

**Nebraska Technical Advisory Committee Meeting**  
**Nebraska Department of Education**  
**May 27, 2021**

**60 minutes: Through Year Model Simulation Results**

During the NDE, NWEA and TAC through-year test model working sessions, the NWEA psychometrics team received feedback on configurations that stakeholders would like applied to test models for simulations so that the test design will match the desired outcomes. Based on the feedback, the current model requires the engine to limit the selection of diagnostic off-grade items to one grade above and below the student's enrolled grade of record. Additionally, the engine focuses on information in the diagnostic section from domain areas where the simulation shows specific areas of weakness. High-performing students are simulated going off grade when the result from the operational section illustrates high performance. NWEA will present the results from the updated simulations for ELA and Mathematics Grades 4, 5, and 6. These results will include the test reliability, blueprint matching performance, diagnostic section results, and examples of scoring options based on the simulations.

1. Based on these results, what additional updates would you recommend in preparation for the Winter Pilot?

**30 minutes: EIDM Standard Setting Method**

The Embedded Item Descriptor Matching (EIDM) method is intended to integrate alignment and standard setting into a single process, taking what is often viewed as disparate processes of assessment development and synergizing them. Under EIDM, teachers align items to range achievement level descriptors (RALDs), which explicate the progression of cognition students need to show for more advanced knowledge and skills in the content area as they develop toward college and career readiness or other appropriate goals. Rater agreement statistics quantify the degree to which teachers agree on and converge in agreement on alignment across each round. The Embedded Standard Setting (ESS) method (Lewis & Cook, 2020) is used to calculate the cut scores based on the alignment information by optimizing the placement of the cut score at the point on the test scale where score interpretations are most accurate. Therefore, this approach offers states critical advantages over treating alignment and standard setting as separate approaches; it adds value to test scores while streamlining processes.

1. What is the TAC's perspective of piloting EIDM with a small number of teachers or NDE SMEs in science with the field test items and adjusting the RALDs based on data?

**30 minutes: Item Difficulty Modeling for ELA Reading Items**

Item difficulty modeling (IDM) was conducted to predict item difficulty for NSCAS ELA reading items using item features that reflect cognitive, content, and stimulus demands of the item, including item type, item RALD level, DOK levels, and passage text complexity. Four models were applied, including the widely used multiple linear regression and three machine learning algorithms: support vector machines (SVM), random forests (RF), and k-nearest neighbor regression (k-NN). Results from the models suggest that the included item features can predict item difficulty to some extent. The linear regression model and the random forests model achieved an R-squared around 0.38 and 0.41 for the reading items, respectively. A moderate relation between item RALD level and item difficulty was discovered, which provided validity evidence of the RALDs and the items.

1. What are ways to communicate results to teachers about the value of RALDs given that the RALDs appear to contribute unique information to the model and are barely to not related to text complexity?
2. How should the range of text complexity be considered in a through-year system? In supplemental analyses, we have found that grade becomes a significant predictor if we remove text complexity. However, with Lexile in the model, Lexile becomes a surrogate for grade level even with the wide overlap of Lexile levels across grades.