is also an important ingredient of engineering design expertise. This significance makes quantification of knowledge structures of individuals essential for understanding design cognition and for designing products and systems that mimic humans [1].

Quantification of individual-specific knowledge structures is crucial from the perspective of both design research and education. Within design research, the availability of individual specific knowledge structures can complement the descriptive models of designers' decisions [2, 3]. The knowledge structures can be particularly helpful in accurately describing the prior beliefs of decision makers. The detailed models of prior knowledge can provide a better understanding of how designers use background knowledge for inductive and deductive reasoning tasks in design [4, 5]. From the educational perspective, a detailed quantitative representation of knowledge structures can support accurate assessment of students' knowledge [6], development of improved educational support tools [7], and development of educational interventions for flexible personalized learning environments.

Despite extensive research on knowledge representation in diverse fields such as computer science, engineering design, and psychometrics, there is a lack of approaches for extracting an individual's knowledge structure. On one hand, knowledge representation models such as those based on function-behavior-structure (FBS) [8] and core [9] models (discussed in Section 2.1) capture the structure but are meant for computational designs support only. Further, these models do not capture scientific theories. On the other hand, psychometric approaches (discussed in Section 2.2) are focused on quantifying human knowledge and skills, but do not capture the details of the knowledge structure of individuals. This gap introduces a need for quantitatively modeling individuals' knowledge structures at a detailed level. We address this need for a specific type of knowledge - *theory-based causal knowledge*.

Our approach is based on a probabilistic graph-based model for representing an individual's concept-specific causal knowledge for a given theory, and employs Bayesian inference for estimating graphs for individuals from their responses to a questionnaire. First, we assume that each individual has a latent knowledge graph that is a subgraph of the ideal theory graph. That is, we assume that the subjects do not make up new causal links, but they may not know some of the real causal links. Second, our model predicts the probability of a correct response to a theoretical question conditional on the causal links that are relevant to the given question. Third, we use Bayesian inference to estimate the posterior over the individual knowledge graphs conditioned on each subject's answers to a series of theoretical questions. We illustrate the approach using the responses of mechanical engineering students to a set of questions involving shaft design problems to quantify their knowledge of fatigue failure.

The results highlight the advantages of causal graphs for accurately quantifying theory-based knowledge and for correctly predicting individuals' test responses. Compared to the commonly used three-parameter logistic (3PL) model from the item response theory [10], the proposed model has better posterior predictive accuracy at the levels of aggregate, each question and individual student. The proposed model also allows identification of parts of the curriculum that a student struggles with and parts they have already mastered which is essential for providing feedback.

The paper is organized as follows. We begin with a review of relevant work from engineering education and design in Section 2 and describe their role in motivating the proposed methodology. Section 3 provides mathematical details of the proposed approach. Section 4 presents the data collection approach implemented in an introductory machine design class. In Section 5, we present key results from the analysis. The implications of the proposed methodology are discussed from engineering education, design and practice in Section 6. Section 7 summarizes the key conclusions.

2 Review of Literature

2.1 Knowledge Representation in Engineering Design

Many studies investigate the role of design knowledge for understanding design creativity and problem framing. They use the function-behavior-structure (FBS) framework as a theoretical basis for representing processes of creativity and conceptual design [8]. The FBS framework separates function and structure to emphasize the role of iteration over prior knowledge of the design requirements (i.e., the function). Extensions of the FBS framework have been proposed which allow preliminary structures of artifact to form based on the design requirements without iterations [11]. However, the existing theoretical frameworks do not shed light on a designer's prior knowledge which clearly plays an important role.

Other studies undertake computational approaches for representing knowledge of design processes and design artifacts, e.g., in the product systems design [12, 13]. The goal of these computational studies is to discover generalized and specialized product knowledge from design databases for supporting tool development for improved analogical design. Dong and Sarakar [14] represent complex products and processes as matrices where nodes are product elements and cells are structural, functional or behavioral relationships between nodes. Then, they derive generalized design knowledge as the macroscopic level information from matrix representations using singluar value decomposition. With the goal of quantifying a product's innovativeness in terms of component-level decisions, Rebhuhn et al. [15] represent the product design process as the hierarchy of product, function, and components. They use multi-agent models to propagate novelty scores of products down to the component level. For understanding functional and surface relatedness between products, Fu at al. [16] analyze the US patent database and discover different structural forms such as hierarchy and ring. Despite this development, computational approaches for representing and estimating an individual's theory-driven causal knowledge are lacking. The proposed methodology addresses this gap by modeling theory-specific causal knowledge as a probabilistic causal graph and estimating person-specific causal graphs using Bayesian inference.

2.2 Psychometric Approaches for the Measurement of Knowledge

Wide ranging techniques such as interviews, protocol analysis, case studies, decision analysis and behavioral experiments are employed in the literature for knowledge elicitation [17]. These methods are broadly categorized into formal and informal methods. The informal methods of knowledge elicitation require moderation from an elicitor agent (human) [18], whereas the formal methods use procedural and analytical methods to reduce interference [19]. Given the quantitative goal of measuring knowledge, the scope of elicitation techniques relevant to this paper is formal.

Student modeling is an area where the formal approaches for knowledge measurement are extensively studied [20]. Most of these studies are based on the item response theory (IRT) which allows assessment of students' ability using psychometrics. An example of psychometric assessment is the force concept inventory which tests Newtonian concepts along six dimensions such as kinematics, impetus, active force, action-reaction pairs, concatenation of influence, and other influences such as centrifugal forces [21]. The basic structure of the item response theory models involves defining the probability of correct response as a function of single or multi-dimensional ability variable, and question (item)-specific parameters [10]. The models are differentiated based on the number of question-specific parameters, either one, two, or three parameters. Then, the model parameters are estimated from students' observed responses using maximum likelihood estimation methods. Multi-dimensional IRT (MIRT) models that represent ability using more than one dimension are the state-of-the-art in the IRT models [6,22,23]. However, most MIRT models assume that all ability dimensions are required for answering any question correctly, and the probability of correct response increases with any dimension (monotonicity assumption). They also assume that the responses of different questions are independent (local independence assumption). But in the context of engineering design, selective concepts are required to answer a specific question correctly and the responses are correlated because the underlying knowledge dimensions are the same.

2.3 Item Response Theory: Three-parameter Logistic Model

Given the similar knowledge elicitation techniques between the proposed approach and the item response theory (IRT), we use a popular three-parameter logistic (3PL) IRT model as a baseline for comparison. The 3PL model represents dichotomous data where a response is either correct or incorrect.

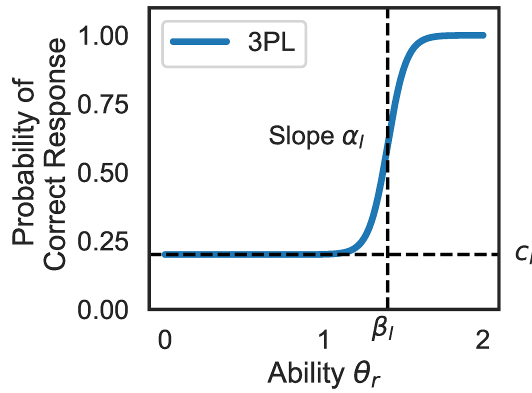


FIGURE 1: Graphical representation of the three parameter logistic (3PL) model.

Assume that R individuals are tested on L problems. The 3PL model has one person-specific ability parameter θ_r, $r = 1,\dots,R$ and L problem-specific parameters:

1. *Problem discrimination* α_l*:* This measures how the probability of answering a question correctly changes with ability.
2. *Problem difficulty* β_l*:* This is a measure of problem difficulty based on the ability required to get the correct answer. A higher ability required to solve a given problem corresponds to a greater problem difficulty.
3. *Pseudo-guessing parameter* c_l*:* This accounts for the probability of getting a correct answer by guessing in a multiple choice problem and is inversely proportional to the number of possible correct answers.

We denote with E_{rl} the answer that individual r gave to question l. The probability of a correct answer (likelihood) is:

$$p(E_{rl} = 1|\theta_r, \alpha_l, \beta_l, c_l) = c_l + (1 - c_l)\,\mathrm{sigm}\left(\alpha_l\left(\theta_r - \beta_l\right)\right), \quad (1)$$

where $\mathrm{sigm}(x) = 1/(1+e^x)$ is the sigmoid function. A graphical representation of the likelihood function under 3PL model is shown in Figure 1. The model is typically trained by maximizing the log-likelihood of all observations, i.e., the sum of the logarithms of Eq. (1). In this work we opt for a Bayesian approach. Assuming that the ability parameter θ_r is a real number, slope α_l is positive and threshold β_l is positive, we use the following priors:

$$\begin{aligned} \theta_r &\sim \mathrm{Normal}(0,1) \\ \alpha_l &\sim \mathrm{Lognormal}(0,1) \\ \beta_l &\sim \mathrm{Lognormal}(0,1) \\ c_l &= 0. \end{aligned} \quad (2)$$

We do not consider pseudo-guessing, $c_l = 0$, because the data utilized to train the model were from a written exam. We talk more about data collection in Section 4.

3 Methodology

3.1 The definition of a knowledge graph

To represent an individual's state of knowledge about engineering concepts, we utilize **graph-based representations** of domain-specific knowledge. Specifically, the representations we use are based on *directed acyclic graphs*, and are suitable for representing causal knowledge, e.g., see Figure 5 which represents the causal knowledge for the theory of fatigue failure.

Let $X = \{x_1, x_2, \ldots, x_N\}$ be the set of physical variables relevant for a given engineering theory. A physical variable can be a discrete or real-valued variable. These physical variables are related to each other through structural equations. In these equations some variables are inputs and some are outputs and the interpretation is that the input variables are causing the output variables. We say that this input-output relationship that appears in the equations is a causal relationship. Putting aside the specific equations, these causal relationships can be represented with an acyclic directed graph, termed the knowledge graph. The true knowledge graph for a specific engineering problem, can be represented as an $N \times N$ binary matrix, $K^{\text{True}} = \{k_{ij}^{\text{True}}\}$, where k_{ij}^{True} is 1 if the variable x_i is a direct cause of x_j and 0 otherwise.

3.2 Prior over knowledge graphs of individuals

Let the $N \times N$ matrix $K_r = \{k_{r,ij}\}$ represent the r-th person's knowledge about the causal links using the same encoding as in K^{True}. We assume that a person's knowledge graph is always a subgraph of the true knowledge graph of the theory. This means that if the theory has no direct link from x_i to x_j, then a person does not make up such a link. This same assumption means that we are only going to focus on whether or not a person has identified correctly the true causal links. Prior to making any observations, we model our belief that individual r knows about the existence of a true link between x_i and x_j by:

$$p(k_{r,ij} = 1 | k_{ij}^{\text{True}} = 1, a_{r,ij}) = a_{r,ij}, \tag{3}$$

where $a_{r,ij}$ is a hyper-parameter taking values in $[0,1]$. Similarly, the probability that the person does not know the causal link exists is $p(k_{r,ij} = 0 | k_{ij}^{\text{True}} = 1, a_{r,ij}) = 1 - a_{r,ij}$. Given the matrix of prior link probabilities $A_r = \{a_{r,ij}\}$, the prior over the causal graph of individual r is:

$$\begin{aligned} p(K_r | K^{\text{True}}, A_r) &= \prod_{i,j:k_{ij}^{\text{True}}=1} p(k_{r,ij} | k_{ij}^{\text{True}} = 1, a_{r,ij}) \\ &= \prod_{i,j:k_{ij}^{\text{True}}=1} a_{r,ij}^{k_{r,ij}} (1 - a_{r,ij})^{1-k_{r,ij}}. \end{aligned} \tag{4}$$

The reader should notice that the product is only over the true causal links.

3.3 The data likelihood

In contrast to the IRT, our model requires detailed knowledge about the subgraph of the true knowledge graph that each question tests. Each question is framed in terms of a given set of design parameters and a set of a single performance parameter to be evaluated. A person answers the question by providing a value of the performance parameter. The knowledge required to answer question l is part of the knowledge graph that connects the design parameters to the performance parameter. Mathematically, we can get the relevant subgraph from the knowledge graph using an $N \times N$ *reduction* matrix Q_l, whose cell value $q_{l,ij}$ is 1 if variable x_i and x_j belongs to the set of relevant design parameters and zero otherwise. Then, the true knowledge subgraph for question l is the Hadamard product (elementwise product) of the reduction matrix Q_l and the true knowledge graph K^{True}, denoted as $Q_l \circ K^{\text{True}}$. In the matrix $Q_l \circ K^{\text{True}}$ irrelevant variables have been replaced by zeros. We assume that r-th individual's response to question l depends only on the relevant subgraph $Q_l \circ K_r$.

To proceed, we postulate that the probability that an individual's response is correct is a function of on how close that individual's relevant knowledge subgraph is to the true relevant knowledge subgraph. In other words, we propose that the fraction of relevant links that a person correctly identifies is representative of the person's problem-specific ability. The number of relevant links is $\| Q_l \circ K^{\text{True}} \|_F$, where $\| B \|_F = \sum_{i,j} b_{ij}$ is the Frobenius matrix norm. Notice since the individual's knowledge graph is a subgraph of the true graph, the number of correctly matched links is simply $\| (Q_l \circ K_r) \circ (Q_l \circ K^{\text{True}}) \|_F$. Thus, the fraction of correctly identified links is:

$$\phi(Q_l \circ K_r, Q_l \circ K^{\text{True}}) := \frac{\| (Q_l \circ K_r) \circ (Q_l \circ K^{\text{True}}) \|_F}{\| (Q_l \circ K^{\text{True}}) \|_F}. \tag{5}$$

Finally, we model the probability of individual r giving the correct response to question l as:

$$\begin{aligned} &p(E_{rl} = 1 \mid Q_l, K_r, c_l, \alpha, \beta, K^{\text{True}}) \\ &\quad = c_l + (1 - c_l) \operatorname{sigm}\left(\alpha(\phi(Q_l \circ K_r, Q_l \circ K^{\text{True}}) - \beta\right), \end{aligned} \tag{6}$$

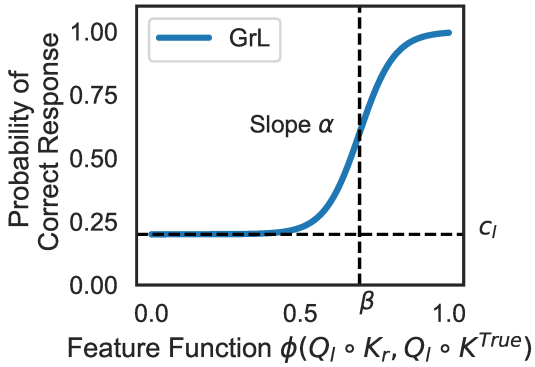


FIGURE 2: Graphical representation of the graph based logistic (GrL) model.

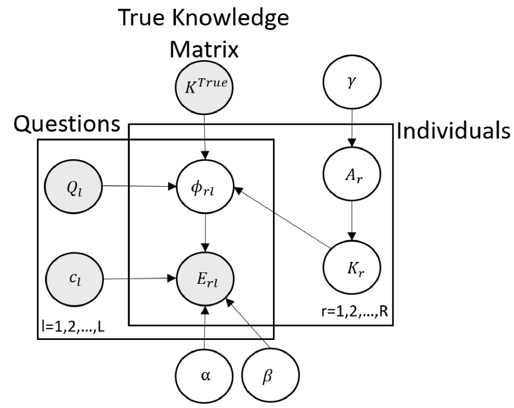


FIGURE 3: The plate-notation diagram for Bayesian hierarchical representation of the graph based logistic (GrL) model. The filled nodes represent the observed variables.

where c_l is a problem-specific parameter capturing the probability of guessing the correct answer, while α and β are global hyper-parameters determining the slope and the threshold of the sigmoid, respectively. We call this likelihood function as the *graph based logistic* (GrL) model. Finding that α is of the order of 10 and β close to 1 would be an indication that our model works as expected. A value for α of order of 10 signifies a steep change in the probability of correctness of response as a function of fraction of correctly identified links. Further, β close to 1 would imply that in order to correctly answer a given question, an individual needs to know all the relevant links for that question correctly.

See Figure 2 for a visualization of the probability of answering correctly as a function of ϕ.

Assuming zero chances of pseudo-guessing, we take $c_l = 0$ for all questions. Since α can only take positive values and β should be close to 1, the priors α and β are as follows:

$$\begin{aligned}
&\alpha \sim \text{Lognormal}(1,1), \\
&\beta \sim \text{Beta}(5, 0.1), \\
&c_l = 0.
\end{aligned}$$

3.4 Assigning hyper-priors

To complete the model we need to assign a hyper-prior on the link probability matrices A_r. The hyperpriors represent our beliefs about the population's knowledge about link probabilities. For a link between variables i and j, since $a_{r,ij}$ is between 0 and 1, and γ_{ij} is always positive the hyperprior is defined on hyper-parameter γ_{ij} as follows, assuming that the link exists in the true knowledge graph:

$$\begin{aligned}
&\gamma_{ij} \sim \text{Lognormal}(0,1), && \\
&a_{r,ij} | \gamma_{ij} \sim \text{Beta}(1, \gamma_{ij}) && \text{if } k_{ij}^{\text{True}} = 1, \\
&a_{r,ij} = 0 && \text{if } k_{ij}^{\text{True}} = 0.
\end{aligned}$$

Figure 3 shows the plate-notation for the proposed methodology.

Also, to reduce the number of parameters, we assume that some link probabilities are the same based on our belief about whether or not they require knowledge of same concepts. For example, Figure 5 represents a knowledge map in which different subgroups of variables are enclosed in separate boxes. Then, for any pair of boxes the link probability $a_{r,ij}$ connecting any variable i in the first box to any other variable j in second box is constant. The probability of links between Marin Factors and variable *Se* are the same.

3.5 Sampling from the posterior

The final step is specify how to sample from the joint posterior of all parameters. Because the Markov Chain Monte Carlo algorithm is generally slow for running inference on large graphs [24] (such as the fatigue knowledge graph), we employ the No-U-Turn sampler (NUTS) that is an extension of the Hamiltonian Monte Carlo method [25]. The NUTS can be implemented in a Python environment using the PyMC3 library [26]. To speed-up the inference, we reparameterize the binary link variable $k_{r,ij}$ into a continuous variable using a sigmoid function of latent variable $\lambda_{r,ij}$, while ensuring that the link probability $a_{r,ij}$ remains the same. The specifics of the reparameterization

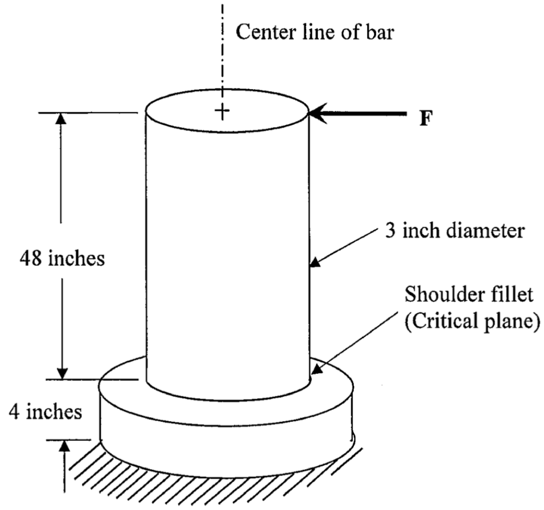


FIGURE 4: A steel shaft under cyclic loading F

are as follows:

$$\lambda_{r,ij}|a_{r,ij} \sim \text{Normal}(\Phi^{-1}(a_{r,ij})), 1),$$
$$k_{r,ij} = \text{sigm}(50\lambda_{r,ij}).$$

Function Φ^{-1} is the inverse cumulative density function of the standard normal distribution. Since the continuous reparameterization of the binary link variable is very close to either 0 or 1, $k_{r,ij}$, is used directly for Eq. (5).

4 Data Collection

We collected student response data during the final exam of a machine design course at a major university. This dataset involved 205 student subjects. The exam tested the concepts of fatigue theory using a shaft design problem. In this problem, the objective is to perform fatigue analysis to ensure the shaft can support a set of loads and is safe against yielding and fatigue failure. The geometry of the shaft, dynamically loaded at one end and fixed at the other is shown in Figure 4. Figure 5 highlights the mapping between the input-output variables for the theory of fatigue failure. Variable F represents the external loading applied to the steel shaft with geometry G which is being operated at room temperature T. The external loading, F, causes the bar to develop bending moment M. Variable R is the reliability requirement for the bar. The ultimate tensile strength S_{ut} is a material property. The theoretical endurance limit Se_p is defined in terms of the ultimate tensile strength S_{ut} using empirical relations given in Ref. [27, sec. 6-7]. The nominal stress σ_o is adjusted by multiplying with the fatigue stress-concentration factor for bending K_f. The adjusted stresses are shows as σ. The endurance limit Se_p is adjusted through multiplication by *Marin Factors* for different conditions of surface finish, size, loading, temperature and miscellaneous factors. This adjusted endurance limit is denoted as Se. Finally, the factor of safety (FOS) is shown as n_f.

TABLE 1: Causal links required to answer the questions

Question	Design Parameters	Output Parameter	Relevant Causal Links
Q_1	F,G,M	σ_o	(G,M), (G,σ_o), (F,M),(M,σ_o)
Q_2	F,G	M	(G,M), (F,M)
Q_3	F,G	M	(G,M), (F,M)
Q_4	G,M,σ_o,K_f	σ	(G,M), (G,σ_o), (M,σ_o), (σ_o,σ), (K_f,σ)
Q_5	G,M,σ_o,K_f	σ	(G,M), (G,σ_o), (M,σ_o), (σ_o,σ), (K_f,σ)
Q_6	F,G,M	σ_o	(G,M), (G,σ_o), (F,M),(M,σ_o)
Q_7	S_{ut}	Se_p	(S_{ut},Se_p)
Q_8	G	k_a	(G,k_a)
Q_9	F,S_{ut}	k_b	(F,k_b),(S_{ut},k_b)
Q_{10}	R	k_e	(R,k_e)
Q_{11}	F,T	k_c,k_d	(F,k_c), (T,k_d)
Q_{12}	Se_p, k_a, k_b, k_c, k_d, k_e	S_e	(Se_p,S_e), (k_a,S_e), (k_b,S_e), (k_c,S_e), (k_d,S_e), (k_e,S_e)
Q_{13}	σ,Se	n_f	(σ,n_f), (Se,n_f)

The student subjects were provided with a total of 13 fatigue theory questions. Each question included input variables and expected the students to calculate an output parameter; assuming that input variables causes the output variables. Answers to a given question depend on parents or a set of ancestral nodes. Consider question 2 and question 12 for which the relevant nodes and links are highlighted by using loosely spaced dashes and densely spaced dots respectively in Figure 5. In question 2, nodes F and G are the input variables and M is the output variable. Similarly, for question 12, nodes Se_p, k_a, k_b, k_c, k_d, and k_e correspond to the input variables and endurance limit Se as the output variable. For the sake of brevity, the actual questions are not included in this paper but can be made available upon request. Further, Table 1 shows input variables, output variables and the relevant causal links all 13 questions.

5 Results

We present the results of model parameter estimation and compare predictive accuracy of both, the three-parameter logistic

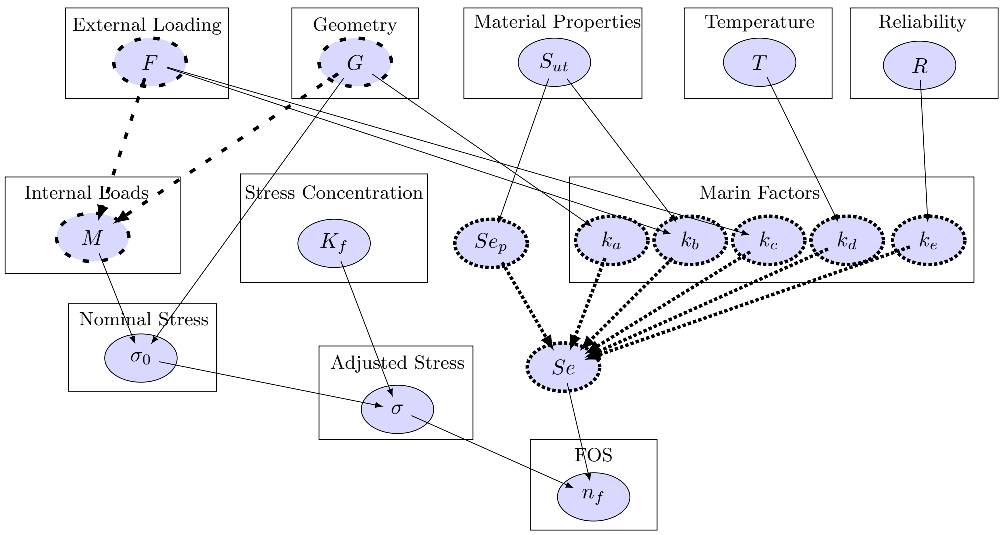


FIGURE 5: The true directed-acyclic graph for the theoretical knowledge of fatigue failure. The loosely dashed and densely dotted nodes represent the relevant variables for questions 2 and 12 respectively.

(3PL) model, and the graph-based based logistic (GrL) model. The results are divided into four parts: i) model checking, ii) representation of person-specific aggregate ability, iii) representation of problem difficulty, and iv) representation of individuals' knowledge graphs.

5.1 Model Checking

We utilize the Watanabe-Akaike information criterion (WAIC) for calculating the in-sample prediction accuracy of the two models. The WAIC estimates the log pointwise predictive density of observed data and adds a correction term based on the effective number of model parameters to adjust for overfitting [28, 29]. Because the goal is to compare two models with different number of parameters, it is essential to account for the natural ability of a model with more number of parameters to fit data better [30]. The lower the WAIC, the better the predictive accuracy. The results in Table 2 indicate that the GrL model has the lower WAIC as compared to 3PL model and, thus, the better predictive accuracy.

TABLE 2: Model comparison based on Watanabe-Akaike information criterion (WAIC)

Model	Ability Parameters	Problem parameters	WAIC
3 Parameter Logistic	1	3	1929.49
Graph Based Logistic	16	26	721.60

For visual checks, posterior predictive checking is performed on the individual responses in Figure 6 and on important test quantities in Figure 7. We see that the posterior predictions using the GrL model match the patterns in the observed responses much more closely that the 3PL model. At the population level, in Figure 7, the prediction of the total number of correct responses is explained adequately by both models with Bayesian p-values close to 0.5 for both models, see Ch. 6 of [31]. However, for questions Q_4 and Q_5 which require knowledge of the same subgraph, the GrL model better explains the the number of students who get both the questions correct than the 3PL model.

Further, we compare the question-specific and person-specific posterior predictive accuracy using test quantities such as the number of correct responses. Suppose $s_l = \sum_{r=1}^{R} E_{rl}$ (summed over all students) is the number of correct responses for a given question l, then the percentage predictive accuracy is the fraction of predicted responses that match with the observed responses, $\frac{\sum_{i=1}^{5000} 1[\hat{s}_{l,i}=s_l]}{5000}$, where $s_{l,i}$ is the prediction of the test quantity by i^{th} posterior sample. As seen from Figure 8, the GrL model has higher question-specific posterior prediction accuracy. Further, the test quantity for comparing the person-specific predictive accuracy is the number of correct responses by individual students. If the number of correct responses by person r is $s_r = \sum_{l=1}^{L} E_{rl}$ (summed over all questions), then the person-specific percentage predictive accuracy is $\frac{\sum_{i=1}^{5000} 1[\hat{s}_{r,i}=s_r]}{5000}$, where $\hat{s}_{r,i}$ is the predicted number of correct responses for person r and $1[A]$ is an indicator function which is 1 if condition A is true and 0 otherwise. In Figure 9, the distribution of pooled percentage accuracy for all

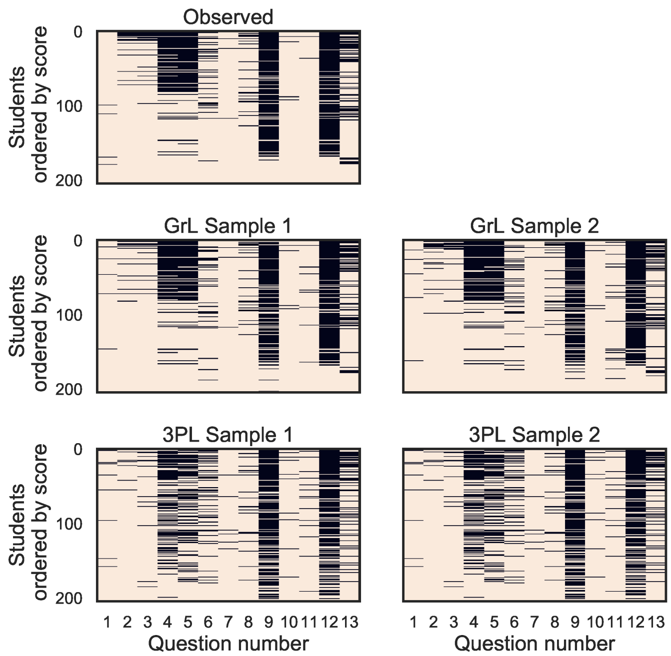


FIGURE 6: Posterior predictive checking on all responses. Black color represents incorrect responses.

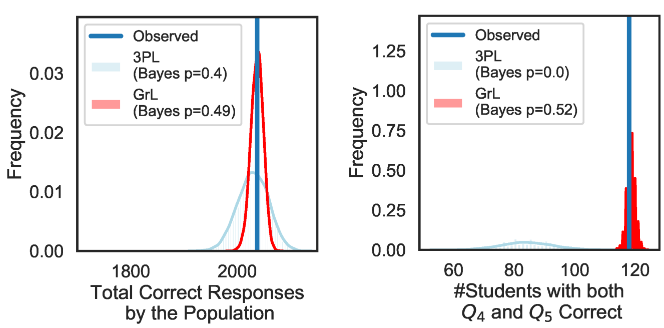


FIGURE 7: Posterior predictive checking on test quantities: 1) the total number of correct responses for the population, and ii) the number of students who get both Q_4 and Q_5 correct.

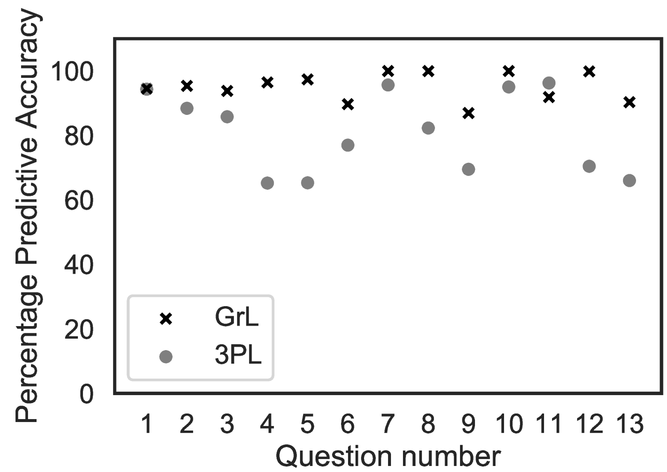


FIGURE 8: Percentage predictive accuracy on the number of correct responses in different questions.

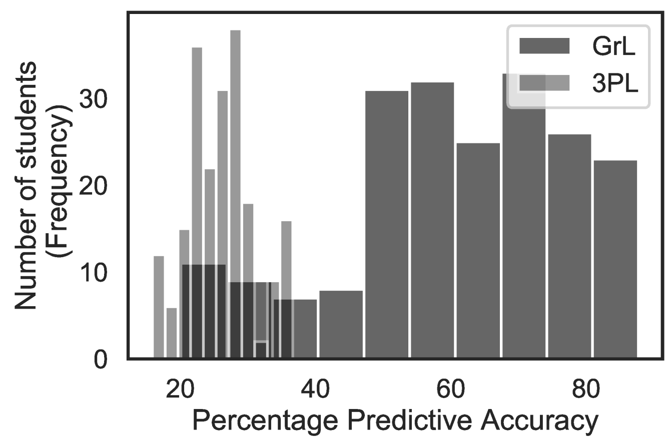


FIGURE 9: Percentage predictive accuracy on the number of correct responses by individual students.

students indicates that the GrL model has higher accuracy compared to the 3PL model for predicting an individual's number of correct responses. Note that the average person-specific accuracy for the GrL model is 0.7, which is smaller than 1, indicating the GrL model may need further improvements.

5.2 Representing Aggregate-level Ability

As expected, both the GrL and 3PL models show a positive correlation between the estimated aggregate ability and the students' exam score . Recall that threshold θ represents an individual's ability in the 3PL model. For the GrL model, the total number of matched links with respect to the true knowledge matrix quantify an individual's overall ability which is given as

$$\#\text{Matched links} = \| (Q_l \circ K_r) \circ (Q_l \circ K^{\text{True}}) \|_F \,. \tag{7}$$

These estimated ability parameters are in general agreement with the exam score, which is commonly regarded as the baseline for students' overall knowledge. Figure 10 shows that the estimated θ increases as the exam score increases. Similar trend is observed for the estimated number of matched links in Figure 11. The GrL model seems to have more variation that 3PL model. One possible reason could be that 3PL model represents one-dimensional θ variable whereas GrL model represents the num-

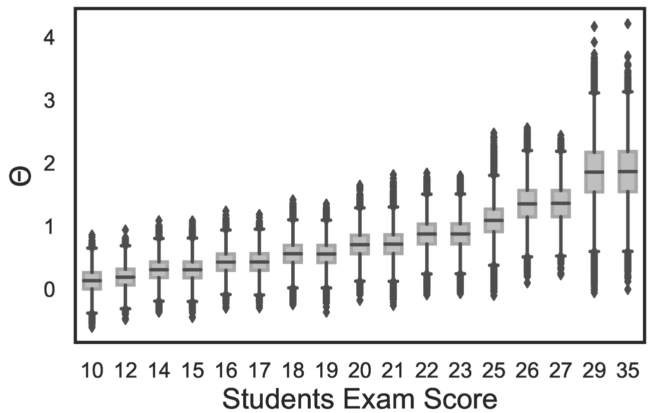


FIGURE 10: Estimated person-specific ability θ in the three parameter logistic (3PL) model as a function of exam scores

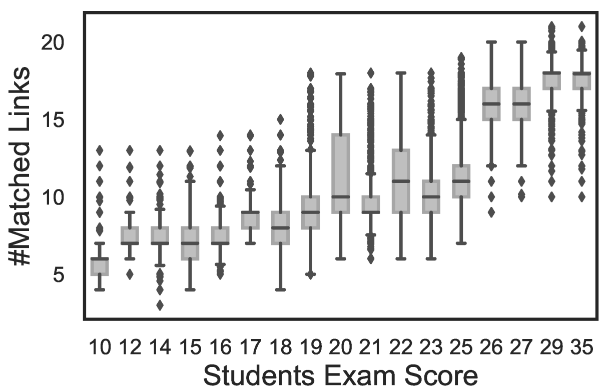


FIGURE 11: Estimated total number of matched links from the graph-based logistic (GrL) model as a function of exam scores

ber of matched links which are calculated from multiple causal link probabilities.

5.3 Representing Problem Difficulty

Further, the posterior estimates of the question-specific difficulty β_l for the 3PL are shown in Figure 12. Across the population, the 3PL model identifies questions 1, 7, 10 and 11 as easy problems and questions 9 and 12 as difficult. The estimation of difficulty is mostly along the lines of percentage wrong responses, for instance high fractions of the students, 44% and 45% respectively, get questions 9 and 12 wrong. Interestingly, the questions requiring knowledge of the same links has similar β_l estimates, e.g. questions 2 and 3, and questions 4 and 5.

On the other hand, the GrL does not quantify question difficulty explicitly using one parameter. Parameter β in the GrL represents the threshold fraction of required links necessary for answering any question correctly. A higher value of β in the GrL could be interpreted as average difficulty perceived by a population taking that exam. However, it is important to emphasize that problem difficulty may be different for different students based on their ability. We should also consider the individuals' concept-specific knowledge for indicating problem difficulty. The results from quantification of concept-level knowledge as knowledge graphs is presented in Section 5.4.

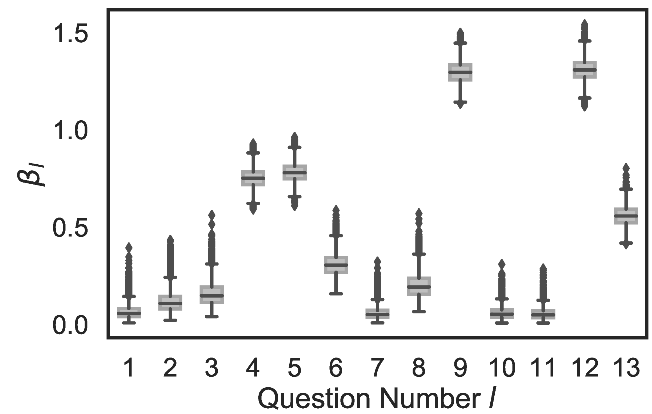


FIGURE 12: Posterior estimates of problem difficulty parameter β_l for three parameter logistic (3PL) model.

5.4 Representing Causal Knowledge

Unlike the 3PL model, the GrL model can quantify an individual's causal knowledge in terms of link probabilities. Figure 13 represents the estimated posterior link probabilities for all causal links across the student population. It shows that the student population has better knowledge of some links than the others. For example, the students know links $(G,M),(G,\sigma_o),(G,K_a)$ with high certainty. Conversely, for some causal links, such as (K_f,σ) and (S_{ut},k_b), the population density is skewed towards 0; signifying that the population has poor knowledge of these links.

The GrL model can also help categorize the subjects based on their knowledge of causal links. Consider two students, a high-scoring student who answered all 13 questions correct versus a low-scoring student who answered 5 questions correct. Figure 14 compares the knowledge structures of these two students. The differences in the knowledge structures are evident from the estimated link probabilities. The high-scoring student has high knowledge all concepts such as external loading, internal loading, internal stresses and fatigue factor of safety. The low scor-

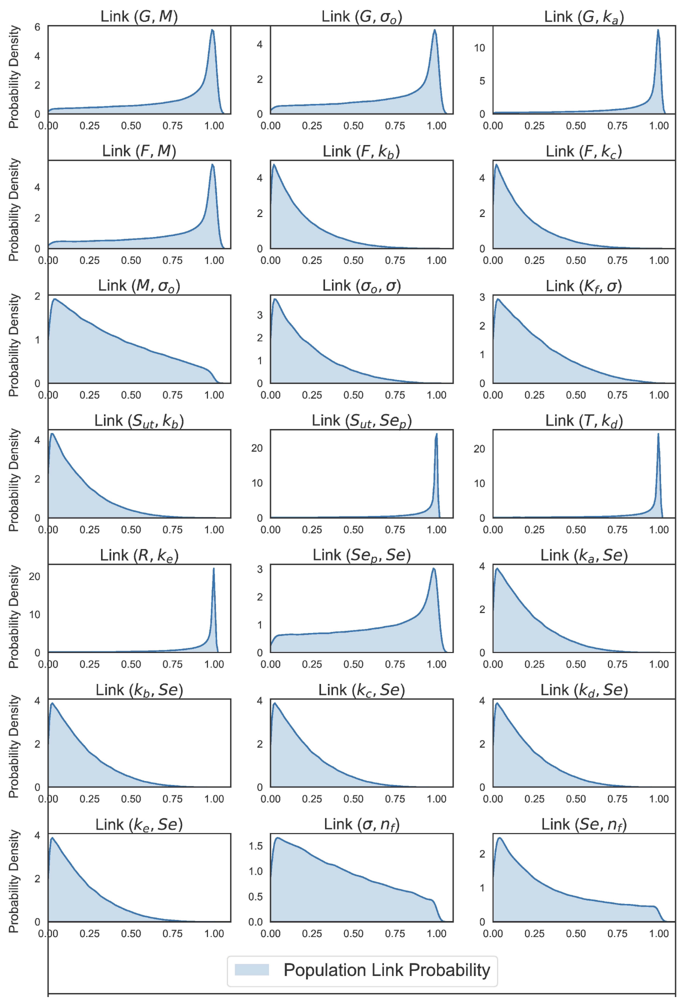
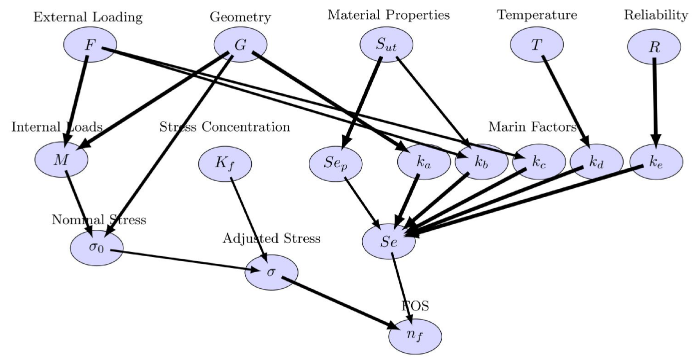


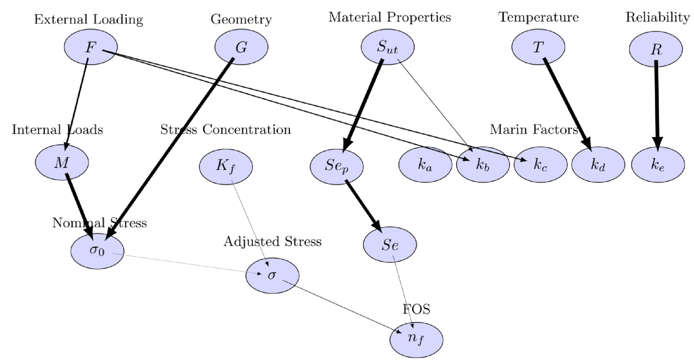
FIGURE 13: Population-wide posterior distribution of link probability parameters a_{ij} in the graph-based logistic model.

ing student knows some links, e.g. calculations of Marin factors k_d, k_e and nominal endurance limit Se_p, but does not know other links. For the low scoring student, the GrL model estimates the following link probabilities as zero: Marin factors to the adjusted endurance limit Se, geometry G to k_a, and geometry G to bending moment M.

Getting link-level probabilities is particularly useful for predicting concepts and questions that an individual might get wrong. For example, consider question 7, for which (S_{ut}, Se_p) is the relevant causal link. The probability densities for this link is close to 1, signifying that the students are likely to get this link correct. This is also reflected by the student score; with 98% of students getting question 7 correct. Further, consider question 9, for which (F, k_b) and (S_{ut}, k_b) are the relevant causal links whose estimated probability densities are skewed towards 0. Then, the lack of knowledge about the relevant links is reflected in the student performance for question 9; with only 30% of students getting question 9 correct. As we traverse down the the knowledge graph, we observe that individuals are more likely to get incorrect answers because latter concepts require more detailed understanding (e.g. question 13) and based on more parent concepts (e.g. question 12). It is important to note here that for some causal links, such as (K_a, Se),(K_b, Se), etc., have identical population density because of the assumption that some causal links (in same sub-groups) share the same probability.



(a) Estimated knowledge graph for a student with all correct responses



(b) Estimated knowledge graph for a student with 5 correct responses

FIGURE 14: Comparison of the directed-acyclic graphs representing knowledge structures for high-scoring and low-scoring students. Larger line thickness represents greater probability of knowing the link correctly.

6 Implications for Engineering Education, Design Research and Practice

From the engineering education perspective, the graph-based logistic model can provide an in-depth understanding of the concepts that a given population has poor knowledge of as well as the concepts that the population knows well. From observing the estimated probability density of individual causal links, instructors can provide personalised feedback on specific concepts that help improve students' knowledge. Also, instructors could potentially use a different combinations of causal links to generate questions with varying difficulty. This would not only allow instructors to test the same concepts using different questions but also help reveal the true knowledge of the concepts for students.

The computational modelling technique can be coupled with the existing design research for capturing design knowledge and predicting human behaviors. The expertise research may benefit from quantifying granular design knowledge for comparison of design theories such as novice designers implement situation-independent rules while experienced designer tend to think in a pattern-based way [32]. The model can be applied for developing more realistic representation of human behaviors such as similarity assessment [4], design decisions [2], and design performance [33].

For practitioners, a better understanding of the knowledge structures can help in reducing the inefficiencies caused by poor comprehension of relevant physical variables for a given design problem. This would enable better design support tools that support human designers in decision making and information provision, better computational design methods that mimic human reasoning, better design of human-machine interaction (e.g., co-robotics), and improved design of automated artificial intelligence (AI) based products and systems [1], and partially automated augmented intelligence systems that work with humans.

7 Conclusions

The proposed graph based logistic (GrL) model presents a new approach for quantifying an individual's causal knowledge. The model is theory driven and can be implemented for any theory with causal relations. The model quantifies the experimenter's belief about an individual's ability in terms of link-specific knowledge. The individual-specific ability also helps in inferring individual-specific problem difficulty. Additionally, this model employs Bayesian framework for accurate estimation of individuals' granular knowledge. The model can be used to capture a given population's overall knowledge structure. Using the indications of which concepts a given population understands well and poorly, instructors can design test questions for accurate knowledge assessment, researchers can predict design behaviors, and practitioners can build computational tools to support design.

In future, we can compare the graph-based logistic model to different multidimensional models of the item response theory. To test the predictive accuracy of the model on external observations, we plan to predict similarity judgments from the estimated knowledge graphs and compare those to available similarity decisions of the same subject population, which were reported earlier in Ref. [4]. The methodology can also be implemented on scientific theories other than the theory of fatigue failure for assessing where engineering subjects have weakest and strongest knowledge.

ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from the National Science Foundation through NSF CMMI Grants No. 1662230 and No. 1728165. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J., 2017. "Building machines that learn and think like people". *Behavioral and brain sciences,* ***40***.

[2] Chaudhari, A. M., Bilionis, I., and Panchal, J. H., 2020. "Descriptive models of sequential decisions in engineering design: An experimental study". *Journal of Mechanical Design,* ***142***(8).

[3] Safarkhani, S., Bilionis, I., and Panchal, J. H., 2020. "Toward a theory of systems engineering processes: A principal–agent model of a one-shot, shallow process". *IEEE Systems Journal*.

[4] Chaudhari, A. M., Bilionis, I., and Panchal, J. H., 2019. "Similarity in engineering design: A knowledge-based approach". In ASME 2019 International Design Engineering Technical Conferences-Design Theory and Methodology.

[5] Safarkhani, S., Bilionis, I., and Panchal, J. H., 2018. "Understanding the effect of task complexity and problem-solving skills on the design performance of agents in systems engineering". In ASME 2018 44th Design Automation Conference, American Society of Mechanical Engineers Digital Collection.

[6] Millán, E., DescalçO, L., Castillo, G., Oliveira, P., and Diogo, S., 2013. "Using bayesian networks to improve knowledge assessment". *Computers & Education,* ***60***(1), pp. 436–447.

[7] Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez-De-La-Cruz, J. L., and Ríos, A., 2004. "Siette: A web-based tool for adaptive testing". *International Journal of Artificial Intelligence in Education,* ***14***(1), pp. 29–61.

[8] Gero, J. S., 1990. "Design prototypes: a knowledge representation schema for design". *AI magazine,* ***11***(4), pp. 26–26.

[9] Fenves, S. J., Foufou, S., Bock, C., and Sriram, R. D., 2008. "Cpm2: a core model for product data". *Journal of computing and information science in engineering,* ***8***(1).

[10] De Ayala, R. J., 2013. *The theory and practice of item response theory*. Guilford Publications.

[11] Christophe, F., Bernard, A., and Coatanéa, É., 2010. "Rfbs: A model for knowledge representation of conceptual design". *CIRP annals,* ***59***(1), pp. 155–158.

[12] Chandrasegaran, S. K., Ramani, K., Sriram, R. D., Horváth, I., Bernard, A., Harik, R. F., and Gao, W., 2013. "The evo-

lution, challenges, and future of knowledge representation in product design systems". *Computer-aided design,* ***45***(2), pp. 204–228.
[13] Wu, D., and Gary Wang, G., 2020. "Knowledge-assisted optimization for large-scale design problems: A review and proposition". *Journal of Mechanical Design,* ***142***(1).
[14] Dong, A., and Sarkar, S., 2014. "Generalized design knowledge and the higher-order singular value decomposition". In *Design Computing and Cognition'12*. Springer, pp. 415–432.
[15] Rebhuhn, C., Gilchrist, B., Oman, S., Tumer, I., Stone, R., and Tumer, K., 2014. "A multiagent approach to evaluating innovative component selection". *Design, computing, and cognition*.
[16] Fu, K., Cagan, J., Kotovsky, K., and Wood, K., 2013. "Discovering structure in design databases through functional and surface based mapping". *Journal of mechanical Design,* ***135***(3).
[17] Shadbolt, N. R., Smart, P. R., Wilson, J., and Sharples, S., 2015. "Knowledge elicitation". *Evaluation of human work*, pp. 163–200.
[18] Schiuma, G., Gavrilova, T., and Andreeva, T., 2012. "Knowledge elicitation techniques in a knowledge management context". *Journal of Knowledge Management*.
[19] Cooke, N. J., 1994. "Varieties of knowledge elicitation techniques". *International Journal of Human-Computer Studies,* ***41***(6), pp. 801–849.
[20] Self, J. A., 1994. "Formal approaches to student modelling". In *Student modelling: The key to individualized knowledge-based instruction*. Springer, pp. 295–352.
[21] Hestenes, D., Wells, M., and Swackhamer, G., 1992. "Force concept inventory". *The physics teacher,* ***30***(3), pp. 141–158.
[22] Reckase, M. D., 2009. "Multidimensional item response theory models". In *Multidimensional item response theory*. Springer, pp. 79–112.
[23] Millán, E., Loboda, T., and Pérez-De-La-Cruz, J. L., 2010. "Bayesian networks for student model engineering". *Computers & Education,* ***55***(4), pp. 1663–1683.
[24] Goudie, R. J., and Mukherjee, S., 2016. "A gibbs sampler for learning dags". *The Journal of Machine Learning Research,* ***17***(1), pp. 1032–1070.
[25] Hoffman, M. D., and Gelman, A., 2014. "The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo.". *Journal of Machine Learning Research,* ***15***(1), pp. 1593–1623.
[26] Salvatier, J., Wiecki, T. V., and Fonnesbeck, C., 2016. "Probabilistic programming in python using pymc3". *PeerJ Computer Science,* ***2***, p. e55.
[27] Budynas, R. G., Nisbett, J. K., et al., 2008. *Shigley's mechanical engineering design*, Vol. 8. McGraw-Hill New York.
[28] Bernardo, J., and Smith, A., 2007. *Bayesian Theory*. Wiley Series in Probability and Statistics. Wiley.
[29] Akaike, H., 1998. "Information theory and an extension of the maximum likelihood principle". In *Selected papers of hirotugu akaike*. Springer, pp. 199–213.
[30] Gelman, A., Hwang, J., and Vehtari, A., 2014. "Understanding predictive information criteria for bayesian models". *Statistics and computing,* ***24***(6), pp. 997–1016.
[31] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B., 2013. *Bayesian data analysis*. Chapman and Hall/CRC. Ch.6.
[32] Cross, N., 2004. "Expertise in design: an overview". *Design studies,* ***25***(5), pp. 427–441.
[33] Shergadwala, M., Bilionis, I., Kannan, K. N., and Panchal, J. H., 2018. "Quantifying the impact of domain knowledge and problem framing on sequential decisions in engineering design". *Journal of Mechanical Design,* ***140***(10).