

# EXHIBIT A

## **Rebuttal Expert Report of Gregory J. Cizek, PhD**

### **In Re Delaware Public Schools Litigation (C.A. No. 2018-0029-VCL)**

**May 29, 2020**

#### **A. INTRODUCTION**

I was asked by counsel for the state defendants to review and evaluate the expert report submitted by Dr. Andrew Ho in the above matter. The primary purpose of this report is to respond to certain conclusions offered in that report; this rebuttal also addresses to a lesser extent similar points contained in the expert report submitted by Dr. Clive Belfield. This rebuttal comprises seven specific responses presented in Section B, below, and Conclusions (Section C).

#### **B. ANALYSES**

##### **1) Points of Agreement Related to Delaware State Testing**

*The report submitted by Dr. Ho makes numerous points regarding the Delaware/Smarter Balanced Assessment Consortium (SBAC) statewide assessment program that concur with points made in my previously-submitted expert report.*

Regarding the psychometric quality of the Delaware/SBAC assessment program design and standard setting procedures, Dr. Ho and I appear to agree on at least six major points:

- \* the content standards upon which the Delaware/Smarter Balanced assessments are based detail relevant, rigorous instructional targets for Delaware students;
- \* Delaware/SBAC assessments demonstrate strong alignment to the content standards;
- \* the process used to develop the Delaware/SBAC tests resulted in assessments that are fair, accessible, and free of bias;
- \* clear, well-specified achievement level descriptors (ALDs) were in place for setting performance standards on Delaware/SBAC assessment and that those ALDs were relied upon in the standard-setting process;
- \* the setting of performance standards (i.e., cut scores) for the Delaware/SBAC assessments used well-established, psychometrically sound procedures; and
- \* the actual standard-setting process was conducted faithfully to those procedures.

##### **2) Erroneous Use of the Term “Adequacy”**

*The report submitted by Dr. Ho (and Dr. Belfield) erroneously uses the term adequacy. The term is used in two different ways which are used interchangeably—and incorrectly so—in the report.*

The first way in which the term “adequacy” is used is in the context of the present Delaware Public Schools Litigation. In that context, it is the adequacy of Delaware’s public education *system* that is at issue.

The second way in which the term “adequacy” is used is in the context of describing what *level of*

*content mastery* Delaware students must demonstrate in order to be classified into one of the four performance levels (Level I - Minimal, Level II - Partial, Level III - Adequate; Level IV - Thorough). In the Delaware/SBAC standard setting activities, participants in that process were explicitly asked to consider adequacy in this second sense, not the first.<sup>1</sup> That is, they were explicitly asked to make judgments about the levels of knowledge and skill required for Delaware students to be labeled as demonstrating minimal mastery over the tested content, partial mastery over the tested content, adequate mastery over the tested content, or thorough mastery over the tested content. Participants in this process were explicitly directed to make their judgments exclusively as to levels of knowledge and skill to be demonstrated with respect to the content covered by the Common Core standards and not to attend to any policy considerations, political issues, personal notions of desirable goals, system considerations, or current levels of student performance.<sup>2</sup>

The Ho report conflates the two usages. For example, his report states that “Delaware teachers and subject-matter experts participated in the SBAC standard setting process and came to consensus about what constituted adequacy from their experiences with the Delaware students they taught.”<sup>3</sup> Documentary evidence from the standard setting process and observational evidence clearly indicates the specific design, intention, and conduct of the SBAC standard setting activities focused solely on adequacy of student subject matter knowledge and skill for purposes of determining performance levels, not on educational system adequacy. The report submitted by Dr. Belfield makes the same error.<sup>4</sup>

Further, beyond what participants were directed to affirmatively attend to, directions to participants in the standard setting activity specifically directed them *not* to incorporate attention to political implications of their judgments, current student performance levels, or any factors outside of purely content mastery judgments. The procedures explicitly directed participants *not* to attend to notions such as system adequacy. It is erroneous to conclude that participants were making any judgments about system adequacy.

### **3) Erroneous Understanding of the Nature of Exhortatory Performance Standard Setting**

*The report submitted by Dr. Ho mischaracterizes the intention of the Delaware/SBAC performance standards established by the standard setting participants.*

In my expert report, I characterized the purpose of the SBAC standard setting process as one designed to result in performance standards that represented what are sometimes called “stretch

---

<sup>1</sup> The author of this report has first-hand knowledge of these facts. I was retained by Smarter Balanced to serve as an independent, external evaluator of the entire standard setting process. Thus, I personally observed the activities described here. My observations and conclusions of the Smarter Balanced standard setting are documented in Cizek and Koons (2014).

<sup>2</sup> Documentation of the standard setting procedures indicates, for example, that workshop facilitators “reminded participants of the focus on claims across subjects, with the four ELA claims focusing on 1) Reading, 2) Writing, 3) Speaking/Listening, and 4) Research/Inquiry; and mathematics claims focusing on 1) Concepts and Procedures, 2) Problem Solving, 3) Modeling and Data Analysis, and 4) Communication and Reasoning... and indicated that the focus in the workshop would be on the threshold ALDs describing the knowledge and skills of students just entering Levels 2, 3, and 4.” See Cizek, G., & Koons, H. (2014, October), *Observation and report on Smarter Balanced standard setting*. Available from <https://portal.smarterbalanced.org/library/en/standard-setting-observation-andreport.pdf> (p. 10, emphasis added).

<sup>3</sup> Expert Report of Andrew Ho, Ph.D. (nd). p. 13.

<sup>4</sup> According to the Belfield report: “[W]e also identify many high school graduates as being inadequately educated. There are a number of indicators that high school graduation is not sufficient. First, there is the evidence on test scores of Delaware students. Many students in each tested grade are not considered proficient...” Belfield, C. R. (2020, March 12). *The economic benefits of increasing educational attainment of public schools in Delaware*, p. 4.

goals” for educational systems, educators, parents, and students. Other labels used to describe such standard setting goals are *aspirational* or *exhortatory*.

In his report, Dr. Ho questions this purpose of the Delaware/SBAC performance standards, asking “If this SBAC Level 3 description were intended to be an aspirational long-term goal, I cannot explain why it was described as ‘adequate,’ nor why a higher aspiration was not set...”<sup>5</sup>

Two facts provide the explanation. First, the performance description is characterized as “adequate” precisely because panelists were directed to derive a long-term goal for student achievement. The Level 3 performance level they sought to identify was one that would represent adequate performance for the future. That is, it was a level expected to be achievable over a period of time in which educational systems, teacher preparation, educator professional development, curriculum materials, and other components of educational systems would evolve to fully incorporate the teaching and learning of the more rigorous and challenging content standards represented by the Delaware/Common Core State Standards. Evidence of this is found in the SBAC standard setting report cited by Dr. Ho. In that report, it is documented that an orientation to the standard setting activity was provided to all panelists by Deborah Sigman (Co-chair of the Smarter Balanced executive committee). It is noted that Dr. Sigman “thanked participants for their involvement and their efforts with Smarter Balanced to improve teaching and learning for students.”<sup>6</sup> It is clear that the task set before standard setting participants was one of establishing challenging goals for the *future*—goals that would stimulate improvement in teaching and learning—and not simply reflecting current performance levels.

Second, an even higher aspiration was not set because the process used to set SBAC/Delaware standards considered the percentages of students that would *currently* be classified into the designated performance levels. As is best practice in standard setting, panelists (and, subsequently, the chief state school officers of states adopting the SBAC assessments) explicitly considered the impact of the panelists’ recommended cut scores for their reasonableness. That is, they were presented with the expected percentages of students that would be classified into each of the performance levels based on the panel-recommended cut scores, and they considered what levels of expectations would be so high as to be counterproductive. Panelists’ attention to this information (called *impact data*) is considered essential and is routinely incorporated into procedures for setting performance levels.

A third point is also relevant in response to Dr. Ho’s query: The even higher aspirational standard contemplated by Dr. Ho would fail to accomplish the intention of the “stretch” goals established by Delaware/SBAC standard setting participants, their chief state school officer and policy makers. It is well known that, in order for exhortatory standards to actually stimulate greater educational achievement, they cannot be set at a level that is so high as to discourage progress to the established goals. This principle was clearly articulated by Linn, Baker, and Betebenner who noted that:

“[H]aving a goal that is unobtainable no matter how hard teachers try can do more to demoralize than to motivate greater effort. Goals need to provide a challenge but not be set so high that they are unachievable.”<sup>7</sup>

---

<sup>5</sup> Expert Report of Andrew Ho, Ph.D. (nd). p. 14.

<sup>6</sup> See Cizek and Koons, 2014, p. 11.

<sup>7</sup> Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left behind Act of 2001. *Educational Researcher*, 31(6), p. 12.

#### **4) Erroneous Equating of SBAC Performance Levels as Defining Educational Adequacy in Delaware**

*The report submitted by Dr. Ho mischaracterizes the performance levels resulting from the SBAC standard setting process as equivalent to statements about educational adequacy in Delaware.*

In his report, Dr. Ho correctly observes that “The standards I have reviewed are Delaware state standards. The executive state summary described Delaware’s active participation in the achievement level setting process... The Delaware State Board of Education’s adoption of the achievement levels in January 2015, represents a legitimization of these standards and descriptions as those of Delaware, not some external entity.”<sup>8</sup>

Whereas Delaware educators were among the 482 total participants in the SBAC standard setting procedure, they comprised less than 5% (n =24) of the that group. Further, although the chief state school officers of SBAC member states voted unanimously (with two abstentions) to accept the performance standard recommendations,<sup>9</sup> the standards endorsed by SBAC member states and jurisdictions are appropriately considered to be consortium-wide performance standards. The performance standards ultimately adopted by the chief state school officers of each SBAC member were the result of negotiation and compromise among members spanning at least two in-person meetings (November 6, 2014 and November 14, 2014) at which review of proposed performance standards, discussion, negotiation, and eventually voting, occurred.<sup>10</sup> As the documentary evidence indicates, at no point during the process of considering the performance levels to be adopted by SBAC members was there any intention to create state-specific standards; rather, the intention was to create consortium-wide standards. Further, as was highlighted in Sections B3 and B4 above, when the performance standards were adopted by SBAC members, it was done with the singular purpose of gauging student achievement vis-à-vis the Common Core content standards.

#### **5) Erroneously Proposes Using a Test for a Purpose for Which It Has Not Been Validated**

*Although the Delaware/SBAC assessments have been designed to gauge mastery of content judged necessary for college or careers, the report submitted by Dr. Ho incorrectly accepts the use of Delaware/SBAC test results for the purpose of determining college readiness and educational adequacy—proposed uses that are considered professionally inappropriate.*

The report submitted by Dr. Ho implicitly recognizes the professional authority of the *Standards for Educational and Psychological Testing*.<sup>11</sup> For example, the report reflects acceptance of the authority of the *Standards* when it is asserted that “The *Standards for Educational and Psychological Testing* list the relationship between test scores and other variables as an important source of validity evidence.”<sup>12</sup>

However, validity evidence is only relevant with respect to a specific, intended interpretation of

---

<sup>8</sup> Expert Report of Andrew Ho, Ph.D. (nd). pp. 14-15.

<sup>9</sup> See *Smarter Balanced Technical Report*, Chapter 10, Achievement Level Setting, p. 28. Available at: <http://www.smarterbalanced.org/wp-content/uploads/2015/08/Chapter-10-Achievement-Level.pdf>

<sup>10</sup> *Smarter Balanced Technical Report*, Chapter 10, Achievement Level Setting, p. 28.

<sup>11</sup> American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

<sup>12</sup> Expert Report of Andrew Ho, Ph.D. (nd). p. 17.

test scores. To consider the validity evidence in support of Delaware/SBAC assessments, it is first essential to identify the specific, intended score meaning(s) to be validated.<sup>13</sup> The specific intended score meanings for the Delaware/SBAC assessments are clearly and carefully identified in the SBAC technical documentation. In addition to several related purposes (such as for use as accountability indicators, aiding educators in planning and improving instruction, etc.), the primary intention of Delaware/SBAC scores is identified as being the provision of information about “students’ ELA/literacy and mathematics achievement with respect to the CCSS measured by the ELA/literacy and mathematics summative assessments in grades 3 to 8 and high school.”<sup>14</sup>

Indeed, an abundance of evidence for the validity of Delaware/SBAC test scores has been obtained based on examining the relationships among those scores and other relevant variables<sup>15</sup> and it can be concluded from this evidence that there is ample support for the primary intended score meaning related to students’ ELA/literacy and mathematics achievement with respect to the Common Core State Standards (CCSS).

However, the same *Standards* cited in the Ho report regarding relevant sources of validity evidence also contain strict admonitions regarding inappropriate test use. Of particular note is that the *Standards* strongly discourage the use of test scores for any purpose for which they have not been validated. According to the *Standards*:

“If validity for some common or likely interpretation for a given use has not been evaluated...potential users should be strongly cautioned about making unsupported interpretations” and “[t]he improper use of tests...can cause considerable harm to test takers and other parties affected by test-based decisions.”<sup>16</sup>

The Delaware/SBAC assessments were designed to gauge mastery of content judged important for students’ subsequent pursuit of a college degree or for entry into diverse careers. For example, one of the listed intended purposes of grade 3-8 SBAC scores is to provide information as to “whether students prior to grade 11 have demonstrated sufficient academic proficiency in ELA/literacy and mathematics to be on track for achieving college-readiness.”<sup>17</sup>

It must again be stressed that claims about score meaning are *clearly and carefully* worded. Thus, the above claim regarding SBAC scores is that they provide information about *academic proficiency* and support interpretations about the extent to which students are *on track*. Thus, whereas there are many factors that contribute to college or career readiness, the Delaware/SBAC scores are only intended to reflect the academic knowledge and skills deemed important in post-secondary contexts. Further, the intended interpretation of Delaware/SBAC scores is *not* that they indicate college readiness, but that they indicate the likelihood of students being “*on track* for achieving college-readiness.”

This careful wording of the intended score meaning is evident in numerous locations. For example, Delaware Department of Education (DDoE) score interpretation guidelines describe the following

---

<sup>13</sup> See Cizek, G. J. (2020). *Validity: An integrated approach to test score meaning and use*. New York, NY: Routledge/Taylor and Francis.

<sup>14</sup> Smarter Balanced Assessment Consortium. (2019). *2017–18 summative technical report*. (p. 4). Available from <https://www.smarterbalanced.org/wp-content/uploads/2019/08/2017-18-Summative-Assessment-Technical-Report.pdf>

<sup>15</sup> See Smarter Balanced Assessment Consortium. (2019). *2017–18 summative technical report, chapter 1, validity*.

<sup>16</sup> *Standards for Educational and Psychological Testing* (2014) pp. 23, 1.

<sup>17</sup> Smarter Balanced Assessment Consortium. (2019). *2017–18 summative technical report*. (p. 4)

meaning of Level 3 performance:

“Level 3 - The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills of state standards in English language arts/literacy.”<sup>18</sup>

Another DDoE source indicates that:

“The Level 3 student demonstrates adequate understanding of and ability to apply the English language arts and literacy (mathematics) knowledge and skills needed for success in college and career, as specified in the Common Core State Standards.”<sup>19</sup>

In a guide for parents to aid their understanding of their children’s test scores, Delaware Secretary of Education Susan Bunting states:

“Unlike multiple choice tests of the past, these assessments put more emphasis on writing, solving problems and critical thinking. They were created specifically to measure students’ progress toward mastery of Delaware’s academic standards.”<sup>20</sup>

The intended meaning of Delaware/SBAC score is purposefully articulated in these careful constructions in order to avoid likely misinterpretations. Delaware/SBAC scores are *not* intended to be interpreted as students being college ready, but as mastering one aspect of that goal; namely, the Common Core based knowledge and skills that have been judged as necessary for students to be on-track for college readiness.

Several points in the report submitted by Dr. Ho reinforce this caution. For example, with respect to validity evidence for any claims about college readiness, the report recognizes that “information relating Delaware student test scores to future college and career outcomes is not yet available systematically” and “existing Delaware SBAC data do not enable prediction of college or career outcomes.”<sup>21</sup> These facts further support the conclusion that Delaware/SBAC scores cannot validly be used to support claims about system adequacy for ensuring students’ college or career readiness.

Finally, a review of the intended interpretations of Delaware/SBAC scores provided in the SBAC technical report<sup>22</sup> indicates that those assessments have not been validated for any purpose related to evaluation of educational programs or judgments about educational system adequacy.

*Similarly, although the SAT is designed to gauge acquisition of content judged necessary for college admission, the report submitted by Dr. Ho incorrectly equates performance on the SAT as*

---

<sup>18</sup> Sources: Smarter Balanced Assessment Consortium. (2013). *Initial achievement level descriptors and college content-readiness policy*. Santa Cruz, CA: Author. Retrieved from <https://portal.smarterbalanced.org/library/en/mathematics-alds-and-college-content-readiness-policy.pdf> and Delaware Smarter ELA/Literacy, Mathematics, and DCAS Science and Assessment Interpretive Guide: Understanding Your Child’s Test Scores, Spring 2017, p. 2 Available from [https://de.portal.airast.org/core/fileparse.php/2667/urlt/DE\\_Smarter\\_DCAS\\_Family\\_Guide.pdf](https://de.portal.airast.org/core/fileparse.php/2667/urlt/DE_Smarter_DCAS_Family_Guide.pdf).

<sup>19</sup> Delaware Department of Education. (2019, September). Delaware system of student assessments (DeSSA) executive state summary 2018-2019 administration. Dover, DE: Author. pp. 11-12.

<sup>20</sup> Delaware Smarter ELA/Literacy, Mathematics, and DCAS Science and Assessment Interpretive Guide: Understanding Your Child’s Test Scores, Spring 2017, p. 2. Available from [https://de.portal.airast.org/core/fileparse.php/2667/urlt/DE\\_Smarter\\_DCAS\\_Family\\_Guide.pdf](https://de.portal.airast.org/core/fileparse.php/2667/urlt/DE_Smarter_DCAS_Family_Guide.pdf).

<sup>21</sup> Expert Report of Andrew Ho, Ph.D. (nd). pp. 17, 18.

<sup>22</sup> Smarter Balanced Assessment Consortium. (2019). *2017–18 summative technical report*. (p. 4)

*the single measure of college readiness, and incorrectly accepts the use of SAT test results for the purposes of determining college and career readiness and educational adequacy—proposed uses that are considered professionally inappropriate.*

Both the Delaware/SBAC and SAT assessments have been designed to cover content judged important for students' subsequent pursuit of a college degree or for entry into diverse careers. However, although for convenience, a single score may be indicated as a college readiness benchmark, all such benchmarks—including the performance standards for the SAT adopted by Delaware—are established statistically as convenient but arbitrary reference points and are appropriately regarded as general guidelines, not strict indicators of college or career readiness.

In support of these points, it is necessary only to refer to the information provided by the College Board (responsible for the SAT). The College Board has described their reference points in terms of statistical probabilities of earning a specified grade in certain introductory college-level courses:

“The SAT Math benchmark is the SAT Math section score associated with a 75% chance of earning at least a C in first-semester, credit-bearing, college-level courses in algebra, statistics, precalculus, or calculus. The SAT Evidence-Based Reading and Writing benchmark is the SAT Evidence-Based Reading and Writing section score associated with a 75% chance of earning at least a C in first-semester, credit-bearing, college-level courses in history, literature, social science, or writing.”<sup>23</sup>

It is of note that the above are recently revised definitions of the benchmarks. Prior to 2017, the College Board specified the benchmarks as “a 65% probability of achieving an overall first-year college GPA of B– or higher.”<sup>24</sup>

Further, the College Board cautions users specifically *not* to interpret its readiness benchmarks as strictly defining “college ready” or “not college ready.” According to the College Board:

“It is important to note that college readiness is a continuum—students scoring below the SAT benchmarks can still be successful in college, especially with additional preparation and perseverance.”<sup>25</sup>

This caution from the College Board about avoiding strict (mis)interpretations of a specific benchmark as an indicator of college or career readiness reflects current research and policy. For example, Professor David Conley, who is widely regarded as one of the foremost experts on college and career readiness has indicated that:

“A student who is ready for college and career can qualify for and succeed in entry-level, credit-bearing college courses leading to a baccalaureate or certificate, or career pathway-oriented training programs without the need for remedial or developmental coursework. However, not every student requires the same proficiency in all areas. A student's interests and post-high school aspirations influence the precise knowledge and skill profiles necessary to be ready for postsecondary studies. Therefore, a single cut score on

---

<sup>23</sup> College Board. (nd). *College and career readiness benchmarks for the SAT suite of assessments*. New York, NY: Author. (p. 1). Available from <https://collegereadiness.collegeboard.org/pdf/educator-benchmark-brief.pdf>

<sup>24</sup> College Board. (nd). p. 2.

<sup>25</sup> College Board. “About Benchmarks.” Retrieved from <https://collegereadiness.collegeboard.org/about/scores/benchmarks>



a test given to high school students does not take into account this individualization of the match between knowledge and skills on the one hand, and aspirations on the other.”<sup>26</sup>

Because the concept of college or career readiness has too often been oversimplified as a single academic benchmark, Dr. Conley cautions that readiness considerations should consider “other important factors not addressed by the definition, such as positive citizenship, parental support and peer group influence, and, perhaps most importantly, student financial capability to attend college.”<sup>27</sup> As illustrated in Figure 1, his work provides a full accounting of the constellation of factors necessary to support conclusions about readiness, of which content knowledge is only one.

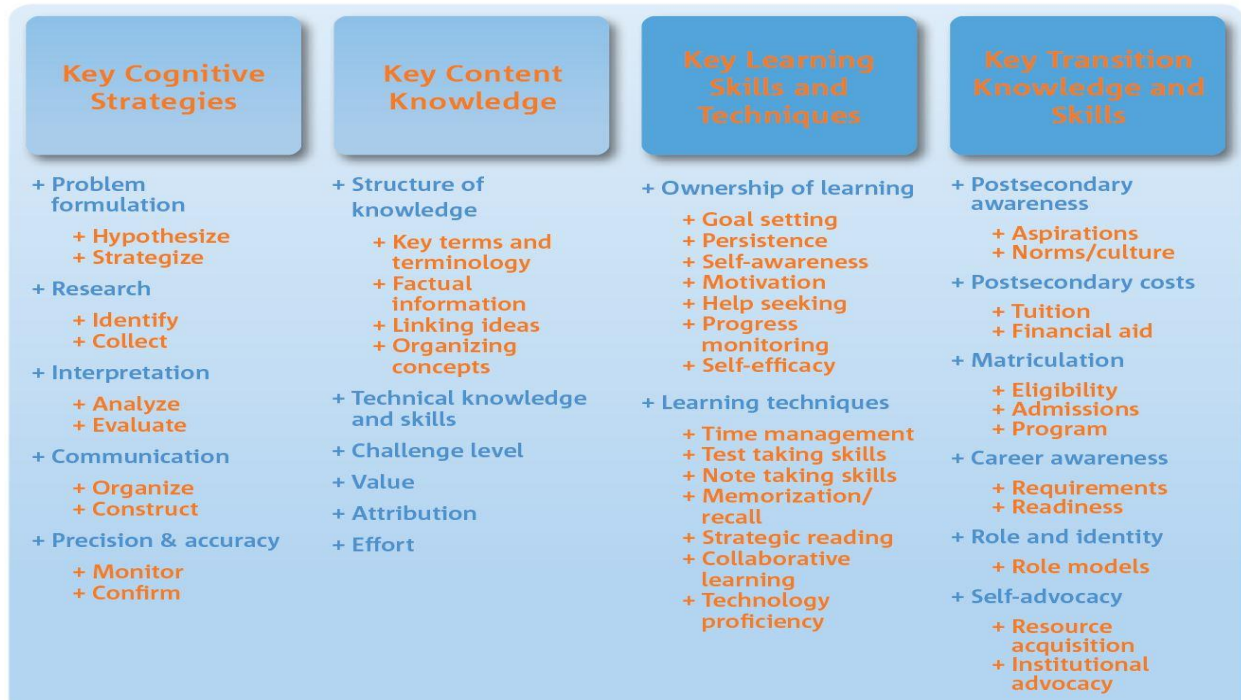


Figure 1.  
Four Keys to College and Career Readiness (Conley, 2012, p. 3)

Finally, as was the case with the Delaware/SBAC assessments, the SAT was not developed for and has not been validated for use in making determinations about educational system adequacy or other broader outcomes. A recent research brief from the College Board summarizes the validation of the SAT as limited to predicting freshman year college GPA, cumulative GPA, first year college course grades, college retention, and graduation.<sup>28</sup> It is clear from these sources of validity evidence (in addition to the content validity evidence described in the summary) that the SAT was not designed nor validated for predicting later life outcomes or for educational system evaluations.

The performance standards for the Delaware SAT were established by a consortium of states using

<sup>26</sup> Conley, D. T. (2012). A complete definition of college and career readiness. Eugene, OR: Educational Policy Improvement Center. (p. 1)

<sup>27</sup> Conley, 2012, p. 4.

<sup>28</sup> See Shaw, E. J. (2015). *An SAT validity primer*. New York, NY: The College Board.

a procedure similar to that used for the Delaware/SBAC assessments. According to the DDoE:

“Educators from Connecticut, Delaware, Maine, and New Hampshire reviewed item by item on the operational test form; discussed the expectations of student performance specified in the Achievement Level Descriptors (ALDs) for each achievement level and the impact data; and determined the cut points on the reporting scale based on the Modified Angoff approach.”<sup>29</sup>

Regarding validity evidence for its use in the Delaware System of Student Assessments, there is no evidence that supports the use of the SAT for system adequacy evaluation or similar policy purposes. To the contrary, it is specifically stated in DDoE documentation that “[t]he change [to using the SAT as the high school measure of academic achievement] originated at the request of legislators as the state continued to look for ways to reduce testing time, particularly for 11th grade students”<sup>30</sup> with the same goal (i.e., measurement of student mastery of state-adopted content standards) as the previous generation of content-based statewide assessments.

## **6) Ignores Evidence of Timeline for Enabling Students to Reach Higher Performance Levels**

*The report submitted by Dr. Ho fails to recognize that the performance standards and ambitious goals for increased Level 3 (or greater) performance in Delaware were intended to be accomplished on an extended timeline.*

As part of the original standard setting process and in the course of subsequent consideration of the Delaware/SBAC performance standards recommended to the SBAC-member chief state school officers, *impact data* were provided that showed the estimated percentages of students who would be classified into each of the four performance levels. These estimates—subsequently borne out following operational administration of the tests—indicated that large percentages of students would not initially reach Level 3 or higher. As described in my original report, these levels of performance were judged to be acceptable given the exhortatory nature of the performance standards and expectations. It was anticipated that it would take substantial time for educators, educational systems, and students/parents to adjust to the increased rigor of the Common Core standards and the performance level expectations in order to reach those goals. Thus, it is not only inappropriate to judge current performance against a future-referenced standard, but it is also inappropriate to fail to recognize progress toward that standard.

Evidence regarding such progress is available from various sources. First, as shown in my initial report, Delaware has made steady progress across the years of Delaware/SBAC test administrations 2015-2019.<sup>31</sup> Similar progress was demonstrated on an independent, external assessment, the National Assessment of Educational Progress (NAEP).<sup>32</sup>

It is noteworthy that the NAEP is named the National Assessment of Educational *Progress*,<sup>33</sup>

---

<sup>29</sup> Delaware Department of Education. (2019). *Delaware system of student assessments (DeSSA) executive state summary, 2018-2019 administration*. Dover, DE: Author. (p. 32).

<sup>30</sup> Delaware Department of Education. (2019). *Delaware system of student assessments (DeSSA) executive state summary, 2018-2019 administration*. Dover, DE: Author. (p. 29)

<sup>31</sup> See Figures 14-15 in Cizek, G. J. (2020, March). *Analysis of relationships between academic performance standards and educational adequacy*. (In Re Delaware Public Schools Litigation, C.A. No. 2018-0029-VCL.

<sup>32</sup> See Cizek, 2020, Figures 10-14.

<sup>33</sup> Italics added only for emphasis.

because judgments about the effectiveness of educational programs across the U.S. and on a state-by-state basis are not supported by percentages of students classified in a given performance category in any single year, but by examination of trends over time. Not only do the Delaware/SBAC and NAEP assessment trends reveal the progress made by Delaware within the state over time, and compared to other states over time, but Delaware’s *Every Student Succeeds Act* (ESSA) accountability plan also highlights the fact that educational progress is best evaluated longitudinally. The long-term achievement goals included in Delaware’s ESSA plan reveal increased expectations and a next generation of challenging, exhortatory goals for student performance by the year 2030. These goals were noted in my original report related to Smarter Balanced testing in grades 3-8.<sup>34</sup> The same ESSA plan also commits Delaware to challenging, exhortatory, long-term goals at the high school level; these goals are shown in Table 1 for all students, as well as for student subgroups, with long term goals indicated in bold.<sup>35</sup> For example, the 2030 goal for all students in English language arts is a 23.8% increase in proficiency (from the 2015-2016 starting point of 52.39% to 76.20%); the goal for mathematics is a 34.35% increase (from 31.31% to 65.66%).

Table 1.  
Delaware Long-Term ESSA Goals, SAT

Subgroup	ELA	ELA	Mathematics	Mathematics
	Starting Point (2015-2016)	Long-Term Goal (2030)	Starting Point (2015-2016)	Long-Term Goal (2030)
All students	52.39%	<b>76.20%</b>	31.31%	<b>65.66%</b>
Economically disadvantaged students	32.65%	<b>66.33%</b>	13.68%	<b>56.84%</b>
Children with disabilities	11.12%	<b>55.56%</b>	3.99%	<b>52.00%</b>
English learners	6.21%	<b>53.11%</b>	5.26%	<b>52.63%</b>
African American	32.50%	<b>66.25%</b>	13.36%	<b>56.68%</b>
American Indian or Alaska Native	70.37%	<b>85.19%</b>	33.33%	<b>66.67%</b>
Asian	74.38%	<b>87.19%</b>	62.93%	<b>81.47%</b>
Native Hawaiian/ Other Pacific Islander	40.00%	<b>70.00%</b>	10.00%	<b>55.00%</b>
Hispanic or Latino	38.71%	<b>69.36%</b>	17.53%	<b>58.77%</b>
White	64.93%	<b>82.47%</b>	42.36%	<b>71.18%</b>

<sup>34</sup> See Cizek, 2020, Table 5. It should be clarified that the goals cited in the original report were overall goals for the entire system, including grades 3-8 and high school, not exclusively goals for SBAC performance.

<sup>35</sup> Source: Source: Delaware Department of Education. (2019, June 10). *Delaware consolidated state plan under the Every Student Succeeds Act*. Dover, DE: Author. (p. 8)

## 7) Ignores Relevant External Comparisons

*The reports submitted by Dr. Ho and Dr. Belfield fail to take into account Delaware’s progress with respect to other states.*

As indicated previously, Delaware’s educational progress can be shown by examining increases in performance over time. However, additional perspective on that progress can be gained by considering Delaware’s standing among those in the Smarter Balanced Consortium. Two such analyses are described below.

First, it is useful to consider the starting point for gauging progress by examining Delaware’s standing compared to other states in the Smarter Balanced consortium at the time that the initial SBAC performance standards were established in 2014.

Figures 2 and 3 show the estimated impact of the Level 3 performance standards for English language arts and mathematics, respectively, recommended and endorsed by the SBAC standard setting panelists; that is, the figures show the estimated percentages of students that would be classified as Level 3 or higher based on the panelists’ recommendations. These data were endorsed by the panelists as reasonable at the conclusion of the standard setting process and, with only minor modifications, also endorsed by the chief state school officers of SBAC member states.

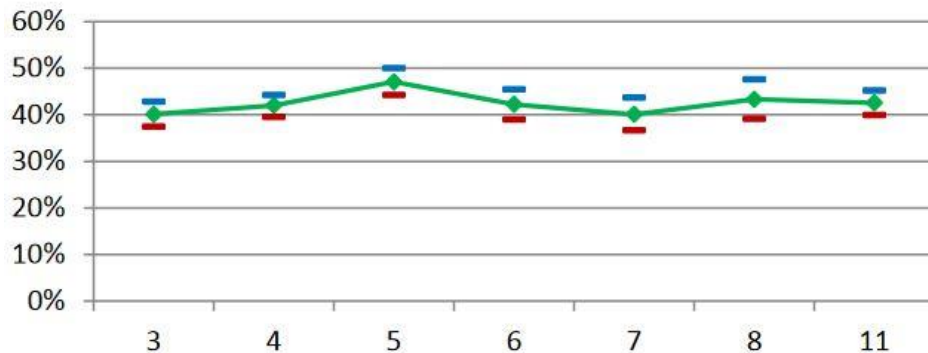


Figure 2.  
SBAC Estimated Impact Data, English Language Arts/Literacy<sup>36</sup>

<sup>36</sup> Smarter Balanced Assessment Consortium. (2015). *Achievement level setting final report*. Available from <https://portal.smarterbalanced.org/library/en/achievement-level-setting-final-report-with-appendix.pdf> (p. 11).

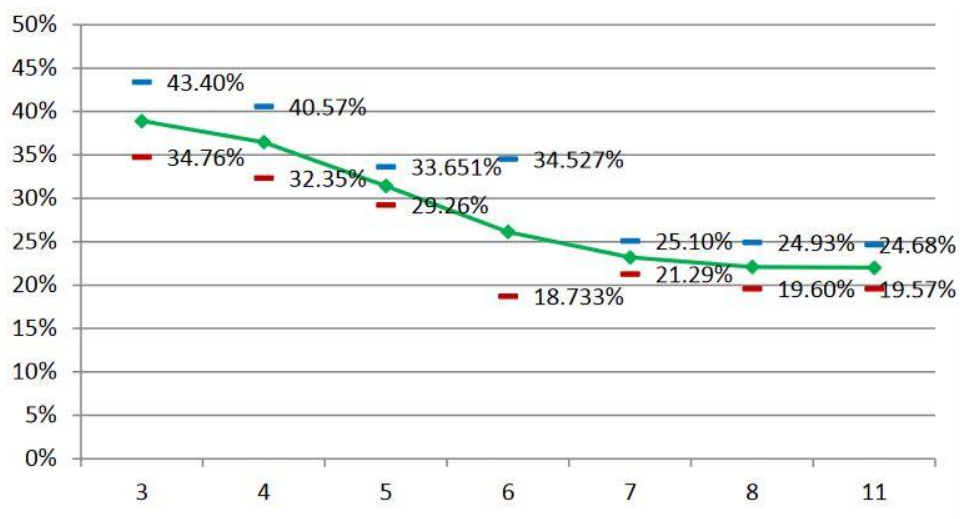


Figure 3.  
SBAC Estimated Impact Data, Mathematics<sup>37</sup>

The green lines in each figure show the estimated percentages of students that would be classified as Level 3/Adequate or above at grades 3-8 and 11. Figure 2 shows that the estimated impact for English was roughly between 40% and 48%, depending on grade level; the estimated impact for Mathematics ranged from slightly less than 40% for third grade to approximately 22% for 11th grade. Again, these estimated percentages of students that would be classified as Level 3 or above were presented, reviewed, and endorsed not only by the standard setting panelists, but also by the chief state school officers of each SBAC member state. It was clear at the time that Delaware's performance standards were established that only modest percentages of students would reach those standards initially. In sum, not only was stimulating progress toward increased performance an explicit goal of the standard setting activity, but there was a shared understanding among participants and state chiefs that adoption of the performance standards would result in modest percentages of students being classified at Level 3 or above in the initial years of implementation.

Second, it is relevant to view educational progress in Delaware since the initial standard setting in 2014 by comparing the percentages of students classified as Level 3 or above in Delaware with the percentages of such students in other SBAC member states. Such comparisons are essential to contextualizing educational progress in Delaware; both the Ho and Belfield reports fail to examine relevant educational progress measures in Delaware in the context of what is occurring nationally (with respect to NAEP results) and in the Smarter Balanced Consortium.

Figures 4 and 5 show the actual percentages of students across all grade levels who were classified at Level 3 or above in English language arts and mathematics, respectively, for members of the Smarter Balanced consortium of states since 2015.<sup>38</sup>

<sup>37</sup> Smarter Balanced Assessment Consortium. (2015). *Achievement level setting final report*. Available from <https://portal.smarterbalanced.org/library/en/achievement-level-setting-final-report-with-appendix.pdf> (p. 12).

<sup>38</sup> Source: Delaware Department of Education. (2019). *Smarter national comparison* [PowerPoint presentation]. The results shown in this report are based on data provided only by states that administered the Smarter Balanced tests in the same configuration as administered in Delaware.

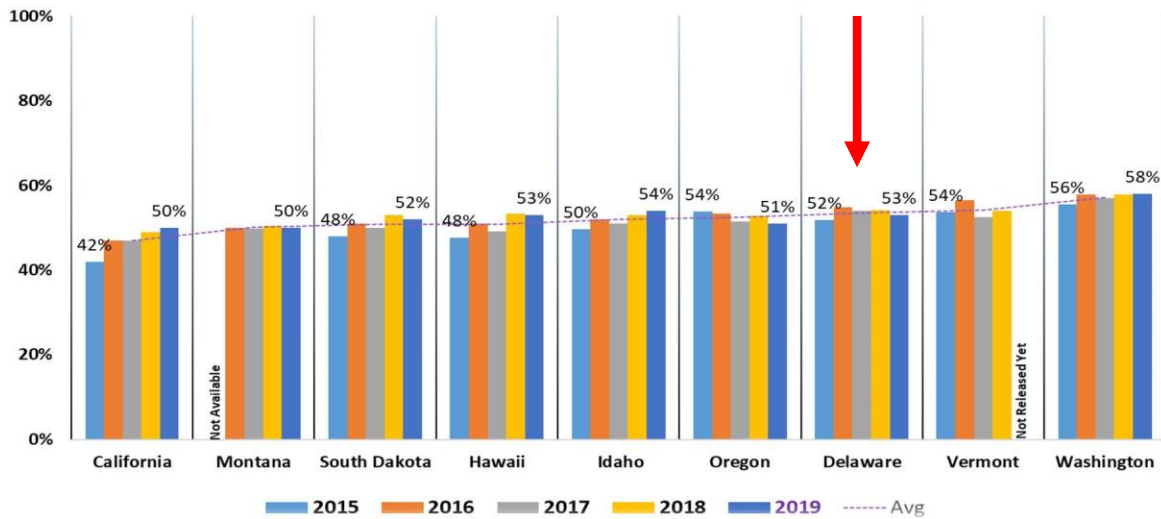


Figure 4.  
Level 3 and above Performance, Smarter Balanced Consortium, English Language Arts/Literacy, 2015-2019

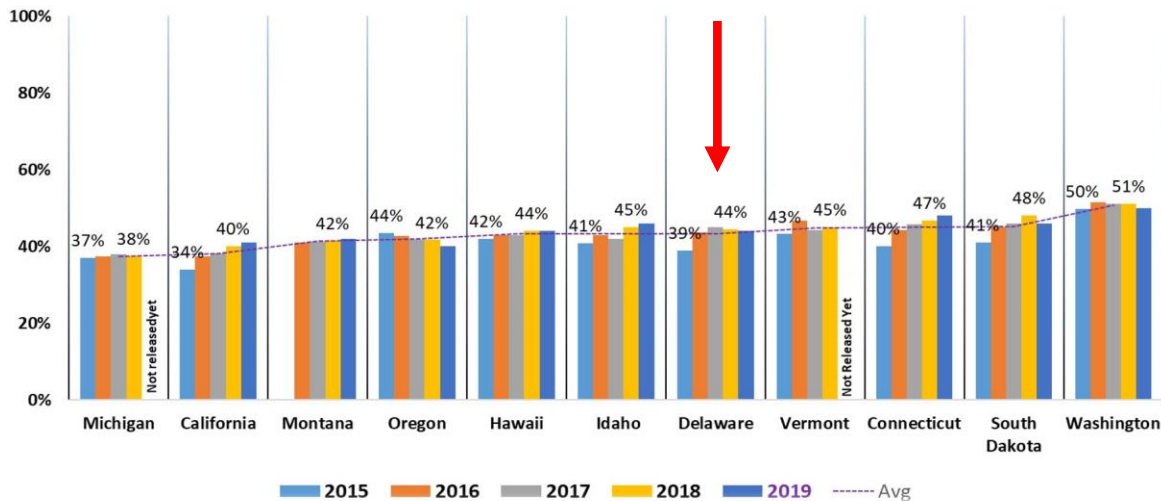


Figure 5.  
Level 3 and above Performance, Smarter Balanced Consortium, Mathematics, 2015-2019

The red arrows in Figures 4 and 5 shows Delaware’s position among SBAC member states in average performance since 2015. As can be seen in the Figures, Delaware ranks third from the top (behind Washington, Vermont) in English language arts and fifth in mathematics among SBAC member states.

### C. CONCLUSIONS

The data and documentation reviewed in Section B support the following seven conclusions:

- 1) *The report submitted by Dr. Ho substantially concurs with my initial findings regarding the quality and psychometric defensibility of the Delaware/Smarter Balanced Assessment Consortium (SBAC) statewide assessment program in general and the Delaware/SBAC performance standards in particular.*
- 2) *The reports submitted by Dr. Ho and Dr. Belfield erroneously conflate the meanings of the term adequacy as used in the Delaware Level 3 achievement level descriptions (i.e., to describe adequacy of student mastery of content) with the term as used related to the focal issue of this lawsuit (i.e., to describe adequacy of the Delaware system of public schools).*
- 3) *The report submitted by Dr. Ho mischaracterizes the intention of the Delaware/SBAC performance standards established by the standard setting participants.*
- 4) *The report submitted by Dr. Ho erroneously characterizes the Delaware/SBAC performance levels as germane to evaluating state-level educational adequacy.*
- 5) *Contrary to professional norms and guidelines, the report submitted by Dr. Ho incorrectly accepts the use of Delaware/SBAC and SAT test results for uses that they have not been developed or validated to support.*
- 6) *The report submitted by Dr. Ho fails to consider the longitudinal progress demonstrated by Delaware and the extended timeline for progress that was recognized by Delaware/SBAC standard setting participants and state-level policy makers.*
- 7) *The reports submitted by Dr. Ho and Dr. Belfield fail to contextualize Delaware's progress with respect to other states.*



---

Gregory J. Cizek

2020, May 29

Date

### Compensation

For my work on this case, I am being compensated at a rate of \$300.00 per hour. As of May 29, 2020, I have worked a total of 130 hours.



\_\_\_\_\_  
Gregory J. Cizek

\_\_\_\_ 2020, May 29 \_\_\_\_  
Date