Abstract: In modern psychological and biomedical research with diagnostic purposes, scientists often formulate the key task as inferring the fine-grained latent information under structural constraints. These structural constraints usually come from the domain experts' prior knowledge or insight. The emerging family of Structured Latent Attribute Models (SLAMs) accommodate these modeling needs and have received substantial attention in psychology, education, and epidemiology. SLAMs bring exciting opportunities and unique challenges. In particular, with high-dimensional discrete latent attributes and structural constraints encoded by a design matrix, one needs to balance the gain in the model's explanatory power and interpretability against the difficulty of understanding and handling the complex model structure.

In the first part of this talk, I present identifiability results that advance the theoretical knowledge of how the design matrix influences the estimability of SLAMs. The new identifiability conditions guide real-world practices of designing diagnostic tests and also lay the foundation for drawing valid statistical conclusions. In the second part, I introduce a statistically consistent penalized likelihood approach to selecting significant latent patterns in the population. I also propose a scalable computational method. These developments explore an exponentially large model space involving many discrete latent variables, and they address the estimation and computation challenges of high-dimensional SLAMs arising from large-scale scientific measurements. The application of the proposed methodology to the data from an international educational assessment reveals meaningful knowledge structure of the student population.