ACS VENTURES
BRIDGING THEORY & PRACTICE

Alignment of the ACT to the Wisconsin Academic Standards in Science

Final Report

Susan Davis-Becker, Ph.D.

Table of Contents

Executive Summary

The Wisconsin Student Assessment System (WSAS) includes a series of assessments designed to measure what students know and can do in core academic areas. At the high school level (grade 11), this system includes the ACT tests of Reading, Math, English, Science, and Writing. As a part of the validation effort for the WSAS, the Wisconsin Department of Public Instruction (WDPI) sought an independent alignment study evaluating how the content of the ACT Science test aligned with the Wisconsin Standards for Science for High School (WSSH). ACS Ventures, LLC (ACS) was contracted to complete this independent alignment study in September of 2018. This report documents the processes, results, and findings from the alignment study.

The WSSH were translated into *Specifications for the Wisconsin State Science Assessment* which sets expectations for the Science Assessment in terms of what content would be measured and how performance would be reported. The specific process for this study was created to meet the needs of the WDPI and gather the information needed for documentation, reporting, and supporting the claims made based on the administration of the ACT to Wisconsin high school students as set forth in these *Specifications*. Subject matter experts (panelists) were asked to review the connections between the WSSH and the ACT by aligning the ACT College and Career Readiness Standards (CCRS) and the ACT content (items) to the WSSH. As a first activity in this process, panelists were provided training on the alignment study purpose, process, and the specific judgments they were to make. Over two days, the panelists completed two alignment tasks which included a combination of independent judgments and group consensus discussions.

Two overarching questions were identified to evaluate how well the ACT Science Test fulfilled the expectations in the *Specifications*:

1) Does the ACT Science test measure a sample of the knowledge, skills, and abilities included within the WSSH?
   The panel found that the ACT CCRS supported the measurement of knowledge and skills incorporating the Crosscutting Concepts and the Science and Engineering Practices within the context of the Disciplinary Core Ideas. In addition, the ACT Science items were all linked to multiple dimensions and provided a distribution of measurement across the DCIs, the SEPs, the CCCs, and the reporting categories. A few of the specific findings deviated from the targeted ranges outlined in the *Specifications* which should be reviewed when developing future forms of the test.

2) Is the content included in the ACT Science test at an appropriate cognitive level given the performance indicators in the WSSH?
   Overall, the panel found that the CCRS and ACT Science test items largely measured knowledge and skills in a way that aligned to the grade level expectations for cognitive process outlined within the standards. Some measurement spanned beyond the targeted range but likely reflects the larger purpose and design of the ACT.

An evaluation framework for validity evidence was applied to the process and results. Overall, there was a substantial amount of validity evidence supporting the outcomes of this study for use by WDPI.

Introduction

Alignment has been characterized several ways – one of the most common is from Webb (1997; 2007) who described alignment as "the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expected to know and do" (p. 3). This definition – and other similar ones used in published literature – suggest that alignment information should be considered a key source of validity evidence for the use and interpretation of educational test scores. The unified perspective of validity suggests evaluating sources of evidence based on the intended use and interpretation of test scores (e.g., Kane, 2006; Messick, 1989). As a key source of evidence in the test development process, it is important to ensure that test content supports these inferences by representing a sampling of the domain of the educational program (e.g., content framework, standards, test blueprint).

The ACT is designed to measure similar topics at the same grade level as Wisconsin's academic content standards. The specific findings from a formal independent alignment study are required for accountability and will provide valuable information for stakeholders about the use of the ACT.

Alignment Approach

The overall alignment process was designed through a collaborative effort between ACS and WDPI (including input from their Technical Advisory Committee). The specific judgments to be made were determined based on the organization of the WSSH, the organization of the ACT, and the use of the ACT as the Wisconsin high school assessment.

Wisconsin Standards for Science for High School

The Wisconsin Standards for Science "emphasize students should be engaging in three-dimensional science learning from kindergarten through grade 12, meaning they learn the content by engaging in the scientific and engineering practices while using the perspectives of the crosscutting concepts to think like scientists". (p. 2, Standards for Science document). The Standards are organized into sections: Crosscutting concepts (CCCs), Science and engineering practices (SEPs), and Disciplinary Core Ideas (DCIs). The CCCs and SEPs represent specific dimensions, but the DCI section is further divided into four disciplines: Life Science, Physical Science, Earth and Space Science, and Engineering, Technology, and the Application of Science. Within each discipline there are a number of Standards (29 in total) with each Standard statement indicating that students are engaging in three-dimensional learning. Specific exemplars of the expectations for student by grade span are organized by learning elements (topic areas, concepts) and performance indicators (exemplars only, not an exhaustive list). Given this organization, it is understood that any evaluation of alignment to these Standards should be designed in a way to evaluate how the assessment will measure multi-dimensional thinking.

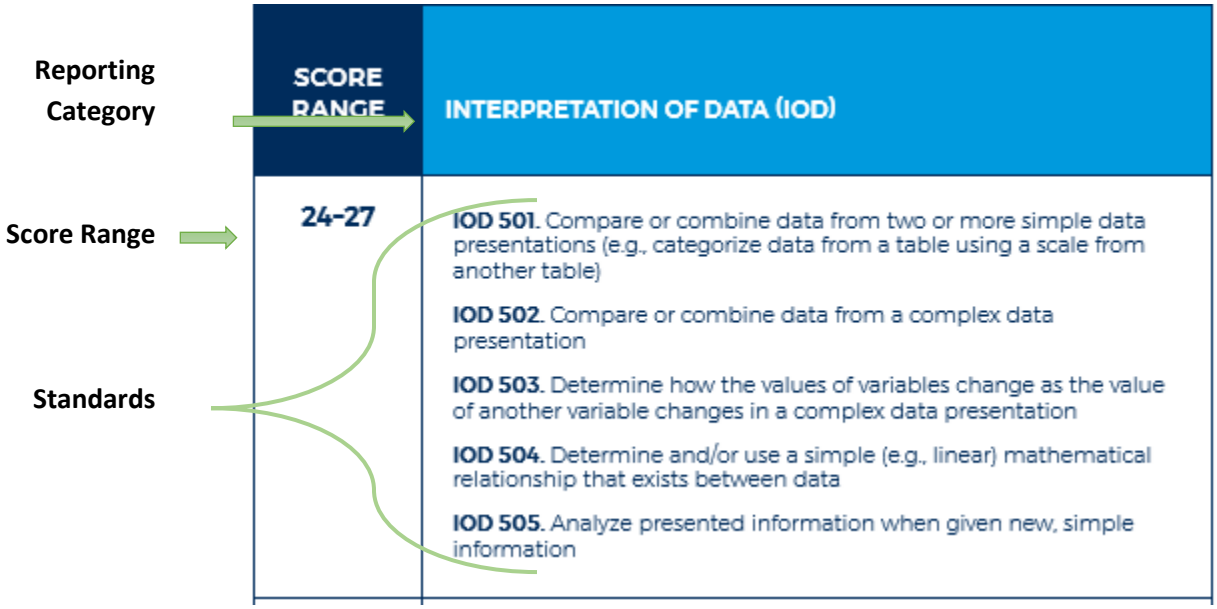
ACT Science Test

The ACT Science test has 40 multiple-choice questions per form that are organized into three reporting categories:

- Interpretation of Data
- Scientific Investigation
- Evaluation of Models, Inferences & Evaluation of Results

The knowledge and skills assessed within each reporting category are outlined in the ACT College and Career Readiness Standards (CCRS) by reporting category and score range: 13-15, 16-19, 20-23, 24-27, 28-32, 33-36 (see example in Figure 1).

Figure 1. Organization of ACT Content Categories, Score Ranges, and Standards



Assessment Specifications

Given the depth and breadth of the WSSH, WDPI created the *Specifications for the Wisconsin State Science Assessment* to describe how the range of expectations within the WSSH will be sampled for an assessment. These specifications combine the range of expectations across all standards with the reporting categories that are the basis for the ACT. At a very high level, WDPI set targets for the distribution of content across the three primary content areas (Life Science, Physical Science, Earth and Space Science) and the planned reporting categories (Interpretation of Data, Scientific Investigation, Evaluation of Models, Inferences, and Experimental Results). These target ranges are outlined in Table 1.

Table 1. Planned Distribution of Content for Wisconsin HS Science Assessment.

	Percent of Test
Distribution by Content	
Life Science	30-35%
Physical Science	30-53%
Earth and Space Science	15-35%
Distribution by Reporting Category	
Interpretation of Data	45-55%
Scientific Investigation	20-30%
Evaluation of Models, Inferences, & Experimental Results	25-35%

Figure 2 (taken from the *Specifications* document) shows the expected relationship among the reporting categories, Crosscutting Concepts, Science and Engineering Practices, and Disciplinary Core Ideas. As is shown in the Figure, the Disciplinary Core Ideas serve as the foundation through which students can demonstrate their ability to use the Crosscutting Concepts and Science and Engineering Practices to demonstrate the skills in the three reporting areas (Interpretation of Data, Scientific Investigation, and Evaluation of Models, Inferences, & Experimental Results).

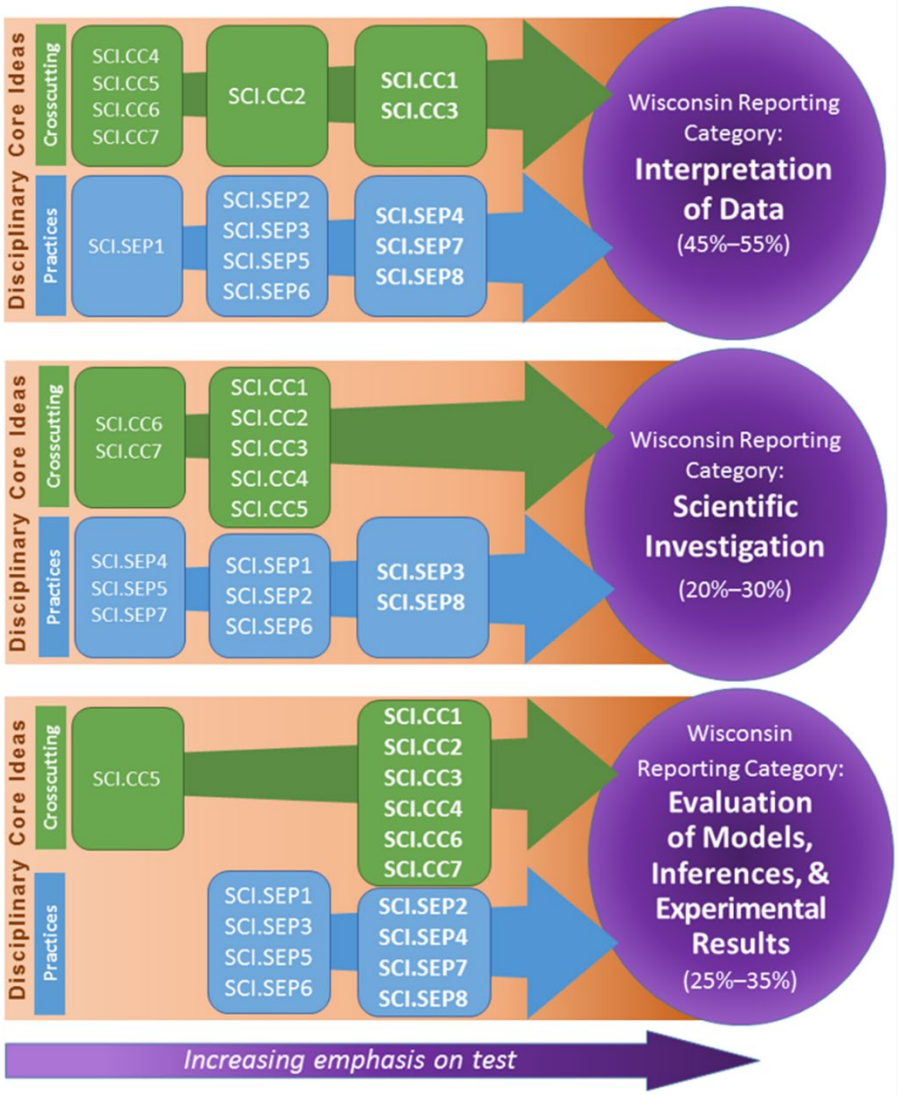


Figure 2. Intended Alignment of the WSSH to the Wisconsin Science Reporting Categories.

Alignment Approach

Given that the *Specifications* represent how the WSSH should be assessed, the approach to evaluating alignment in this study was designed around determining whether the ACT Science test met those specifications.

To evaluate this connection, two questions were identified to be addressed:

1) **Does the ACT Science test measure a sample of the knowledge, skills, and abilities included within the WSSH?**
   The first question targets whether the various dimensions of the WSSH are sampled. This includes measuring student ability within the Disciplinary Core Idea, integrating the Crosscutting Concepts, using the Science and Engineering Practices, and generating evidence that supports reporting in the three defined categories. To answer this question, panelists reviewed ACT Science test materials (standards, items) to find the connections to each of the dimensions and the reporting categories.

2) **Is the content included in the ACT Science test at an appropriate cognitive level given the performance indicators in the WSSH?**
   This second question evaluates whether the KSAs being measured are at the right level of rigor for high school students. While traditional alignment approaches focus on evaluating cognitive complexity as a match (e.g., DOK of a standard matched that of an item), aligning to multidimensional standards makes the question a bit more complex. In addition, the wording of the WSSH in each area intentionally integrates a high level of cognitive process as it incorporates the three dimensions. Therefore, to answer this question, panelists were tasked with evaluating whether the cognitive demand of the ACT Science test content fell within the grade-level expectations for cognitive processing outlined in the performance indicators for the aligned WSSH standards.

Alignment Tasks

As noted above, the review of ACT Science test materials included an evaluation at multiple levels in order to develop a comprehensive picture as to how the knowledge and skills outlined in the *Specifications* representing the WSSH were measured by the ACT. The first level of focus was the ACT CCRS[1] - panelists were asked to review each of the CCRS (56) and identify the alignment to the different areas of the WSSH. Because the CCRS describe a skill, the panelists were tasked with evaluating the measurement possibilities (i.e., how the CCRS could be assessed in conjunction with CCCs, SEPs, DCIs) by answering the following questions:

- **Which Crosscutting Concepts could be assessed in the context of this CCRS?**
  It was anticipated each ACT Science CCRS could be assessed incorporating a variety of Crosscutting Concepts.
- **Which Science and Engineering Practices could be assessed in the context of this CCRS?**
  It was anticipated each ACT Science CCRS could be assessed incorporating one or more Science and Engineering Practices.
- **In which Domains could this CCRS be assessed?**
  It was anticipated that a given ACT Science CCRS could be measured in multiple domains as these are written to be context free.
- **Do the knowledge and skills described in this CCRS fall within the range of expectations for students in high school? Or are they below/above the expectations for students in high school?**

[1] Full copies of the ACT Standards can be found in Appendix B or accessed here: http://www.act.org/content/act/en/education-and-career-planning/college-and-career-readiness-standards.html

It is expected that the cognitive demand of most CCRS will be targeting the high school grade band some would measure prerequisite skills (i.e., those CCRS that are targeted to lower score ranges). Panelists were tasked with identifying whether the CCRS knowledge/skill was within the performance indicators outlined for high school students within the aligned elements, above those indicators, or below those indicators.

In the second task, panelists reviewed three forms of the ACT Science test and identified the alignment to the *Specifications* (40 items per form X 3 forms = 120 total items). This rating task was similar to the CCRS task but with more specific judgments as to how the measurement in the item actually represents the different elements of the specifications. Panelists provided judgment on the alignment of each item by answering the following questions:

- **Which Crosscutting Concepts are being assessed by this item?**
  It was anticipated that most ACT Science test items would incorporate a Crosscutting Concept.
- **Which Science and Engineering Practices are being assessed by this item?**
  It was anticipated that most ACT Science test items would incorporate a Science and Engineering Practice.
- **In which Domain does this item fall?**
  It was anticipated that each ACT Science test item would measure knowledge within a content domain as each item is linked to a passage about a Science topic or activity.
- **Do the knowledge and skills measured by this item fall within the range of expectations for students in high school? Or are they below or above the expectations for students in high school?**
  It is expected that some items will measure pre-requisite skills (i.e., those targeted to lower score ranges) but that most will be targeting expectations that fall within those identified for the high school grade band standards. Panelists were tasked with identifying whether the knowledge/skills required to answer an item correct were within the performance indicators outlined for high school students within the aligned elements, above those expectations, or below those expectations. To note, the WSSH are written for high school but not a particular grade.

Data Collection

The expert judgments required to evaluate this alignment were gathered during an in-person meeting held in September of 2018. ACS was responsible for preparing all materials and facilitating the meeting. ACT staff provided access to the secure testing materials that were used throughout the study. WDPI recruited subject matter experts to serve as panelists based on their experience and expertise. Five panelists participated in the study representing different districts from across the state and had an average of 23 years of experience[2]. Three held master's degrees and two held Doctorate degrees (see Appendix A for detailed panelist information).

On the first day, panelists were welcomed by ACS (Dr. Susan Davis-Becker) who explained the purpose of the meeting, the rules regarding test security, and provided the panelists with an orientation to the overall process

[2] WDPI recruited a total of seven panelists but two panelists were unable to attend at the last minute due to unforeseeable circumstances.

and training on the specific judgments they were to make. During this first part of the process, panelists signed a confidentiality agreement with ACT and completed a demographic form documenting their expertise and experience.

To begin the first task (CCRS to WSSH), the ACS facilitator allowed the panel to align the first few CCRS together by discussing the targeted knowledge and skills, and how measurement of these might connect to the different dimensions of the WSSH. The panel then proceeded by working through the remaining CCRS one section at a time. For each CCRS, the panel made independent judgments as to the alignment and then came to consensus as a group. The final group consensus was recorded by the facilitator. When this task was complete, the facilitator provided additional training on the second task (ACT items to the WSSH) by reviewing a few sample items and allowing the panel to discuss thoughts on alignment as a group. The panel then moved through the remainder of the first form of the ACT (passage by passage) by making and recording independent alignment judgments and then discussing to come to consensus. The final consensus ratings were recorded by the facilitator. On the second day of the meeting, the panel repeated the process for Task 2 with the second and third form of the ACT. To complete the study, the panelists responded to an evaluation survey.

Results

The results are summarized by alignment question within this section of the report. The tables within this section provide a summary of the results from which the major findings and conclusions are drawn. The full set of consensus ratings can be found in Appendix B.

Does the ACT Science test measure a sample of the knowledge, skills, and abilities included within the WSSH?

The purpose of this alignment question was to evaluate how the various dimensions of the WSSH could be sampled (based on a review of the CCRS) and how exemplar ACT Science test forms did sample from the WSSH.

CCRS

The values in Table 2 indicate what percent of the CCRS, within each reporting category, could be measured integrating each CCCs, SEPs, and DCI. These do not indicate a balance or proportional representation on a test form as the CCRS were created from a large sample of ACT items. Rather, these findings indicate that a reasonable sample is possible across these three dimensions. As shown in the table:

- Almost all CCRS could possibly be assessed with all CCCs. The only exceptions were some CCRS within the *Interpretation of Data* category that were not linked to the *Structure and Function* CCC.
- All CCRS were linked to one or more SEPs. Within each performance category, the panel found more possible measurement opportunities with some SEPs than others.
- All CCRS could be assessed within any of the DCIs.

Table 2. CCRS Alignment Results to CCCs, SEPs, and DCIs.

	Interpretation of Data	Scientific Investigation	Evaluation of Models, Inferences & Experimental Results
Crosscutting Concepts			
CC1. Patterns	100%	100%	100%
CC2. Cause and effect	100%	100%	100%
CC3. Scale, proportion, and quantity	100%	100%	100%
CC4. Systems and models	100%	100%	100%
CC5. Energy and matter	100%	100%	100%
CC6. Structure and function	90%	100%	100%
CC7. Stability and change	100%	100%	100%
Science and Engineering Practices			
SEP1. Ask questions and define problems	14%	56%	6%
SEP2. Develop and use models	86%	28%	100%
SEP3. Plan and carry out investigations	5%	100%	0%
SEP4. Analyze and interpret data	100%	28%	65%
SEP5. Mathematics and computational thinking	62%	11%	12%
SEP6. Construct explanations and design solutions	24%	39%	41%
SEP7. Engage in argument from evidence	5%	11%	41%
SEP8. Obtain, evaluate and communicate information	100%	22%	47%
Disciplinary Core Ideas			
Life Sciences	100%	100%	100%
Physical Sciences	100%	100%	100%
Earth and Space Sciences	100%	100%	100%

Items

At the item level, the panelists identified the specific aspects of each dimension that were being measured and aligned an item to a reporting category. The results of this activity are shown in Table 3 by form. The percentages provided for the DCIs indicate the proportion of items on each form that were aligned to each of the content domains. The percentages provided for each reporting category indicate the proportion of items on each form that were aligned to each reporting category. Finally, the counts provided for the CCCs and SEPs indicate the number of items linked to each element within that dimension.

- The panel found some items on each form to align to each Domain and some items were identified as aligning to multiple content domains. In most cases, this was when an item was aligned to one of the target content domains in combination with the *Engineering, Technology, and the Application of Science* domain. The percentage of items, aligned to each domain, are shown in Table 3 compared to the target range from the *Specifications* document. As shown in the table, most domains on each form fell within the target range (these values are shown as ***bold and italicized***).

- Panelists identified the reporting category they felt an item was most likely targeting[3]. The percentage of items on each form linked to each reporting category is shown in Table 3 compared to the target range from the *Specifications* (these values are shown as ***bold and italicized***).
- Almost all items (93%) were found to incorporate one or more Crosscutting Concepts. Several of the CCCs were aligned more frequently (e.g., *Systems and Models*, *Cause and Effect*) but overall, content on each form was linked to four or more CCCs.
- All items were found to incorporate one or more of the Science and Engineering Practices. Several of the SEPs were aligned more frequently (e.g., *Develop and Use Models*, *Analyze and Interpret Data*) but overall, content on each form was linked to six or more CCCs.

Table 3. Item Alignment Results to DCIs, Reporting Categories, CCCs, and SEPs (by form)

	Target	Form A	Form B	Form C
Disciplinary Core Idea (Content Domain)				
Life Sciences	30-35%	***35%***	***33%***	25%
Physical Sciences	30-53%	***50%***	***53%***	***48%***
Earth and Space Sciences	15-35%	***18%***	10%	***18%***
Engineering, Technology, and the Application of Science	--[4]	33%	35%	25%
Reporting Categories				
Interpretation of Data	45-55%	***48%***	60%	***53%***
Scientific Investigation	20-30%	15%	***20%***	18%
Evaluation of Models, Inferences, & Experimental Results	25-35%	38%	20%	***30%***
Crosscutting Concepts				
CC1. Patterns		7	0	3
CC2. Cause and effect		5	5	12
CC3. Scale, proportion, and quantity		0	14	1
CC4. Systems and models		20	12	19
CC5. Energy and matter		0	1	1
CC6. Structure and function		0	8	3
CC7. Stability and change		0	0	1
None		8	0	0
Science and Engineering Practices				
SEP1. Ask questions and define problems		0	0	0
SEP2. Develop and use models		6	2	4
SEP3. Plan and carry out investigations		3	6	8

[3] During item development, ACT identified the targeted reporting category. This was compared to the panelists' judgments post study. Overall, the panel agreed with the ACT judgment 71% of the time. Overall, the panel often found themselves debating the nuances between particular aspects of two reporting categories and these minor differences may have resulted in some of the disagreement with the intended alignment.

[4] This domain (ETS) was identified as being measured in conjunction with other domains and therefore there is not a specific target for the percent of items.

Target	Form A	Form B	Form C
SEP4. Analyze and interpret data	24	26	19
SEP5. Mathematics and computational thinking	1	3	0
SEP6. Construct explanations and design solutions	0	1	2
SEP7. Engage in argument from evidence	1	1	1
SEP8. Obtain, evaluate and communicate information	5	1	6
None	0	0	0

In summary, the evidence from the alignment of the CCRS and the ACT Items suggest that the ACT Science test can and does measure from a sample of Disciplinary Core Ideas (Content Domains), CCCs, and SEPs in a way that supports the intended reporting scheme with minor deviations from the targets in the *Specifications*.

Is the content included in the ACT Science test at an appropriate cognitive level given the performance indicators in the WSSH?

This second question was designed to evaluate whether the KSAs being measured are at the right level of rigor for high school students. While traditional alignment approaches focus on evaluating cognitive complexity as a match (e.g., DOK of a standard matched that of an item), aligning to multiple multidimensional standards makes the question a bit more complex. In addition, the wording of the WSSH in each area intentionally integrates a high level of cognitive process as it incorporates the three dimensions. However, each standard includes a series of performance indicators for high school (and other grade bands). These performance indicators outline the level of knowledge and skills that students are expected to apply when demonstrating competency related to that standard. Panelists were asked to consider the collective performance indicators for the aligned dimensions/standards and determine if the knowledge and skills described in each CCRS or measured by each item were within the range of expectations outlined for high school, above those expectations, or below those expectations.

CCRS

Table 4 below shows the results of the grade level evaluation for the CCRS within each reporting category. Approximately half of the CCRS within each reporting category were found to be within the grade-level expectations for cognitive processing. The remaining CCRS were found to represent knowledge and skills that were either above or below the grade level expectations. For the *Interpretation of Data* and *Scientific Investigations*, these were roughly balanced (1/4 of CCRS below grade level, 1/4 of CCRS above). The CCRS for the *Evaluation of Models, Inferences, & Experimental Results* appears to include more knowledge/skills that are above the grade level expectations rather than below. When looking across the score scale range (categories defined by ACT organization of the CCRS), the expected pattern emerges. The CCRS identified as measuring knowledge and skills below the grade level expectations for cognitive processes represent the lower score ranges on the ACT, those within the grade level expectations for cognitive processes represented the middle of the ACT score scale, and finally those representing knowledge and skills above grade level represent the highest ACT score ranges.

Table 4. Evaluation of CCRS against Grade Level Expectations (by score range)

	Total	13-15	16-19	20-23	24-27	28-32	33-36
Interpretation of Data							
Below	24%	3	2	0	0	0	0
Within	52%	0	2	4	5	0	0
Above	24%	0	0	0	0	3	2
Scientific Investigation							
Below	22%	2	1	1	0	0	0
Within	50%	0	2	4	1	1	1
Above	28%	0	0	0	2	1	2
Evaluation of Models, Inferences, & Experimental Results							
Below	6%	1	0	0	0	0	0
Within	53%	0	2	4	3	0	0
Above	41%	0	0	0	2	3	2

Items

At the item level, panelists identified whether the assessed knowledge and skills fell within the grade level expectations described for each of the aligned standards. As shown in Table 5, the majority of items on each form were found to fall within the grade-level expectations for cognitive processes. The overall distribution was very similar for Forms B and C whereas Form A had more items falling Below and Above grade level.

Table 5. Evaluation of Items against Grade Level Expectations (by Form)

	Form A	Form B	Form C
Below	20%	18%	13%
Within	58%	73%	78%
Above	23%	10%	10%

In summary, the evidence from the evaluation of grade level appropriateness of the ACT Science test suggests that the majority of content on the exam does target the grade-level knowledge and skills expected for students in high school based on the WSSH. The content that falls outside this range is likely a reflection of the purpose of the ACT which is to measure college readiness of students across a range of abilities.

Panelist Evaluation

The results of the post-study evaluation are shown in Table 6. Panelists provided feedback as to how prepared they felt for each task, the time allocated to each part of the training, their confidence in the tasks they completed, the time allocated to complete the tasks, and the overall success of the study. As is shown in the Table, the panel generally felt prepared for each task, the right amount of time was allocated to each training component, confidence in their completed work, and that there was an appropriate amount of time allocated to completing each task. The panelists were also allowed to provide comments on their perceptions of the alignment study which are included in Appendix C.

Table 6. Evaluation results by Panel

	Median
Rate how well the Training prepared you for each of the alignment tasks *1=Not Prepared, 2=Somewhat Prepared, 3=Very Prepared*	
Task 1 - Alignment of the ACT CCRS to the WSSH	3
Task 2 - Alignment of the ACT test forms to the WSSH	3
Rate the Time allocated to Training *1=Not enough time, 2=The right amount of time, 3=Too much time*	
Task 1 - Alignment of the ACT CCRS to the WSSH	2
Task 2 - Alignment of the ACT test forms to the WSSH	2
Rate your confidence in each of the alignment tasks *1=Not Very Confident, 2=Somewhat Confident, 3=Very Confident*	
Task 1 - Alignment of the ACT CCRS to the WSSH	3
Task 2 - Alignment of the ACT test forms to the WSSH	3
Rate the time allocated to each task *1=Not enough time, 2=The right amount of time, 3=Too much time*	
Task 1 - Alignment of the ACT CCRS to the WSSH	2
Task 2 - Alignment of the ACT test forms to the WSSH	2
Overall success of the alignment study *1= Totally Unsuccessful, 2=Unsuccessful, 3=Successful, 4=Totally Successful*	4

Overall Findings

Based upon the review of the ACT Science test at multiple levels and comparison of the measured knowledge and skills to the *Specifications* set for assessment of the WSSH, several key findings were identified for the WDPI as they evaluate the use of the ACT. Among the critical findings are:

1) **Does the ACT Science test measure a sample of the knowledge, skills, and abilities included within the WSSH?**
   The panel found that the ACT CCRS supported the measurement of knowledge and skills incorporating the Crosscutting Concepts and the Science and Engineering Practices within the context of the Disciplinary Core Ideas. In addition, the ACT Science items were all linked to multiple dimensions and provided a distribution of measurement across the DCIs, the SEPs, the CCCs, and the reporting categories. A few of the specific findings deviated from the targeted ranges outlined in the *Specifications* which should be reviewed by WDPI and ACT when developing future forms of the test.

2) **Is the content included in the ACT Science test at an appropriate cognitive level given the performance indicators in the WSSH?**
   Overall, the panel found that the CCRS and ACT Science test items largely measured knowledge and skills in a way that aligned to the grade level performance indicators outlined within the WSSH. As expected, some measurement spanned beyond targeted range but likely reflects the larger purpose and design of the ACT.

Based on the results of the evaluation, the panel comprised to complete this alignment activity indicated that they felt comfortable completing the work and that they indicated a high degree of confidence in the results.

Evaluation of Validity Evidence

To evaluate the alignment study, we applied the validation framework suggested by Davis-Becker and Buckendahl (2013). Within this framework, the authors suggested four sources of evidence that should be considered in the validation process: procedural, internal, external, and utility. Threats to validity observed in these areas should mitigate policymakers' judgments regarding the usefulness of the results and the validity of the interpretation. Evidence within each of these areas that was observed in this study is discussed here.

Procedural

Procedural evidence was available when considering panelist selection and qualifications, choice of methodology, application of the methodology, and panelists' perspectives about the implementation of the methodology. For this study, the recruited panel included experienced educators in various roles from across the state. In addition, the panelists were independent of any development and validation activities for the ACT. Completing two levels of alignment judgments allowed for the analyses to provide two perspectives on how the ACT aligned with the *Specifications* for measuring the WSSH. In addition, modifications to the traditional judgments of cognitive complexity were made to reflect the organization and multidimensional nature of the WSSH. Panelists' perspectives on the process were collected and the evaluation responses were consistently positive.

Internal

The internal evidence for alignment studies can be evaluated by examining the consistency of panelists' ratings and the convergence of the recommendations. For this study, the rating tasks and decision rules were based on agreement/consensus judgments and each of the panels were consistently able to reach consensus judgments for each of the tasks assigned to them. Although the results should not be interpreted as unanimous support for every judgment by the panelists, the panelists worked well together in evaluating differences of opinion to calibrate amongst themselves on each judgmental task and determine the most appropriate consensus judgment. In addition, their evaluation ratings suggested they were confident in the results and for the consensus activities this includes the known final judgments on alignment.

External

External evidence is often the most difficult to obtain but in the design of this study, the ability to look at the results across levels provided one source of such evidence. Between the two markers, several indicators of consistency were found:

- All ACT CCRS and ACT Science items were aligned to multiple dimensions of the WSSH.
- At the Content Domain level, the panel identified the CCRS as being written in a way that could support measurement in any Domain and that the items were distributed across the domains.
- Across the Crosscutting Concepts, the panel found the CCRS could integrate most (if not all) of these concepts and found that the items on each form targeted a range of concepts.
- Across the Science and Engineering Practices, the panel found that each CCRS could likely integrate one or more practice and found that the items on each form targeted a range of practices.

- When evaluating the ACT for grade-level appropriateness, both levels of evaluation yielded a distribution largely focused on grade-level expectations with the remaining content split below and above the grade level expectations.

Utility

Evidence of utility is based largely on the extent to which the summative and formative feedback can be used to inform policy and operational decisions related to the interpretation of exam scores. This study reviewed the alignment of the ACT to the *Specifications* for measurement of the WSSH, using both ACT items and the ACT CCRS across the entire score range. As a result, we believe that the summative information from the study provide the evidence necessary for Wisconsin to move forward with the use of the ACT as their high school statewide assessment for Science representing a sampling of the WSSH. The deviations observed from the targets in the *Specifications* and some of the between form difference should be investigated further by WDPI and ACT as they design future test forms.

References

Davis-Becker S. & Buckendahl, C. W. (2013). A proposed framework for evaluating alignment studies. *Educational measurement: Issues and practice, 32*(1), 23-33.

Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education (Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6). Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education, 20*(1), 7-26.

Appendix A: Demographic Information of Panelists

Table A.1 provides additional detail on the demographic information for the five[5] panelists.

Table A.1 Panelist Demographic Information

Panelist	Employer	Title	Grades	Years of Experience	Highest Degree
1	Milwaukee Public Schools	Director of Research, Assessment, and Data	All	32	Ph.D.
2	Waunakee Community School District	Director of Curriculum and Instruction	PK-12	28	M.A.
3	School District of Oakfield	Science Teacher	10-12	15	M.A.
4	MAPSD (Medford)	Teacher	9-12	6	M.A.
5	Madison Metropolitan School District	Director of Assessment	K-12	33	Ed.D.

[5] WDPI recruited a total of seven panelists but two panelists were unable to attend at the last minute due to unforeseeable circumstances.

Appendix B: Consensus Alignment Ratings

The file embedded below includes the consensus ratings for each task completed by the panelists.

Consensus Ratings.xlsx

Appendix C: Evaluation Comments

- Great job!
- Great team. Valuable work done these past 2 days.
- Very well organized and facilitated.
- Enjoyed the experience very much and this deepened my understanding of the ACT specifications and the WI Science standards.