**Personal, Background, and Future Goals Statement**

**Background.** Reading Comprehension (RC) is a fundamental skill essential for academic success and social mobility, yet significant disparities persist among traditionally underserved communities.1 These disparities perpetuate deeply rooted inequalities that traditional educational interventions have struggled to address.2 RC models, such as the Componential and Active View models address factors that include individual and social differences, inference processing strategies, background knowledge, and motivation, among other factors.3 However, there is difficulty applying these models in school settings and identifying the aforementioned factors that influence RC ability.

This difficulty is for good reason. RC is one of the most basic yet challenging cognitive processes students learn.4 How individuals are able to transform linguistic symbols into meaningful concepts, retain that information, and then draw on it as needed has been the focus of intense theoretical and empirical study for decades.5 While robust, existing models of RC are relatively untested within dynamic environments where factor influence fluctuates as tasks and contexts vary. Recent advancements in artificial intelligence (AI) and machine learning (ML) offer new possibilities to address these limitations.

Included in these advancements are neural networks (NN), which are statistical modeling systems that utilize data processing strategies to create useful representations of objects and features. After learning from a particular task, a NN is able to leverage its learned representations to make predictions within dynamic systems.6 This includes RC models. However, in order for NNs to accurately predict RC, the underlying factors that contribute to RC must be accessible to the NN. This can be achieved by retraining NNs (fine-tuning) and introducing RC factors in their architecture. This would allow NNs to improve their ability to predict and represent the underlying factors of RC, which can be achieved with Hierarchical Graph Neural Networks (HGNN).7

HGNNs are a novel approach within the field of AI and provide a powerful framework for modeling the complex cognitive processes underlying RC. By utilizing nodes to represent cognitive and contextual factors from various RC models and edges to encode the relationships between them, HGNNs may offer a more dynamic and comprehensive analysis of RC.8 This research aims to develop and validate HGNNs to capture both individual and social differences across RC models, with the ultimate aim of identifying relevant factors that can be used to improve student comprehension outcomes.

**Proposed Research.** Despite extensive research into RC, there has been little focus on creating dynamic frameworks to examine the variation and accuracy of these models in real time, dynamic environments. This is compounded by the difficulty of monitoring and tracking the litany of factors involved.9 Thus, current RC models fail to capture the dynamic cognitive processes that underlie text comprehension and their change over time. These models primarily focus on surface-level language processing and are not equipped to account for the complexity of individual cognitive variations shaped by factors such as socioeconomic background, linguistic diversity, or educational context.10 As a result, these models inadequately reflect the RC outcomes especially for historically underrepresented groups.

This research aims to address this gap by developing and validating HGNNs that can more accurately model the factors underlying digital text-based RC. This approach will provide a deeper understanding of how individual differences in cognitive processing affect reading outcomes, particularly in populations that have been systematically underserved by traditional educational assessments.

R**esearch Questions**. *(1) To what extent do current empirical and theoretical models of RC overlap and can be represented via HGNNs.* *(2) How do key psychological constructs, such as emotional valence, motivation, and reasoning, influence RC across populations?* *(3) To what extent do HGNNs accurately predict individual RC outcomes compared to traditional models, particularly for diverse populations? (4) How well do the HGNN-based cognitive models align with self-reported cognitive processes in participants from diverse backgrounds? (5) Can HGNN models inform the development of more equitable and personalized educational assessments and interventions for improving digital RC outcomes in diverse populations?*

**Methods and Resources.** This research will be structured into four studies that integrate factors from existing RC models, contributing to the development and validation of HGNNs for modeling RC.

**Study 1** is a scoping review of the current literature. It addresses ***Research Questions 1 and 2*** by analyzing existing RC models (e.g., Componential, Active View, KReC models) to identify key factors and models that will serve as the basis for implementing HGNN models in subsequent studies.

**Study 2** addresses ***Research Questions 1, 2, and 3*** by modeling existing RC frameworks and into HGNNs. These HGNNs will be trained on existing RC data and the predictive accuracy of HGNN models will be compared to traditional RC models. HGNNs will be refined through optimization techniques, including loss and error rates, and tested on unseen data to ensure generalizability. We hypothesize that HGNNs will outperform traditional RC assessment in addition to increased factor interpretability.

**Study 3** addresses ***Research Question 4*** by developing self-report measures to assess how participants’ cognitive processes (e.g., emotional engagement, reasoning) align with HGNN predictions. Participants will complete questionnaires pre and post reading tasks. These self-reports will be compared to HGNN and traditional model predictions. We hypothesize that self-report strategies will show moderate to strong alignment with HGNN predictions and outperform traditional RC model predictions.

**Study 4** addresses ***Research Questions 3, 4, and 5*** by validating the HGNN models developed in Study 2 using the self-report data gathered in Study 3. Through reinforcement learning and human feedback, the models will be adapted to better capture cognitive processes and predict comprehension outcomes across diverse populations. We hypothesize that the refined HGNNs will provide more accurate predictions across diverse populations compared to traditional assessments.

**Intellectual Merit.** This research will significantly advance our understanding of the cognitive processes that underlie RC. Current models in cognitive science have largely overlooked the complexity and variability of cognitive processes across demographic groups. The dynamic modeling of hierarchical relations among factors and constructs allows for a deeper understanding of their interactions and influences on comprehension outcomes. Resultantly, this research will contribute to more accurate and nuanced modeling of reading comprehension. To the researcher’s knowledge, the proposed studies will be one of the first to incorporate such constructs into HGNNs designed to predict text-based comprehension.

The innovative use of HGNNs in this study provides a novel way to model the complexity of cognitive processes during RC. Unlike traditional NN models, HGNNs can better capture the dynamic and hierarchical nature of cognition, allowing for a more sophisticated analysis of how diverse learners process and interpret digital texts. This approach will offer greater interpretability, as well as account for individual differences in cognitive processing. Such research addresses a critical gap in current assessment practices, which often fail to account for the cognitive diversity of learners.

**Broader Impacts.**  There are biases inherent to AI systems that must be acknowledged. However, the outcomes of these proposed studies should allow researchers and educators to identify and address components in student RC that are difficult to monitor for both individuals and at scale. This research should lead to more effective, personalized feedback systems and adaptive learning technologies that support diverse cognitive processes. By improving the accuracy and equity of RC assessments, this research will contribute to broader societal efforts to reduce educational disparities. The development of more sophisticated RC models will help ensure that students from all backgrounds, including traditionally underrepresented communities, receive the support they need to succeed across their lifetime. Additionally, this research will foster greater inclusivity in the design of AI and educational technologies, promoting educational access and opportunity for all learners, regardless of their background.

**References:** 1Romeo et al., 2022; 2Duke and Carlisle 2010, 3Aaron and Joshi, 2000; Duke and Cartwright, 2021; 4Kendeou et al., 2014; 5van den Broek et al., 2005; 6Perconti and Plebe 2020; 7Zhong et al. 2022; 8Sobolevsky, 2021; 9McMaster and Kendeou 2023; 10James et al., 2023