HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION

2025 No. 084

Review of Revised Iowa Test Specifications

Technical Memorandum

Prepared for: Catherine J. Welch
The University of Iowa
240 South Madison Street
Iowa City, IA 52242

Prepared under: Contract # 00000994

Authors: Emily R. Dickinson
Caroline R. H. Wiley
Jon S. Twing
Arthur A. Thacker

Date: July 31, 2025

Review of Revised Iowa Test Specifications

Executive Summary

In 2024, the Iowa State Board of Education revised Iowa's Academic Standards. Revised test specifications for the Iowa Statewide Assessment of Student Progress (ISASP) were created to guide the development of new test forms measuring student performance on the revised standards. Iowa Testing Programs (ITP) at the University of Iowa requested that the Human Resources Research Organization (HumRRO) conduct an independent, external review of the revised test specifications.

HumRRO used a combination of Large Language Model (LLM) tools and human review to evaluate the extent of the changes to the revised standards. First, we calculated the cosine similarity index of each original and revised standard pairing. Second, we prompted HumRRO's proprietary and secure Artificial Intelligence (AI) engine, RRObot, to compare each original and revised standard and provide a rating of the likelihood of item transferability for each comparison. HumRRO staff then reviewed the AI-based results and provided a final rating. These ratings, along with targets from the ISASP test specifications, were then used to identify whether the test specifications support the production of valid and reliable scores.

Results from this process addressed the following three research questions:

Research Question 1. Do the revised content standards reflect the same overall content domain as the original?

We found the revised content standards to reflect the same overall content domain as the original standards. The ELA standards are essentially identical in terms of their composition; both are organized around the College and Career (CCR) readiness anchor standards. The main source of difference in the two sets of standards was changes in the DOK level associated with some standards. All substantive changes to the content of the math standards are at the high school level. However, most standards were not changed enough to render available items unacceptable for measuring those standards.

Research Question 2. Does the organization of the test specifications reflect the organization of the standards?

We found that the test specifications clearly reflect the organization of the Iowa Academic Standards for ELA and mathematics while accounting for grade specific assessments. The coding schema of the standards is used in the test specifications, allowing for a clear delineation showing that the test specifications cover the content outlined in the standards. In high school math, the prior high school math standards contained a single Algebra topic, whereas the revised standards divide the content into Algebra 1 and Algebra 2 and the standards are organized by topic (e.g., Algebra 1, Geometry, Algebra 2). Test specifications are appropriately organized by grade level assessment such that a mapping of the non-grade standards to grade specific specifications was required. We were able to map the topic-based standards to the grade level-based test specifications to analyze the test specifications relative to reporting with relative ease.

Research Question 3. Does the organization of the test specifications support the production of valid and reliable scores at the level reported?

The ISASP item pool consists of items developed to measure the original standards. The alignment of items to the original standards was also verified via a traditional alignment study (Dickinson, et. al., 2019). This established the alignment between items and standards and served as the basis for evaluating how revisions to the standards might impact the suitability (i.e., transferability) of items to measure the new standards and, by extension, the ability to meet specifications and produce defensible scores. We found that for nearly all reporting categories, the transferability of items should allow for the targets outlined in the test specifications to be met.

The notable exception was the high school Statistics and Probability reporting category, where all the standards within a single cluster were changed substantially. Although the minimum number of items needed for this reporting category could be sourced from other clusters, the validity of the category scores may be limited by the omission of a cluster (i.e., breadth of coverage).

Review of Revised Iowa Test Specifications

Table of Contents

List of Tables

Review of Revised Iowa Test Specifications

Background

In 2024, the Iowa State Board of Education revised Iowa's Academic Standards. Revised test specifications for the Iowa Statewide Assessment of Student Progress (ISASP) were created to guide the development of new test forms measuring student performance on the revised standards. Iowa Testing Programs (ITP) at the University of Iowa requested that the Human Resources Research Organization (HumRRO) conduct an independent, external review of the revised test specifications.

Changes to the academic content standards pose potential risks to the utility of the pool of existing test items and associated testing apparatus. If changes are minor, then items developed to measure the original standards can be considered "transferable" to the new standards. More substantial revisions to the standards could mean that not all existing items will fully measure the revised standards. This threatens a testing program's capability to construct the best tests with aligned test forms without some amount of item revision, new item development, or other changes to operational delivery and reporting requirements.

HumRRO's review focused on the extent of changes to the Iowa Academic Standards for English Language Arts (ELA) and mathematics, how these changes impact "item transferability," and the implications of these for constructing test forms that meet the targets outlined in the test specifications. In addition, the review details the impact this might have on the collection of validity evidence and the test reliability needed to produce scores at the levels reported.

Research Questions

HumRRO's review was designed to address the following research questions:

1. Do the revised content standards reflect the same overall content domain as the original?
2. Does the organization of the test specifications reflect the organization of the standards?
3. Does the organization of the test specifications support the production of valid and reliable scores at the level reported?

Methods

To evaluate the extent of changes to the revised Iowa Academic Standards for ELA and math (Research Question 1), HumRRO used a combination of Large Language Model (LLM) tools and human review. Specifically, each original standard and its corresponding revised standard were compared via two LLM-based approaches, with review by HumRRO staff.

First, we calculated the cosine similarity index of each original and revised standard pairing. The cosine similarity index indicates how similar two different pairs of text are in terms of their semantic similarity and can be interpreted like a Pearson correlation coefficient (Butterfuss & Doran, 2024).

Second, we prompted HumRRO's proprietary and secure Artificial Intelligence (AI) engine, RRObot, to compare each original and revised standard and provide a rating of the likelihood of item transferability for each comparison, along with a detailed rationale for each rating. Standards comparisons were rated using the following scale selected by HumRRO:

- 3 =The revised standard is either identical to the original standard or contains changes to grammar and/or organization, but the key academic content elements are the same. An item written to measure the original standard is likely appropriate for measuring the revised standard.
- 2 =The revised standard contains some of the same key academic content elements as the original standard, but some are slightly different. An item written to measure the original standard might only partially measure the revised standard, or the standard might include additional content not directly addressed by existing items.
- 1 = The revised standard contains substantial changes to the key content elements contained in the original standard. An item written to measure the original standard is likely not appropriate for measuring the revised standard.

HumRRO staff then reviewed the AI-based results and provided a final rating. Specifically, they focused their review on any standards flagged based on a cosine similarity value less than 0.90 and/or an AI-based rating of 2 or 1. In addition, ELA standards were flagged for any differences in the Depth of Knowledge (DOK) levels associated with a standard. If HumRRO reviewers disagreed with the AI-generated rating, they recorded their rationale for doing so.

Research Question 2 was informed by visual inspection of the standards and test specifications. HumRRO considered the utility of the test specifications for determining which standards, clusters, domains, and/or reporting categories would be assessed and the number of items specified to assess them.

Ratings of item transferability were then used, along with targets from the ISASP test specifications, to identify whether the test specifications support the production of valid and reliable scores (Research Question 3). Specifically, after accounting for any changes to the content domain reflected in the revised standards, we evaluated whether the test specifications include item targets that would support valid and reliable scores.

Results

Tables 1 through 17 summarize the analysis of the test specifications. Each table presents the reporting categories and the percentage of the domains and/or clusters within each reporting category that should have available items based on the extent of changes to the associated standards. The minimum and maximum numbers of items for each reporting category indicated in the test specifications are also presented, along with an evaluation of whether available items should allow for all test specification targets to be met, as well as an evaluation of whether the number of items for each reporting category supports valid and reliable scores at that reporting level. Meeting test specifications would be flagged if, for example, the number of standards rated as having low to no item transferability would likely impact whether the minimum number of items targeting a cluster and/or domain would be available within the current item pool.

Table 1. Evaluation of Proposed Grade 3 ELA Specifications

Reporting Category	% Standards Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Key Ideas and Details	100% (6)	14	22	Yes	Yes
Craft and Structure	100% (6)	11	16	Yes	Yes
Integration of Knowledge and Ideas	100% (5)	3	9	Yes	Yes

Table 2. Evaluation of Proposed Grade 4 ELA Specifications

Reporting Category	% Standards Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Key Ideas and Details	100% (6)	15	22	Yes	Yes
Craft and Structure	100% (6)	11	17	Yes	Yes
Integration of Knowledge and Ideas	100% (5)	3	9	Yes	Yes

Table 3. Evaluation of Proposed Grade 5 ELA Specifications

Reporting Category	% Standards Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Key Ideas and Details	100% (6)	15	23	Yes	Yes
Craft and Structure	100% (6)	11	17	Yes	Yes
Integration of Knowledge and Ideas	80% (4)	3	10	Yes	Yes

Table 4. Evaluation of Proposed Grade 6 ELA Specifications

Reporting Category	% Standards Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Key Ideas and Details	100% (6)	16	24	Yes	Yes
Craft and Structure	100% (6)	12	18	Yes	Yes
Integration of Knowledge and Ideas	100% (5)	3	8	Yes	Yes

Table 5. Evaluation of Proposed Grade 7 ELA Specifications

Reporting Category	% Standards Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Key Ideas and Details	100% (6)	16	25	Yes	Yes
Craft and Structure	100% (6)	12	19	Yes	Yes
Integration of Knowledge and Ideas	100% (5)	3	8	Yes	Yes

Table 6. Evaluation of Proposed Grade 8 ELA Specifications

Reporting Category	% Standards Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Key Ideas and Details	100% (6)	16	25	Yes	Yes
Craft and Structure	100% (6)	12	19	Yes	Yes
Integration of Knowledge and Ideas	100% (5)	3	8	Yes	Yes

Table 7. Evaluation of Proposed Grade 9-10 ELA Specifications

Reporting Category	% Standards Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Key Ideas and Details	100% (6)	16	25	Yes	Yes
Craft and Structure	100% (6)	12	19	Yes	Yes
Integration of Knowledge and Ideas	100% (5)	3	8	Yes	Yes

Table 8. Evaluation of Proposed Grade 11 ELA Specifications

Reporting Category	% Standards Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Key Ideas and Details	100% (6)	16	25	Yes	Yes
Craft and Structure	100% (6)	12	19	Yes	Yes
Integration of Knowledge and Ideas	100% (5)	3	8	Yes	Yes

Table 9. Evaluation of Proposed Grade 3 Math Specifications

Reporting Category	% Standards Transferable (N)	% Clusters Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Operations & Algebraic Thinking	100% (9)	100% (4)	16	18	Yes	Yes
Number & Operations in Base Ten	100% (3)	100% (1)	6	14	Yes	Yes
Number & Operations - Fractions	100% (3)	100% (1)	8	11	Yes	Yes
Measurement & Data	100% (8)	100% (4)	10	23	Yes	Yes
Geometry	100% (2)	100% (1)	5	11	Yes	Yes

Table 10. Evaluation of Proposed Grade 4 Math Specifications

Reporting Category	% Standards Transferable (N)	% Clusters Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Operations & Algebraic Thinking	100% (5)	100% (3)	10	12	Yes	Yes
Number & Operations in Base Ten	100% (6)	100% (2)	9	11	Yes	Yes
Number & Operations - Fractions	100% (4)	100% (3)	12	14	Yes	Yes
Measurement & Data	100% (7)	100% (3)	7	9	Yes	Yes
Geometry	100% (3)	100% (1)	4	6	Yes	Yes

Table 11. Evaluation of Proposed Grade 5 Math Specifications

Reporting Category	% Standards Transferable (N)	% Clusters Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Operations & Algebraic Thinking	100% (3)	100% (2)	5	7	Yes	Yes
Number & Operations in Base Ten	100% (7)	100% (2)	14	16	Yes	Yes
Number & Operations - Fractions	100% (7)	100% (2)	14	16	Yes	Yes
Measurement & Data	100% (5)	100% (3)	7	9	Yes	Yes
Geometry	100% (4)	100% (2)	5	7	Yes	Yes

Table 12. Evaluation of Proposed Grade 6 Math Specifications

Reporting Category	% Standards Transferable (N)	% Clusters Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Ratios and Proportional Relationships	100% (3)	100% (1)	10	12	Yes	Yes
The Number System	100% (8)	100% (3)	9	11	Yes	Yes
Expressions and Equations	100% (9)	100% (3)	16	18	Yes	Yes
Geometry	100% (4)	100% (1)	5	7	Yes	Yes
Statistics and Probability	100% (5)	100% (2)	5	7	Yes	Yes

Table 13. Evaluation of Proposed Grade 7 Math Specifications

Reporting Category	% Standards Transferable (N)	% Clusters Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
Ratios and Proportional Relationships	100% (3)	100% (1)	11	14	Yes	Yes
The Number System	100% (3)	100% (1)	10	13	Yes	Yes
Expressions and Equations	100% (4)	100% (2)	13	15	Yes	Yes
Geometry	100% (6)	100% (2)	8	10	Yes	Yes
Statistics and Probability	100% (8)	100% (3)	5	7	Yes	Yes

Table 14. Evaluation of Proposed Grade 8 Math Specifications

Reporting Category	% Standards Transferable (N)	% Clusters Transferable (N)	Min Number of Items	Max Number of Items	Meets All Test Specs	Supports Valid Measurement
The Number System	100% (2)	100% (1)	4	6	Yes	Yes
Expressions and Equations	100% (8)	100% (3)	16	18	Yes	Yes
Functions	100% (5)	100% (2)	11	13	Yes	Yes
Geometry	100% (9)	100% (3)	9	11	Yes	Yes
Statistics and Probability	100% (4)	100% (1)	4	9	Yes	Yes

Table 15. Evaluation of Proposed Grade 9 Math Specifications

Reporting Category	% Standards Transferable (N)	%Domains Transferable (N)	%Clusters Transferable (N)	Min N of Items	Max N of Items	Meets All Test Specs	Supports Valid Measurement
Number and Quantity	100% (8)	100% (3)	100% (5)	5	7	Yes	Yes
Algebra	100% (23)	100% (4)	100% (10)	18	24	Yes	Yes
Functions	96.0% (24)	100% (4)	100% (9)	14	18	Yes	Yes
Statistics and Probability	87.0% (20)	100% (4)	87.5% (7)	5	7	No	Yes
Geometry	97.2% (25)	100% (6)	100% (14)	6	8	Yes	Yes

Note. Standards are flagged if rated as having low to no item transferability due to changes in the standard. Domains/clusters are flagged if all standards in the domain/cluster are flagged. Meets all test specs if transferable items are available to meet all required standards/clusters/domains. Supports valid measurement if it can meet the minimum number of items for the reporting category from the available domains, clusters, and standards.

Table 16. Evaluation of Proposed Grade 10 Math Specifications

Reporting Category	% Standards Transferable (N)	%Domains Transferable (N)	%Clusters Transferable (N)	Min N of Items	Max N of Items	Meets All Test Specs	Supports Valid Measurement
Number and Quantity	100% (9)	100% (2)	100% (5)	5	7	Yes	Yes
Algebra	100% (22)	100% (4)	100% (10)	10	14	Yes	Yes
Functions	96.0% (24)	100% (4)	100% (9)	7	11	Yes	Yes
Statistics and Probability	87.0% (20)	100% (4)	87.5% (7)	6	8	No	Yes
Geometry	100% (36)	100% (6)	100% (14)	18	24	Yes	Yes

Note. Standards are flagged if rated as having low to no item transferability due to changes in the standard. Domains/clusters are flagged if all standards in the domain/cluster are flagged. Meets all test specs if transferable items are available to meet all required standards/clusters/domains. Supports valid measurement if it can meet the minimum number of items for the reporting category from the available domains, clusters, and standards.

Table 17. Evaluation of Proposed Grade 11 Math Specifications

Reporting Category	% Standards Transferable (N)	%Domains Transferable (N)	%Clusters Transferable (N)	Min N of Items	Max N of Items	Meets All Test Specs	Supports Valid Measurement
Number and Quantity	100% (9)	100% (3)	100% (5)	5	7	Yes	Yes
Algebra	100% (22)	100% (4)	100% (10)	16	22	Yes	Yes
Functions	96.0% (24)	100% (4)	100% (9)	14	18	Yes	Yes
Statistics and Probability	87.0% (20)	100% (4)	87.5% (7)	5	7	No	Yes
Geometry	100% (36)	100% (6)	100% (14)	8	11	Yes	Yes

Note. Standards are flagged if rated as having low to no item transferability due to changes in the standard. Domains/clusters are flagged if all standards in the domain/cluster are flagged. Meets all test specs if transferable items are available to meet all required standards/clusters/domains. Supports valid measurement if it can meet the minimum number of items for the reporting category from the available domains, clusters, and standards.

Tables 18 and 19 summarize the final ratings about item transferability, based on a comparison of the standards-to-standards crosswalk. Across all grade levels in both content areas, the vast majority of standards were rated as likely to allow for full item transferability. In other words, the revised standards were the same as or similar enough to the past standard that items written to measure the past standard would be expected to measure the revised standard, or transfer. For every grade and content area, some percentage of standards were rated as likely allowing for partial item transferability. In these cases, we anticipate that some, but potentially not all, items in the current item pool that are intended to measure the past standard could also be used to

measure the corresponding revised standard, but that this will require inspection on the part of ITP to determine which items are usable and to what degree. It is also possible that additional items may need to be developed to address new content included in the revised but not the original standards for standards rated as partially transferable. For the three high school math topics, less than 10% of standards were rated as likely allowing for low to no item transferability. We anticipate that substantial item development would be needed to address these more substantive changes in the standards. We have provided ITP with separate files containing standard-level ratings and associated rationales to support this effort.

Table 18. Final Percentage of Item Transferability Ratings for Reading

Grade	Likely Full Item Transferability	Likely Partial Item Transferability	Likely Low to No Item Transferability
3	82.4%	17.6%	0.0%
4	52.9%	47.1%	0.0%
5	58.8%	41.2%	0.0%
6	82.4%	17.6%	0.0%
7	76.5%	23.5%	0.0%
8	88.2%	11.8%	0.0%
9-10	88.2%	11.8%	0.0%
11	64.7%	35.3%	0.0%

Table 19. Final Percentage of Item Transferability Ratings for Mathematics

Grade	Likely Full Item Transferability	Likely Partial Item Transferability	Likely Low to No Item Transferability
3	80.0%	20.0%	0.0%
4	78.6%	21.4%	0.0%
5	88.5%	11.5%	0.0%
6	89.7%	10.3%	0.0%
7	66.7%	33.3%	0.0%
8	78.6%	21.4%	0.0%
Algebra 1	63.6%	27.3%	9.1%
Geometry	84.1%	13.6%	2.3%
Algebra 2	76.9%	21.2%	1.9%

Discussion

Research Question 1. Do the revised content standards reflect the same overall content domain as the original?

We found the revised content standards to reflect the same overall content domain as the original standards. The ELA standards are essentially identical in terms of their composition; both are organized around the College and Career (CCR) readiness anchor standards. The main source of difference in the two sets of standards was changes in the DOK level associated with some standards. All substantive changes to the content of the math standards are at the high school level. However, most standards were not changed enough to render available items unacceptable for measuring those standards.

Research Question 2. Does the organization of the test specifications reflect the organization of the standards?

We found that the test specifications clearly reflect the organization of the Iowa Academic Standards for ELA and mathematics while accounting for grade specific assessments. The coding schema of the standards is used in the test specifications, allowing for a clear delineation showing that the test specifications cover the content outlined in the standards. In high school math, the prior high school math standards contained a single Algebra topic, whereas the revised standards divide the content into Algebra 1 and Algebra 2 and the standards are organized by topic (e.g., Algebra 1, Geometry, Algebra 2). Test specifications are appropriately organized by grade level assessment such that a mapping of the non-grade standards to grade specific specifications was required. We were able to map the topic-based standards to the grade level-based test specifications to analyze the test specifications relative to reporting with relative ease.

Research Question 3. Does the organization of the test specifications support the production of valid and reliable scores at the level reported?

The ISASP item pool consists of items developed to measure the original standards. The alignment of items to the original standards was also verified via a traditional alignment study (Dickinson, et. al., 2019). This established alignment between items and standards served as the basis for evaluating how revisions to the standards might impact the suitability (i.e., transferability) of items to measure the new standards and, by extension, the ability to meet specifications and produce defensible scores and subscores. We found that for nearly all reporting categories, the transferability of items should allow for the targets outlined in the test specifications to be met.

The notable exception was the high school Statistics and Probability reporting category, where all the standards within a single cluster were changed substantially. Although the minimum number of items needed for this reporting category could be sourced from other clusters, the validity of the category scores may be limited by the omission of a cluster (i.e., breadth of coverage).

References

Butterfuss, R., and Doran, H. (2024, October). *An application of text embeddings to support alignment of educational content standards* [Paper presentation]. Generative Artificial Intelligence for Measurement and Education Meeting, Pittsburgh, PA, United States. https://hdoran.github.io/Blog/ContentMapping.pdf

Dickinson, E. R., Michaels, H. R., & Thacker, A. A. (2019). *Evaluating the alignment of the Iowa statewide assessment of student progress ELA, mathematics, and science assessments: Grades 3-11 criteria report* (2019 No. 035). Human Resources Research Organization.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education.* National Institute for Science Education.

Author Bios

Dr. Emily Dickinson is a Principal Scientist with nearly two decades of experience in conducting alignment studies. She routinely works with state departments of education and their testing vendors to create custom alignment criteria that appropriately address the test design. She has conducted alignment studies for California, Colorado, Delaware, Iowa, Kentucky, North Dakota, Mississippi, the National Assessment Governing Board (NAGB), and South Carolina. Her alignment reports have been successfully submitted as evidence for federal peer review. Dr. Dickinson served as the lead author of a peer reviewed publication on the topic of alignment. She has served as a federal peer reviewer since 2022.

Dr. Caroline Wiley has over two decades of experience in educational research and contributes expertise in quantitative and qualitative data analysis, assessment and program evaluation, and large-scale data management. Her background includes serving as technical task lead for projects aligning major testing programs with national and state academic standards, such as the Common Core State Standards, designing and facilitating alignment studies, and conducting interviews and focus groups. Recent leadership roles include directing evaluations of psychometric processes, overseeing rigorous impact evaluations for K-12 programs, and providing general support and guidance on innovative research design, data visualization, and systematic improvements to data management systems.

Dr. Jon Twing is a Chief Scientist at HumRRO. He has over 40 years of experience in experimental psychology, psychometrics, research, and assessment development. He is focused on the development, validity, and efficacy of learning systems that integrate measurement and instruction and the defensibility of measures. Dr. Twing has provided measurement expertise and consulting services to various organizations such as the Organisation for Economic Co-operation and Development (OECD), the Abu Dhabi Education Council, India's Central Board of Secondary Education (CBSE), the Australian Curriculum Assessment and Reporting Authority (ACARA), and the National Center for Educational Statistics (NCES) working on the National Assessment of Educational Progress (NAEP). He is a Research Fellow at the University of Sydney (and previously at the University of Oxford), where he guides staff, students, and policy regarding the scientific measurement of cognitive functions. Dr. Twing serves or has served as a member of numerous boards, committees, and service organizations, including the International Testing Commission (ITC), the College Board Psychometric Panel, the American Educational Research Association (AERA), the National Council on Measurement in Education (NCME), the Council of Chief State School Officers (CCSSO), the Association of Test Publishers (ATP), the American Psychological Association (APA), the University of Iowa College of Education Advisory Board, and the ACT Aspire Board of Directors. Notably, he was one of the pioneers in the application of Large Language Models (LLMs) in assessment, and his research provided validity evidence for the efficacy of automated scoring using latent-semantic analysis via large-language models

Dr. Arthur (Art) Thacker is a Chief Scientist at HumRRO who brings more than 30 years of relevant educational research experience to this project as an assessment researcher. His clients include district, state, and federal education agencies, private education organizations, professional standards boards, and military assessment programs. Dr. Thacker has successfully conducted numerous alignment studies for state education agencies, many of which have been submitted as evidence for federal peer review (some pre-dated federal peer review requirements). He served on a panel of national experts to explore best practices for conducting alignment studies. He served on a technical advisory panel for WIDA to recommend alignment strategies for English language assessments.