IN THE COURT OF CHANCERY OF THE STATE OF DELAWARE

IN RE DELAWARE PUBLIC SCHOOLS LITIGATION	))))	C.A. No. 2018-0029-VCL STATE TRACK

EXPERT REPORT OF ANDREW HO, Ph.D.

Qualifications

1. I am a psychometrician, appointed as the Charles William Eliot Professor of Education at Harvard University. I have a Ph.D. in Educational Psychology and an M.S. in Statistics from Stanford University. I serve as a member of the management committee for the Journal of Educational and Behavioral Statistics, the leading journal in educational statistics. In 2012 and 2016, I was appointed by two different U.S. Secretaries of Education to the National Assessment Governing Board in the role of Testing and Measurement Expert. On that board, I have chaired the standing committee on Standards, Design, and Methodology since 2015. My curriculum vitae is attached as Exhibit A.

2. I am a member of the Boards of Directors for the National Council on Measurement in Education and the Carnegie Foundation for the Advancement of Teaching. I have also served on numerous editorial boards, including the Journal of Research on Educational Effectiveness and the Harvard Data Science Review. I serve on Technical Advisory Committees for state assessment systems in Texas, New York, and Massachusetts. I am a member of the American Educational Research Association, the National Council on Measurement in Education, and the Psychometric Society.

3. I am recognized within my profession as a leading expert on the use of test scores in educational monitoring and accountability systems.

1

4. I have not served as an expert witness previously.

5. I was asked by Plaintiffs' counsel in October 2019 to consult with them, and I have been retained by Arnold & Porter Kaye Scholer LLP (A&P) as an expert witness for this case. A statement for my work and the work of a graduate student who has assisted me in this matter is attached at Exhibit B.

Assignment

6. Counsel have asked me to offer my professional opinion regarding the meaning of aggregate test score results in Delaware. Do percentages of Level 3 and Level 4 test scores for Delaware students, including those for low-income students, students with disabilities, and English Learners (ELs), indicate percentages of students with adequate knowledge and skills for college and career success?

Summary of Conclusion

7. I conclude that the answer to these questions is yes. Substantial evidence supports a straightforward interpretation of Delaware's reported test results for percentages of students scoring at the different levels defined by the Delaware Department of Education: A Level 3 student "demonstrates adequate understanding of and ability to apply the English language and literacy (mathematics) knowledge and skills needed for success in college and career, as specified in the Common Core State Standards." Level 1 and 2 students demonstrate "minimal" and "partial" understanding and ability, respectively.[1]

---

[1] Delaware Department of Education, Delaware System of Student Assessments (DeSSA) Executive State Summary (State Summary), at 11-12 (September 2019)(Available at

8. I conclude that these percentages of students, including, disproportionately, low-income students, students with disabilities, and English Learners (ELs) scoring at Level 1 and Level 2, indicate that large percentages of Delaware students are not on track for college readiness.

9. I reached these conclusions by analyzing evidence from the Delaware System of Student Assessment (DeSSA), the Smarter Balanced Assessment Consortium (SBAC), and the College Board (CB). I also drew from my knowledge of academic publications acquired throughout my education and career, and the professional standards of my field, known as the Standards for Educational and Psychological Testing.[2] These are the consensus standards of three professional organizations, the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education.

Analysis

10. I review the bases for my conclusions in five sections. First, I confirm a necessary precondition for using test scores to define percentages of students on track to college and career readiness: that tested content is relevant to college and careers. Next, I review the process of setting the Level 3 cut score that distinguishes students who are on track from students who are not on track, in each grade and subject. I conclude that this process was conducted according to widely accepted professional standards. The resulting

---

https://www.doe.k12.de.us/cms/lib/DE01922744/Centricity/Domain/111/DeSSA% 20Executive%20State%20Summary%202019-FINAL.pdf).
[2] American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. *Standards for Education and Psychological Testing*, (July 2014).

cut score is legitimate and reasonable, with convincing evidence that reported percentages are unbiased overall and for subgroups. Finally, I review evidence that test scores are relevant to future outcomes and conclude that the current test scores can distinguish statistically between future outcomes.

1) Tested content is relevant

11. The Standards for Educational and Psychological Testing (hereafter, "the Standards") define validity as, "the degree to which evidence and theory support the interpretations of test scores"[3] (p. 11). I have been asked to evaluate the assertion that percentages of Level 3 and 4 students on the Delaware assessment tests indicate percentages of students on track for college and career readiness, and that percentages of Level 1 and Level 2 students indicate percentages not on track. The Standards list five common sources of validity evidence. The first of these is "evidence based on test content." Do test questions assess skills relevant to college and careers?

12. To evaluate test content, I reviewed documents referenced in the SBAC Technical Report[4], the Common Core State Standards in English Language

---

[3] American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. *Standards for Education and Psychological Testing*, 11 (July 2014).
[4] *See* Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium: 2017-2018 Summative Technical Report* (Available at 2017-2018-Summative-Assessment-Technical-Report.pdf).

Arts (ELA)[5] and Mathematics,[6] and the SBAC Content Specifications in ELA[7] and Math[8]. I also reviewed the test specifications for the SAT for evidence-based reading and writing, mathematics, and essay.[9] I determined from these documents that the content is relevant to college and career readiness. Evidence includes a survey of 1815 postsecondary instructors who evaluated the applicability of the Common Core State Standards to their courses and alignment studies that assess whether items measure these standards.[10]

13. As an illustration of the relevance of the tested content to postsecondary outcomes, I include in Exhibit 1 the 4 domains that the SBAC test

---

[5] Common Core State Standards Initiative, *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*, (Available at https://portal.smarterbalanced.org/library/en/english-language-artsliteracy-content-specifications.pdf).

[6] Common Core State Standards Initiative, Common Core State Standards for Mathematics (Available at https://portal.smarterbalanced.org/library/en/mathematics-content-specifications.pdf).

[7] Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium Content Specifications: English Language Arts*, (Available at https://portal.smarterbalanced.org/library/en/english-language-artsliteracy-content-specifications.pdf).

[8] Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium Content Specifications: Math*, (Available at https://portal.smarterbalanced.org/library/en/mathematics-content-specifications.pdf).

[9] *See* College Board, *Test Specifications for the Redesigned SAT*, (Available at https://collegereadiness.collegeboard.org/pdf/test-specifications-redesigned-sat-1.pdf).

[10] EPIC, *Reaching the Goal: The Applicability and Importance of the Common Core State Standards to College and Career Readiness* (2011) (Available at https://files.eric.ed.gov/fulltext/ED537872.pdf).

specifications define for each subject, Mathematics and ELA. These claims show that performance on the SBAC does not only require memorization and understanding, but application, analysis, and evaluation.

Exhibit 1: Mathematics and ELA Claims from SBAC Test Specifications[11]

> • Mathematics Claim #1: Concepts and Procedures "Students can explain and apply mathematical concepts and interpret and carry out mathematical procedures with precision and fluency."
> • Mathematics Claim #2: Problem Solving "Students can solve a range of complex well-posed problems in pure and applied mathematics, making productive use of knowledge and problem-solving strategies."
> • Mathematics Claim #3: Communicating Reasoning "Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others."
> • Mathematics Claim #4: Modeling and Data Analysis "Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems."

---

[11] Smarter Balanced Assessment Consortium, Content Specifications for the Summative Assessment of the Common Core State Standards for Mathematics, 19 (July 2015) (Available at http://www.smarterbalanced.org/wp-content/uploads/2015/08/Mathematics-Content-Specifications.pdf); Smarter Balanced Assessment Consortium, *Content Specifications for the Summative Assessment of the Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*, 29-47 (July 2015) (Available at http://www.smarterbalanced.org/wp-content/uploads/2015/08/ELA_Content_Specs.pdf)

• ELA/Literacy Claim #1: Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts.

• ELA/Literacy Claim #2: Students can produce effective and well-grounded writing for a range of purposes and audiences.

• ELA/Literacy Claim #3: Students can employ effective listening skills for a range of purposes and audiences.

• ELA/Literacy Claim #4: Students can engage in research/inquiry to investigate topics, and to analyze, integrate, and present information.

14. Chapter 1 in the SBAC Technical Report reviews dozens of studies and reports that represent accumulated evidence, both conceptual and empirical, for the relevance of the Common Core State Standards to college and career readiness, and for SBAC as an assessment that produces scores informative about student readiness.[12] These studies include, for example, the judgment of postsecondary instructors that these standards are relevant to postsecondary outcomes, the findings of trained interviewers that examinees answering items correctly were doing so with cognitively appropriate strategies, and the judgment of subject matter experts that test items are aligned to the content standards,[13]

---

[12] Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium: 2017-2018 Summative Technical Report, 11-28.*
[13] Doorey, Nancy and Polikoff, Morgan, Evaluating the Content and Quality of Next Generation Assessments, (February 2016) (Available at files.eric.ed.gov/fulltext/ED565742.pdf, https://files.eric.ed.gov/fulltext/ED537872.pdf,

15. I also reviewed test items, including Delaware practice tests for the SBAC in grades 3-8, in ELA and Mathematics, available at the Smarter Balanced website[14], and SAT practice tests[15], available at the College Board website. As noted above, others who have analyzed the alignment of SBAC to the Common Core State Standards have concluded that these items have good alignment to standards[16]. The depth of knowledge that I observed test items to require in my review of items leads me to the same conclusion. The SBAC Technical Report also reviews the test design process, demonstrating how test questions were developed in a manner aligned with content and complexity appropriate for each grade and across grades.[17] I conclude from this evidence that tested content is relevant to postsecondary readiness.

2) The achievement levels were set according to widely accepted professional assessment standards

16. In educational assessment, we distinguish between content standards and performance standards. Content standards describe what is measured. Performance standards describe levels of achievement. The process of

---

https://portal.smarterbalanced.org/library/en/cognitive-laboratories-technical-report.pdf.

[14] Smarter Balanced Assessment Consortium, Practice and Training Tests, (Available at: http://www.smarterbalanced.org/assessments/samples/).

[15] College Board, SAT Practice Tests, (Available at: https://collegereadiness.collegeboard.org/sat/practice/full-length-practice-tests).

[16] Smarter Balanced Assessment Consortium: Alignment Study Report (April 2016) (Available at: https://portal.smarterbalanced.org/library/en/smarter-balanced-assessment-consortium-alignment-study-report.pdf).

[17] Smarter Balanced Assessment Consortium: Technical Report (Available at: https://www.smarterbalanced.org/wp-content/uploads/2019/08/2017-18-Summative-Assessment-Technical-Report.pdf).

choosing "cut scores" that demarcate achievement levels is called achievement level setting or, briefly, standard setting.

17. I explained in the previous section why the Delaware assessment system measures content relevant for determining whether students demonstrate adequate understanding of and ability to apply the knowledge and skills needed for postsecondary success. Standard setting begins with achievement level descriptors and uses evidence-based judgments from teachers and subject-matter experts to select cut scores demarcating achievement levels. The SBAC developed achievement level descriptors and set cut scores with teachers and subject-matter experts, including participants from Delaware. The Delaware State Board of Education approved these results at the January 2015 State Board session[18].

18. I conclude on the basis of SBAC achievement level setting documentation[19] that the SBAC set achievement levels in an authoritative, transparent, and appropriate manner consistent with the Standards for Educational and Psychological Testing. I agree with the collected statements of support from my colleagues in the field of educational measurement, including those from an independent external auditor and two separate 10-member panels.[20] The Achievement Level Setting Advisory Panel stated in their letter of support,

---

[18] Delaware Department of Education, Delaware System of Student Assessments (DeSSA) Executive State Summary (State Summary), September 2019).

[19] *See* Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium: Achievement Level Setting Final Report (January 2015)* (Available at: https://portal.smarterbalanced.org/library/en/achievement-level-setting-final-report-with-appendix.pdf).

[20] Cizek, Gregory and Koons, Heather, *Observation and Report on Smarter Balanced Standard Setting*, (October 12-20, 2014) (Available at: https://portal.smarterbalanced.org/library/en/standard-setting-observation-and-report.pdf).

that "the design and procedures for the Achievement Level Setting and the Vertical Articulation were implemented as planned, represent a valid process that is consistent with best practices in standard setting, and support the defensibility of the content-based performance standards."[21] I agree with this assessment. The clarity of the achievement level descriptors, the representativeness of the judges, and the documentation of consensus all indicate a legitimate process that meets professional standards.

19. The technical documentation describes a well-established standard setting process. There is a clear description of a borderline Level 3 student that teachers and subject-matter experts crafted. Panelists used evidence from examinee testing responses to anchor this borderline description to the test score scale in each subject and grade. Logically, this anchored scale score location is the cut score that represents the lower border of the Level 3 achievement level. I describe this process in more detail here to emphasize its logic and legitimacy.[22]

20. Rather than defining adequacy relative to some grade level average, the standard setting process begins with an explicit "policy definition" that is broad in scope and applicable across subjects and grades. For example, a student who meets the SBAC Level 3 definition, "demonstrates adequate

---

[21] Smarter Balanced Assessment Consortium, Achievement Level Setting Advisory Panel Letter of Support) (Available at: https://portal.smarterbalanced.org/library/en/statements-of-support-achievement-level-setting-for-the-smarter-balanced-assessments.pdf https://portal.smarterbalanced.org/library/en/standard-setting-observation-and-report.pdf).

[22] *See* Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium: Achievement Level Setting Final Report* (January 2015) (Available at: https://portal.smarterbalanced.org/library/en/achievement-level-setting-final-report-with-appendix.pdf).

understanding of and ability to apply the knowledge and skills associated with college content-readiness"3.[23] Educators and subject matter experts then ground this high-level policy definition in the content learned in each subject and grade. The resulting achievement level descriptors are much more specific and connected to the claims in Exhibit 1 above. Take the Exhibit 1 claim under "Concepts and Procedures" as an example. In Grade 6 mathematics, this claim has one of many nested targets, "Statistics and Probability."[24] For this, there is a specific achievement level description that teachers and subject-matter experts crafted for Level 3 students in the 6th grade. They "should be able to pose statistical questions and understand that the responses to a statistical question have a distribution described by its center, spread, and overall shape..." (p. 388)[25]. There are numerous statements like this for each target and claim.

21. Next, because these statements describe a range, teachers and experts craft an additional description of a student who would be exactly at the lower borderline, or threshold, of Level 3. Teachers and experts consider this borderline description as they review test questions in a booklet. The items in the booklet are ordered by their known difficulty levels, based upon student responses in field trials. The question posed to these panelists was, "Would a student at the threshold have at least a 50% chance of earning this point"[26]?

---

[23] *See* Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium: Achievement Level Setting Final Report*, 753 (January 2015) (Available at: https://portal.smarterbalanced.org/library/en/achievement-level-setting-final-report-with-appendix.pdf).
[24] *See* Exhibit One.
[25] *See* Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium: Achievement Level Setting Final Report*, 388.
[26] *Id.* at 60.

Each panelist proceeds through the booklet, from easier items to harder items, until they reach an item that they believe the threshold student would no longer answer correctly. Logically, this question defines the Level 3 cut score. Because the threshold student description is linked to the boundary of the achievement level, and the question is linked empirically to the score scale, this process establishes empirically a specific Level 3 cut score for each panelist.

22. Naturally, panelist judgments differ. The SBAC technical documentation reports distributions of cut scores from panelists as they converge through three rounds of the consensus-building process. Panelists can be expected to disagree slightly about their final recommended cut scores. When the median was proposed as the consensus, the vast majority (405 out of 438, 92%) of panelists agreed with the statement, "I am confident about the defensibility and appropriateness of the final recommended cut scores." These results are consistent with an authoritative, evidence-based process that resulted in legitimate cut scores.[27]

23. After my review of well-specified achievement level descriptors, anchored via evidence from student responses to a score scale, in a judgmental process with strong expert consensus and confidence, I conclude that SBAC achievement levels meet professional standards. Delaware's active participation in the SBAC achievement level setting process and the Board's explicit adoption of SBAC achievement levels reinforces the legitimacy of these standards. Delaware's reported percentages of Level 3 and higher

---

[27] *See* Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium: Achievement Level Setting Final Report*, 753 (January, 7 2015) (Available at: https://portal.smarterbalanced.org/library/en/achievement-level-setting-final-report-with-appendix.pdf).

students who demonstrate, "adequate understanding of and ability to apply the knowledge and skills associated with college content-readiness," are authoritative.

3) The achievement levels were set at the appropriate level.

24. I have been asked by plaintiffs' counsel to respond to the assertion that the Delaware SBAC Smarter cut scores represent aspirational performance levels designed to push for greater student achievement and, as a result, the percentages of students scoring at Level 1 and Level 2 do not show that those percentages of students are inadequately educated. . That is a non sequitur. 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. The Delaware State Board of Education approved this definition. This consensus definition of adequacy has legitimacy and meaning regardless of any design for greater achievement.

25. One way in which the achievement levels might be "too high" would be if the policy definition for Level 3 were too stringent. One could recognize, as I explained in the previous section, that a Level 3 student demonstrates "adequate understanding," but argue that adequacy is not the standard we hope for students, and that instead some lower level of understanding, perhaps Level 2 (partial) or Level 1 (minimal), is sufficient. As an illustrative contrast, consider an achievement level descriptor that is considered by some to be "too high," that of the National Assessment of Educational Progress (NAEP), "Students performing at or above the NAEP Proficient level on NAEP assessments demonstrate solid academic performance and competency

over challenging subject matter."[28] I think a reasonable person could argue that "competency over challenging subject matter" is a standard that is too high. In other words, even if the description of a student scoring NAEP Proficient were accurate, someone could argue that NAEP Proficient is an unnecessarily high standard.

26. I reject that argument in the case of Delaware, because Delaware's definitions for students in grades 3-8 state that Level 3 students are those who have "adequate understanding of and ability to apply ... knowledge and skills needed." The achievement levels for Level 3 were intended to describe adequate understanding and, as described above, they were set at appropriate levels for that purpose. Inadequate understanding is inadequate.

27. NAEP mapping studies that show the relative stringency of state performance standards reveal that SBAC Level 3 standards are lower than the NAEP Proficient standards. The less rigorous SBAC Level 3 cut score is consistent with the difference between "NAEP Proficient" as defined by "competency over challenging subject matter" and "SBAC Level 3" as defined by "adequate understanding and ability."[29] 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, closer to that of NAEP Proficient.

28. The standards I have reviewed are Delaware state standards. The executive state summary described Delaware's active participation in the achievement

---

[28] See National Assessment of Educational Progress, *Understanding Assessment Results* (Available at: https://nces.ed.gov/nationsreportcard/guides/).
[29] See National Assessment of Education Progress, *Mapping State Proficiency Standards onto the NAEP Scales*, (August 2019) (Available at: https://nces.ed.gov/nationsreportcard/subject/publications/studies/pdf/2019040.pdf)

level setting process, including 48 online participants and 24 in-person participants who represented Delaware.[30] 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.

4) The test scores show no evidence of bias against selected groups.

29. A common concern with tests is that they inflate or misrepresent differences among groups rather than measure differences accurately. This could happen if tests measured irrelevant content with which higher-scoring groups had disproportionate skill or experience. If this were the case, the relatively low percentages Level 3 and above students who are, for example, Black, Hispanic, English Learners, Low Income, or students with disabilities would be even lower than they would be had a more relevant test been given. Colloquially, this is often called "test bias." It is best detected when a more relevant and representative test can be given as an audit.

30. I find no evidence for bias in this case, and I conclude that percentages of students in each achievement level are comparable across reporting categories. The strongest argument against this bias is the relevance of the content itself, as I presented in the first section. I find it difficult to argue that a more relevant and representative test exists for the purpose of measuring these subjects. Beyond content, the SBAC technical report shows that developers employ many standard procedures for ensuring fairness and accessibility, including measuring only the intended construct, establishing

---

[30] Delaware Department of Education, *Delaware System of Student Assessments (DeSSA) Executive State Summary.*

bias and sensitivity review committees, and ensuring that items and administration conditions do not offend or distract students.[31]

31. Secondary evidence exists in the form of the National Assessment of Educational Progress, which reports results from Delaware in Reading and Mathematics, in grades 4 and 8, in odd years. The NAEP results show that Delaware's achievement gaps between traditionally advantaged and disadvantaged groups are commensurate with those reported by the SBAC.[32]

32. There is also a standard method known as Differential Item Functioning (DIF) that was applied in SBAC development to assess whether individual items appeared to favor certain groups over others. Items that exhibited DIF were flagged for a content review panel to evaluate for possible exclusion. The SBAC technical manual presents item DIF statistics,[33] and there is nothing from this presentation that raises any concern about interpreting Level 3 percentages across groups as intended.

33. I also find solid evidence about the fairness of score reporting for students with disabilities and English Learners in the SBAC Technical Manual. The goal in these cases is to allow students to provide evidence of their proficiency while removing "construct-irrelevant barriers.[34]" A list of

---

[31] *See* Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium: 2017-2018 Summative Technical Report.*

[32] *See* Delaware Department of Education, 2019 NAEP Data, (October 28, 2019) (Available at: https://www.doe.k12.de.us/cms/lib/DE01922744/Centricity/Domain/442/2019%20 NAEP%20PreRelease%20presentation%20102519.pdf).

[33] Smarter Balanced Assessment Consortium, *Smarter Balanced Assessment Consortium: 2017-2018 Summative Technical Report* (Available at: https://www.smarterbalanced.org/wp-content/uploads/2019/08/2017-18-Summative-Assessment-Technical-Report.pdf).

[34] *Id.*, at 66.

appropriate accommodations was compiled from the research literature and are deployed depending on the construct or subject being measured. For example, for an English Learner, translations are deployed for mathematics, where inexperience with the English language is a construct-irrelevant barrier to measuring mathematics proficiency. However, translations are not deployed when assessing English Language Arts, because English fluency is the target of measurement. This supports the validity of interpretations of mathematics and English Language Arts scores for English Learners. The technical manual also compares indicators of measurement quality, like standard errors of measurement, across groups and finds them comparable.[35]

5) Test scores are relevant to future outcomes.

34. The Standards for Educational and Psychological Testing list the relationship between test scores and other variables as an important source of validity evidence. In the case of assessments like the SBAC and SAT that claim to measure skills relevant to college and careers, relationships between student test scores and future college and career outcomes is particularly useful. Delaware began using the SBAC in spring of 2015 and the SAT in the spring of 2016. Thus, information relating Delaware student test scores to future college and career outcomes is not yet available systematically. I review national evidence for the SAT's ability to distinguish among groups on

---

[35] *See* Smarter Balanced Assessment Consortium, *Accommodations for English Language Learners and Students with Disabilities: A Research-Based Decision Algorithm*, (February 2013) (Available at: https://portal.smarterbalanced.org/library/en/accommodations-for-english-language-learners-and-students-with-disabilities-a-research-based-decision-algorithm.pdf).

college outcomes and Delaware evidence that scores are strongly correlated from grade to grade.

35. Published reports provide evidence that the SAT predicts college outcome[36]. A 2019 College Board report estimated correlations between SAT scores and first-year college grade-point averages (FYGPA) as well as second-year college retention. These correlations indicate that SAT scores can distinguish between students with higher and lower FYGPA and higher and lower risk of second-year dropout. They also enable the empirical definition of the SAT college benchmark, for students who have "a 75% likelihood of earning a C or better in related introductory, credit-bearing college courses."[37]

36. Existing Delaware SBAC data do not enable prediction of college or career outcomes but show strong correlations from grade to grade. Using a 3-year longitudinal dataset provided by the state for academic years ending 2016-2018,[38] I find 2- and 3-year correlations of 0.8, indicating substantial stability across grades and the potential for accurate long-run predictions.

37. The argument for valid interpretation of percentages is much stronger than these student-level data suggest. At the individual level, test scores have considerable imprecision, and our professional standards have numerous cautions against basing high-stakes decisions about any individual student upon whether that student is "on track" as estimated by a single test score. However, the question is not whether every student who is Level 3 or Level 4

---

[36] Westrick, Paul et al. *Validity of the SAT for Predicting First-Year Grades and Retention to the Second Year*, College Board, (May 2019) (Available at: https://collegereadiness.collegeboard.org/pdf/national-sat-validity-study.pdf).
[37] College Board: *K-12 Educator Brief, The College and Career Readiness Benchmarks for the SAT Suite of Assessments*, (Available at: https://collegereadiness.collegeboard.org/pdf/educator-benchmark-brief.pdf).
[38] SD_0144489_HIGHLY_CONFIDENTIAL.xlsx

is on track. The question is whether the percent of Level 3 and Level 4 students an accurate description of the percent of students on track to postsecondary readiness. These percentages are much more precise. Assuming 10000 students per grade, the margin of error is half a percentage point under standard statistical assumptions.

38. I have reviewed the substantive and statistical evidence that the SBAC and SAT are relevant to college preparedness. By extension, these skills are also necessary for any career that requires a college degree. Literacy and quantitative reasoning are also essential for many careers that do not require a college degree; however, the academic skills necessary for non-college careers are more variable and often poorly measured or poorly conceptualized.[39] If there is a lack of high-quality or consensus proximal measures of readiness for careers that do not require college degrees, this should cast no aspersions on SBAC and SAT standards.

39. I conclude that, in Delaware, the weight of evidence, including a legitimate standard setting process, numerous checks against bias, and supporting longitudinal correlations, supports valid interpretation of the percentage of Level 3 and 4 students as the percentage of Delaware students on track to positive postsecondary outcomes. Therefore, the high percentages of Level 1 and 2 students and disproportionately high percentages of disadvantaged students should be seen as the educational emergency that it is: substantial proportions of Delaware students are not on track.

---

[39] *See* SRI International, *Measuring Career Readiness in High School Literature Scan* (Available at
https://ies.ed.gov/ncee/edlabs/regions/appalachia/events/materials/07-23-19-Career-Readiness-Lit-Scan-508.pdf)

ANDREW DEAN HO

Exhibit A to the Expert Report of Andrew Ho, Ph.D.

Andrew Dean Ho

455 Gutman Library

6 Appian Way

Cambridge, MA 02138

p: 617-496-2408

f: 617-496-3095

Andrew_Ho@gse.harvard.edu

POSITIONS

2018 - present	Charles William Eliot Professor of Education, Harvard Graduate School of Education
2014 - 2018	Professor of Education, Harvard Graduate School of Education
2013 - 2014	Associate Professor, Harvard Graduate School of Education
2009 - 2013	Assistant Professor, Harvard Graduate School of Education
2009 - 2010	Visiting Scholar, Stanford University School of Education
2005 - 2009	Assistant Professor, Educational Measurement and Statistics, University of Iowa

EDUCATION

| 2000 - 2005 | Stanford University | Ph.D. <u>Educational Psychology</u> (2005) M.S. <u>Statistics</u> (2003) |
| 1995 - 1999 | Brown University | Sc.B. *magna cum laude* <u>Neuroscience</u> (1999) |

ACADEMIC AWARDS, HONORS, and FELLOWSHIPS

2018 Annual Award (with Sean Reardon, Ben Shear, Erin Fahle, Demetra Kalogrides, Ken Shores, and Katherine Castellano; National Council on Measurement in Education)

2018 Frank B. Womer Lectureship in Measurement and Assessment (University of Michigan)

2013 Palmer O. Johnson Memorial Award (with Sean Reardon, American Educational Research Association)

2010 Jason Millman Promising Measurement Scholar Award (National Council on Measurement in Education)

2009 National Academy of Education/Spencer Postdoctoral Fellowship

2008 Thomas N. Urban Research Award (Iowa Academy of Education)

2007 Spencer Foundation Exemplary Dissertation Award

2007 Brenda H. Loyd Outstanding Dissertation Award (National Council on Measurement in Education)

2007 Mary Catherine Ellwein Outstanding Dissertation Award (Division D, American Educational Research Association)

2004 Spencer Dissertation Fellowship for Research Related to Education

2004 Harold Gulliksen Psychometric Research Fellowship (Educational Testing Service)

2003 Walter J. Gores Award for Excellence in Teaching (Stanford University Award)

2000 Stanford University School of Education, First Year Fellowship

EXTERNAL GRANTS

Co-Principal Investigator, with Dr. Sean Reardon. *Patterns of achievement and achievement gaps among school districts: New data, new measures, new insights.* The Spencer Foundation. #201500058. $500,000. 9/1/2014-8/31/2016.

Co-Principal Investigator, with Dr. Lori Breslow and Dr. Daniel Hastings. *Understanding the edX MOOC: How can Circuits and Electronics (6.002x) help us understand the MOOC learning experience?* National Science Foundation. DRL-1258448. Rapid Response Research Program. $200,000. 10/1/2012-9/30-2013.

Co-Principal Investigator, with Dr. Sean Reardon. *Addressing Practical Problems in Achievement Gap Estimation: Nonparametric Methods for Censored Data.* Institute of Education Sciences. R305D110018. Statistical and Research Methodology. $697,878. 4/1/2011-3/31/2013

Principal Investigator, with Dr. Stephen Dunbar. *Evaluating the Impact of the Choice of Test Score Scale on the Measurement of Individual Student Growth.* Institute of Education Sciences. R305U070008. Unsolicited Research Grant. $273,844. 9/1/2007-8/31/2010.

Principal Investigator. *Going Long: Metric-Free Representations of Test Score Trajectories over*

Multiple Time Points. The Spencer Foundation. #200800075. $25,000. 8/1/07-1/31/09.

Principal Investigator. *A Framework for Subscale-Level Comparisons of NAEP and State Testing Results Using Metric-Free Gap and Trend Statistics.* Institute of Education Sciences. R902B06007. Secondary Analysis of National Assessment of Educational Progress Data. $63,496. 5/15/06-5/14/07.

BOOKS

Castellano, K. E., & Ho, A. D. (2013). *A practitioner's guide to growth models.* Washington, DC: Council of Chief State School Officers.

JOURNAL ARTICLES (public links, sorted by topic, here)

Reardon, S. F., Ho, A. D., & Kalogrides, D. (2019). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale. *Journal of Educational and Behavioral Statistics.*

Mattern, K., Radunzel, J., Bertling, M., & Ho, A. D. (2018). How should colleges treat multiple admissions test scores? *Educational Measurement: Issues and Practice, 37*(3), 11-23

Jessen, A., Ho, A. D., Corrales, C. E., Yueh, B., & Shin, J. J. (2018). Improving measurement efficiency of the Inner EAR scale with Item Response Theory. *Otolaryngology-Head and Neck Surgery, 158,* 1093-1100.

Reardon, S. F., Shear, B. R., Castellano, K. E., & Ho, A. D. (2017). Using heteroskedastic ordered probit models to recover moments of continuous test score distributions from coarsened data. *Journal of Educational and Behavioral Statistics, 42,* 3-45.

Dede, C., Ho, A. D., & Mitros, P. (2016). Big data analysis in higher education: Promises and pitfalls. *Educause Review, 51*(5), 23-34.

Ho, A. D. (2016). Castles in the clouds: The irrelevance of vertical scales for most practical concerns. *Measurement: Interdisciplinary Research and Perspectives, 14,* 34-38.

Ho, A. D. (2016). The new (educational) statistics: Properties of scales that matter. *Journal of Educational and Behavioral Statistics, 41,* 94-99.

Northcutt, C. G., Ho, A. D., & Chuang, I. L. (2016). Detecting and preventing "multiple-account" cheating in Massive Open Online Courses. *Computers and Education, 100,* 71-80.

Reardon, S. F., & Ho, A. D. (2015). Practical issues in estimating achievement gaps from coarsened data. *Journal of Educational and Behavioral Statistics, 40,* 258-189.

Yee, D. S., & Ho, A. D. (2015). Discreteness causes bias in percent-above-cutoff comparisons: A case study from educational testing. *American Statistician, 69,* 174-181.

Miller, K., Schell, J., Ho, A. D., Lukoff, B., & Mazur, E. (2015). Response switching and self-efficacy in Peer Instruction classrooms. *Physical Review Special Topics – Physics Education Research, 11,* 010104.

Castellano, K. E., & Ho, A. D. (2015). Practical differences among aggregate-level conditional status metrics: From median student growth percentiles to value-added models. *Journal of Educational and Behavioral Statistics, 40,* 35-68.

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing "course": Reconceptualizing educational variables for Massive Open Online Courses. *Educational Researcher, 43,* 74-84.

Ho, A. D. (2014). Variety and drift in the functions and purposes of assessment in K-12 education. *Teachers College Record, 116*(110307), 1-18.

Ho, A. D., & Yu, C. C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement, 75,* 365-388.

Seaton, D. T., Nesterko, S., Mullaney, T., Reich, J., & Ho, A. D. (2014). Characterizing video use in the catalogue of MITx MOOCs. *eLearning Papers, 37,* 33-41.

Ho, A. D. (2013). The epidemiology of modern test score use: Anticipating aggregation, adjustment, and equating. *Measurement: Interdisciplinary Research and Perspectives, 11,* 64-67.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research and Practice in Assessment, 8,* 13-25.

Castellano, K. E., & Ho, A. D. (2013). Contrasting OLS and quantile regression approaches to Student "Growth" Percentiles. *Journal of Educational and Behavioral Statistics, 38,* 190-215.

Ho, A. D., & Reardon, S. F. (2012). Estimating achievement gaps from test scores reported in ordinal "proficiency" categories. *Journal of Educational and Behavioral Statistics, 37,* 489-517.

Furgol, K. E., Ho, A. D., & Zimmerman, D. L. (2010). Estimating trends from censored assessment data under No Child Left Behind. *Educational and Psychological Measurement, 70*(5), 760-776.

Ho, A. D., Lewis, D. M., & Farris, J. L. M. (2009). The dependence of growth-model results on proficiency cut scores. *Educational Measurement: Issues and Practice, 28*(4), 15-26.

Ho, A. D. (2009). A nonparametric framework for comparing trends and gaps across tests. *Journal of Educational and Behavioral Statistics, 34*, 201-228.

Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. *Educational Researcher, 37*, 351-360.

Ho, A. D. (2007). Discrepancies between score trends from NAEP and state tests: A scale-invariant perspective. *Educational Measurement: Issues and Practice, 26*(4), 11-20.

COMMISSIONED PAPERS, TECHNICAL REPORTS, CHAPTERS

Ho, A. D. (2017). Advancing educational research and student privacy in the "Big Data" era. Washington, DC: National Academy of Education.

Haertel, E. H., & Ho, A. D. (2016). Fairness using derived scores. In N. Dorans and L. Cook (Eds.), *Fairness in educational assessment and measurement* (217-237). New York, NY: Routledge.

Chuang, I., & Ho, A. D. (2016). *HarvardX and MITx: Four years of open online courses.*

Ho, A. D., Chuang, I., Reich, J., Coleman, C., Whitehill, J., Northcutt, C., Williams, J. J., Hansen, J., Lopez, G., & Petersen, R. (2015). *HarvardX and MITx: Two years of open online courses* (HarvardX Working Paper No. 10).

Ho, A. D., Reich, J., Nesterko, S., Seaton, D., Mullaney, T., Waldo, J., & Chuang, I. (2014). *HarvardX and MITx: The First Year of Open Online Courses.* (HarvardX and MITx Working Paper No. 1)

Reich, J., Nesterko, S., Seaton, D., Mullaney, T., Waldo, J., Chuang, I., & Ho, A. D. (2014). *Health in Numbers and Human Health and Global Environmental Change: 2012-2013. Harvard School of Public Health course reports* (HarvardX Working Paper No. 2)

Ho, A. D. (2014). Accuracy, transparency, and incentives: Contrasting criteria for evaluating growth models. In R. W. Lissitz and H. Jiao (Eds.), *Value added modeling and growth*

modeling with particular application to teacher and school effectiveness (61-85). Charlotte, NC: Information Age Publishing.

Ho, A. D. (2013). *Variety and drift in the functions and purposes of assessment in K-12 education*. Paper commissioned by the Gordon Commission on the Future of Assessment in Education.

Ho, A. D., & Kane, T. J. (2013). *The reliability of classroom observations by school personnel.* Bill & Melinda Gates Foundation, Seattle, WA.

Haertel, E. H., Beauregard, R., Confrey, J., Gomez, L., Gong, B., Ho, A. D., et al. (2012). *The future of NAEP.* White paper commissioned by the NCES Initiative on the Future of NAEP.

Ho, A. D. (2011). *Supporting growth interpretations using through-course assessments.* Paper commissioned by the Center for K-12 Assessment and Performance Management at ETS.

Hoffer, T. B., Hedberg, E. C., Brown, K. L., Halverson, M. L., Reid-Brossard, P., Ho, A. D., et al. (2011). *Final report on the evaluation of the Growth Model Pilot Project.* Washington, DC: U.S. Department of Education.

Erickson, V., Ho, A. D., Holtzman, D., Jaciw, A., Lukoff, B., Shen, X., Wei, X., & Haertel, E. (2007). *Closing the Gap? A Comparison of Changes Over Time in White-Black and White-Hispanic Achievement Gaps on State Assessments Versus State NAEP* (CSE Report 721). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, University of California, Los Angeles.

Ho, A. D. (2007). Growth models under NCLB: Back to basics. *NCME Newsletter, 15*(4), 5-7.

Ho, A. D., & Haertel, E. H. (2007a). *Apples to apples? The underlying assumptions of state-NAEP comparisons.* Washington, DC: Council of Chief State School Officers.

Ho, A. D., & Haertel, E. H. (2007b). *(Over)-interpreting mappings of state performance standards onto the NAEP scale.* Washington, DC: Council of Chief State School Officers.

Ho, A. D., & Haertel, E. H. (2006). *Metric-Free Measures of Test Score Trends and Gaps with Policy-Relevant Examples* (CSE Report 665). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, University of California, Los Angeles.

Ho, A. D. (2005). *Addressing discrepancies between South Carolina state test score trends on the Palmetto Achievement Challenge Tests and the National Assessment of Educational Progress.* Report provided to the South Carolina Department of Education, Columbia, South Carolina.

CONFERENCE PAPERS AND PRESENTATIONS

Bertling, M., & Ho, A. D. (2019, April 7). Demographics and differential advangates in college admissions retesting. Paper presented at the annual meeting of the National Council on Measurement in Education. Toronto, Canada.

Ho, A. D. (2019, April 6). Big data: Analytics, Privacy, and implications for education research. Invited Dean's Salon. George Washington Graduate School of Education and Human Development. Toronto, Canada.

Thng, Y. T., & Ho, A. D. (2019, April 6). A critique of predictive methods in evidence-based standard setting. Poster presented at the annual meeting of the National Council on Measurement in Education. Toronto, Canada.

Ho, A. D., Shear, B., Reardon, S. F., & Fahle, E. (2018, April 16). Rise and convergence of state proficiency standards in the common core era. Paper presented at the annual meeting of the National Council on Measurement in Education. New York, NY.

Shear, B., Ho, A. D., Reardon, S. F., & Fahle, E. (2018, April 16). Mapping proficiency standards using heteroskedastic ordered probit models and NAEP-based linear linking. Paper presented at the National Council of Measurement in Education. New York, NY.

Yee, D., & Ho, A. D. (2018, April 16). The sensitivity of test-based school growth metrics to transformations of scale. Paper presented at the National Council on Measurement in Education. New York, NY.

Ho, A. D., Reardon, S. F., & Kalogrides, D. (2017, April 28). *Linking U.S. school test score distributions to a common scale.* Paper presented at the annual meeting of the National Council on Measurement in Education. San Antonio, TX.

Quinn, D., & Ho, A. D. (2017, April 29). *An ordinal approach to decomposing test score gaps.* Paper presented at the annual meeting of the National Council on Measurement in Education. San Antonio, TX.

Bertling, M., & Ho, A. D. (2017, April 30). *Demographics and differential advantages in college admissions retesting.* Paper presented at the annual meeting of the National Council on Measurement in Education. San Antonio, TX.

Quinn, D., & Ho, A. D. (2017, March 2). *An ordinal approach to decomposing test score gaps.* Paper presented at the annual meeting of the Society for Research on Educational Effectiveness. Washington, DC.

Jackson, C., & Ho, A. D. (2016, April 12). *Refining and raising the reliability of classroom observations.* Paper presented at the annual meeting of the American Educational Research Association. Washington, DC.

Wong, T., Lichtenstein, D., Whitehill, J., Lopez, G., Chuang, I., & Ho A. D. (2016, April 8). *The impact of instructor interaction in Massive Open Online Courses.* Paper presented at the annual meeting of the American Educational Research Association. Washington, DC.

Reardon, S. F., Kalogrides, D., & Ho, A. D. (2016, March 4). *Mapping U.S. school district test score distributions onto a common scale, 2009-2013.* Paper presented at the spring meeting of the Society for Research on Educational Effectiveness. Washington, DC.

Castellano, K. E., & Ho, A. D. (2015, April 18). *The relevance of normality assumptions in ordered probit models.* Paper presented at the annual meeting of the National Council on Measurement in Education. Chicago, IL.

Ho, A. D. (2014, April 5). *Between the ordinal and the interval: Pliable scales, plausible transformations.* Paper presented at the annual meeting of the National Council on Measurement in Education, Philadelphia, PA

Shear, B., Castellano, K., Reardon, S. F., & Ho, A. D. (2014, March 7). *Simultaneous estimation of multiple achievement gaps from ordinal proficiency data.* Paper presented at the spring meeting of the Society for Research on Educational Effectiveness. Washington, DC.

Reardon, S. F., Ho, A. D., & Kalogrides, D. (2013, April 29). *Addressing measurement error and sampling variability in nonparametric gap estimation.* Paper presented at the annual meeting of the National Council on Measurement in Education. San Francisco, CA.

Ho, A. D., & Reardon, S. F. (2013, April 29). *Practical achievement gap estimation in a nonparametric framework.* Paper presented at the annual meeting of the National Council on Measurement in Education. San Francisco, CA.

Reardon, S. F., Ho, A. D., & Kalogrides, D. (2013, April 30). *Addressing sampling variability in nonparametric gap estimation.* Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

Ho, A. D., & Reardon, S. F. (2013, April 30). *Practical achievement gap estimation in the NCLB era.* Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

Ho, A. D. (2013, April 30). *Off track: Problems with "on track" inferences in empirical and predictive standard setting.* Paper presented at the annual meeting of the National Council on Measurement in Education. San Francisco, CA.

Castellàno, K. E., & Ho, A. D. (2013, April 27). *A practitioner's guide to growth models.* Preconference Workshop. Annual meeting of National Council on Measurement in Education. San Francisco, CA.

Ho, A. D., & Kane, T. J. (2013, March 15). *The reliability of classroom observations by school personnel.* Annual conference of the Association for Education Finance and Policy. New Orleans, Louisiana

Reardon, S. F., & Ho, A. D. (2013, March 8). *Estimating achievement gaps from NCLB-type categories.* Invited Symposium. Society for Research on Educational Effectiveness. Washington, DC.

Ho, A. D. (2011). *Lord's paradox and the use of growth and value-added models for school accountability.* Paper presented at the 2011 International Meeting of the Psychometric Society, Hong Kong, China.

Ho, A. D. (2011). *A psychometrician's perspective on consortia approaches to measuring teacher effectiveness.* Paper presented at the 2011 meeting of the American Educational Research Association, New Orleans, LA.

Ho, A. D., & Reardon, S. F. (2011). *Estimating achievement gaps from test scores reported in ordinal "proficiency" categories.* Paper presented at the 2011 meeting of the American Educational Research Association, New Orleans, LA.

Ho, A. D. (2010). *Beyond the bubble: Disproportionate gains at the "proficiency" cut score across state tests and NAEP.* Paper presented at the 2010 meeting of the American Educational Research Association, Denver, CO.

Ho, A. D. (2009). *Graphical methods for comparing trends, gaps, and gap trends across tests.* Paper presented at the 2009 meeting of the National Council on Measurement in Education, San Diego, CA.

Ho, A. D. (2009). *A three-level hierarchy of perspectives on the vertical-scale-sensitivity of growth models.* Paper presented at the 2009 meeting of the American Educational Research Assocation, San Diego, CA.

Ho, A. D. (2008). *Using parametric assumptions to frame trends in categorical proficiency data under the U.S. No Child Left Behind Act.* Presented at the 2008 International Meeting of the Psychometric Society, Dover, New Hampshire.

Ho, A. D., & Magda, T. R. (2008). *The dependency of growth models on proficiency standards.* Paper presented at the 2008 annual meeting of the National Council on Measurement in Education, New York, New York, and the National Conference on Student Assessment, Orlando, Florida.

Furgol, K. E., & Ho, A. D. (2008). *Is NCLB inflating trends at proficiency? A nonparametric, cross-state analysis of trends across cut scores.* Poster presented at the 2008 annual meeting of the National Council on Measurement in Education, New York, New York

Magda, T. R., & Ho, A. D. (2008). *Comparing trends at different cut scores: Distribution-wide NCLB and NAEP comparisons from 2003 to 2007.* Paper presented at the 2008 annual meeting of the American Educational Research Association, New York, New York.

Thiessen, B. A., Magda, T. R., & Ho, A. D. (2008). *Nonparametric comparisons of high-stakes and low-stakes trends and gaps: 2003-2007.* Paper presented at the 2008 annual meeting of the American Educational Research Association, New York, New York.

Ho, A. D. (2007). *Probability-Probability (PP)-based trend statistics for more than two time points.* Paper presented at the 2007 International Meeting of the Psychometric Society, Tokyo, Japan.

Ho, A. D. (2007). *State-NAEP standard mappings: Cautions and alternatives.* Invited presentation at the National Conference on Large-Scale Assessment, Council of Chief State School Officers, Nashville, Tennessee.

Ho, A. D. (2007). *Describing the pliability of growth statistics under transformations of the vertical scale.* Paper presented at the 2007 annual meeting of the National Council on Measurement in Education, Chicago, Illinois.

Magda, T. R., & Ho, A. D. (2007). *A distribution-wide view of high-stakes test score trends*. Paper presented at the 2007 annual meeting of the American Educational Research Association, Chicago, Illinois.

Ho, A. D. (2006). *Describing the pliability of effect-size based trend and gap statistics under monotone transformations*. Paper presented at the 2006 International Meeting of the Psychometric Society, Montreal, Canada.

Ho, A. D. (2006). *Discrepancies between score trends on NAEP and state tests: An analysis using metric-free statistics*. Paper presented at the 2006 annual meeting of the American Educational Research Association, San Francisco, California.

Lukoff, B., Wei, X., Shen, X., Ho, A. D., & Haertel, E. H. (2006). *Using test content to address trend discrepancies between NAEP and California state tests*. Paper presented at the 2006 annual meeting of the American Educational Research Association, San Francisco, California.

Ho, A. D., & Haertel, E. H. (2005). *Metric-free methods for evaluating and interpreting gaps between test-score distributions*. Paper presented at the 2005 annual meeting of the American Educational Research Association, Montreal, Canada.

Ho, A. D., & DiBello, L. V. (2004). *Policy applications of the fusion model for skills diagnosis*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, California.

Ho, A. D., & Haertel, E. H. (2003). *Criterion-referenced testing: An idea in need of a theory*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, Illinois.

BOARDS, PANELS, COMMISSIONS, and SERVICE

National Council on Measurement in Education (2018-present). Board member

Massachusetts Technical Advisory Committee (2018-present). Committee member

National Assessment Governing Board (2012-present). Board member.
 Committee on Standards, Design, and Methodology. Chair (2015-present).

Carnegie Foundation for the Advancement of Teaching (2017-present). Board member.

Spencer Foundation (2017-2019). Lyle Award review panel.

ACT (2015-present). Technical advisory committee member.

Curriculum Associates (2014-present). Technical advisory committee member.

New York Technical Advisory Committee (2012-present). Committee Member

Texas Technical Advisory Committee (2011-present). Committee Member.

Research Committee, Harvard University Vice Provost for Advances in Learning (2015-present). Chair.

Harvard University Presidential Task Force on Inclusion and Belonging (2016-2018). Subcommittee chair for Demographic Realities.

HarvardX Research Committee (2012-2015). Chair.

Gordon Commission on the Future of Assessment in Education (2011-2013). Commissioner.

Institute of Education Sciences Review Panel (2012)
Statistics and Modeling Review Panel Member.

NCES Initiative on the Future of NAEP (2012). Panel Member.

National Council on Measurement in Education. Annual Meeting Program Co-Chair (2015-2016). Newsletter Advisory Committee (2014-present). Training Session Presenter: Tips for Graduate Students (2006-2012). Brenda H. Loyd Outstanding Dissertation Award Committee (2011-2012)

College Board Pathway Linking Advisory Team (2011-2012)
Advisory Team Member.

Assessing Teaching Collaboratory, Carnegie Foundation for the Advancement of Teaching (2011). Collaborator.

Aspen Institute (2011, April)
Faculty expert at the Aspen Senior Congressional Education Staff Network, on Standards and Assessment: Implications of the Common Core and Assessments

Experts Bringing Evidence to Practitioners (2009, 2010)
IES-funded REL initiative for four talks, in Chicago, Anchorage, Seattle, and Eugene, entitled "Interpreting Test Score Trends and Gaps."

American Educational Research Association. Minority Fellowship Advisory Committee (2006-2009).

National Institute of Statistical Sciences (2008)
  NISS/NESSI Task Force on Full Population Estimates for NAEP (Member)

Reviewer: *AERA Open (Associate Editor), Journal of Research on Educational Effectiveness, Educational Researcher, Journal of Educational and Behavioral Sciences, Educational Measurement: Issues and Practice, Educational Assessment, Science, American Educational Research Journal. National Center for Education Statistics.*

VARIOUS SPEAKING ENGAGEMENTS (2015-2019)

Ho, A. D. (2019, June 26). *How can NCME better serve state assessment professionals?* Panelist. National Conference on Student Assessment. Council of Chief State School Officers. Orlando, FL.

Ho, A. D. (2019, June 14). *The use of test scores in secondary analysis.* Discussant. PIAAC Methodological Seminar. Organisation for Economic Co-operation and Development. Paris, France.

Ho, A. D. (2019, April 17). *Practice-driven data in Chicago.* Reflector. Carnegie Summit. Carnegie Foundation for the Advancement of Teaching. San Francisco, CA.

Ho, A. D. (2019, April 7). *Testing, testing: Retesting and inequality in large-scale college admissions tests.* Chair. Annual meeting of the National Council on Measurement in Education. Toronto, Canada.

Ho, A. D. (2019, April 6). *Practical measurement for improvement science: Principles and applications.* Chair. Annual meeting of the National Council on Measurement in Education. Toronto, Canada.

Reardon, S. F., Ho, A. D., Fahle, E., & Shear, B. (2019, March 6). *Using school-level data from the Stanford Education Data Archive.* Short course. Spring meeting of the Society for Research on Educational Effectiveness. Washington, DC.

Ho, A. D. (2018, August 27). *Big data: The distribution of academic achievement in the United States.* 8x8 Orientation Session. Harvard Graduate School of Education, Cambridge, MA.

Ho, A. D. (2018, June 12). *Statistical and psychometric methods for measurement: Scale development and validation.* Short course. World Bank. Washington, DC.

Ho, A. D. (2018, May 11). *Linking methods for comparing group means and proficiency cut-scores across tests.* Invited presentation. Workshop on Quantitative Methods in

Education, Health, and the Social Sciences. University of Chicago. Chicago, IL.

Ho, A. D. (2018, May 11). *A very short introduction to educational measurement.* Short course. University of Chicago Consortium on School Research. Chicago, IL.

Ho, A. D. (2018, April 30). *Linking 1 million district test score distributions to a common scale: Psychometric methods underlying SEDA.* Invited presentation. Institute of Education Sciences – Predoctoral Interdisciplinary Research Training. New York University. New York, NY.

Ho, A. D. (2018, April 15). Measuring proficiency vs. growth in education; measuring quality vs. improvement in health. Invited session. Annual meeting of the American Educational Research Association. New York, NY.

Reardon, S. F., & Ho, A. D. (2018, April 15). *The Stanford Educational Data Archive (SEDA): Using measurement methods to make public test score data useful.* Invited awards presentation. Annual meeting of the National Council on Measurement in Education. New York, NY.

Ho, A. D. (2018, April 14). Measuring collaboration and engagement using "Big Data": Comments. Discussant. Annual meeting of the National Council on Measurement in Education. New York, NY.

Reardon, S. F., Ho, A. D., Fahle, E., & Shear, B. (2017, April 30). *Stanford Education Data Archive: Training session.* Short course. Annual meeting of the National Council on Measurement in Education. New York, NY.

Ho, A. D. (2018, February 14). *The role of measurement in the "big data" era.* Frank B. Womer Lecture in Measurement and Assessment. University of Michigan. Ann Arbor, MI.

Ho, A. D. (2018, February 14). *Made to be broken: The paradox of student growth prediction.* Frank B. Womer Invitational Keynote. Michigan School Testing Conference. Ann Arbor, MI.

Ho, A. D. (2017, August 25). *The Reach of MOOCs.* Podcast. The Crush, with Davin Sweeney.

Ho, A. D. (2017, August 23). *Big data: The distribution of academic achievement in the United States.* 8x8 Orientation Session. Harvard Graduate School of Education, Cambridge, MA.

Ho, A. D. (2017, June 29). *Measure what matters: Options for high school assessment and accountability.* Invited panelist. National Conference on Student Assessment. Council of Chief State School Officers. Austin, TX.

Ho, A. D. (2017, June 13). *Of telescopes and microscopes: Big data in education.* Invited address.

Curriculum Associates. Billerica, MA.

Ho, A. D. (2017, May 12). *Big data as public good: An example from education.* Keynote address. Delaware Data Science Symposium. University of Delaware. Newark, DE.

Reardon, S. F., Ho, A. D., Fahle, E., & Shear, B. (2017, April 30). *The Stanford Education Data Archive: Using big data to study academic performance.* Short course. Annual meeting of the American Educational Research Association. San Antonio, TX.

Ho, A. D. (2017, April 28). *Advancing educatinal research and student privacy in the "big data" era.* Presidential Session – Using Big Data: The ethics, dilemmas, and possibilities for educational opportunity. Annual meeting of the American Educational Research Association. San Antonio, TX.

Ho, A. D. (2017, April 27). *Advancing your career in measurement or research methodology: Finding the real meaning of growth, path analysis and curves.* Panelist. Division D Graduate Student Seminar. Annual meeting of the American Educational Research Association. San Antonio, TX.

Ho, A. D. (2017, March 13). *A cross-disciplinary look: How do you measure quality?* Panelist. High Quality Health Systems. Harvard T.H. Chan School of Public Health. Boston, MA.

Reardon, S. F., Ho, A. D., Fahle, E., & Shear, B. (2017, March 1). *The Stanford Education Data Archive: Using big data to study academic performance.* Short course. Society for Research on Educational Effectiveness. Washington, DC.

Ho, A. D. (2017, February 6). *Linking U.S. school district test score distributions to a common scale.* Colloquium speaker. Education Policy and Program Evaluation program. Harvard Graduate School of Education. Cambridge, MA.

Reardon, S. F., & Ho, A. D., Fahle, E., & Shear, B. (2017, Jaunary 27). *The Stanford Education Data Archive: Using big data to study academic performance.* Short course. University of California, Los Angeles, CA.

Reardon, S. F., & Ho, A. D. (2017, January 26). *200 million test scores and what do we know?* Featured speakers. Spencer Foundation. Board Meeting. Los Angeles, CA.

Ho, A. D. (2016, December 7). *Measurement: 7 key statistical and psychometric principles for policy researchers.* Causal Inference in Education Policy Research – Short Course. University of Michigan. Ann Arbor, MI.

Ho, A. D. (2016, December 7). *Linking U.S. school district test score distributions to a common scale.* Causal Inference in Education Policy Research Seminar. University of Michigan. Ann Arbor, MI.

Ho, A. D. (2016, December 5). *Pass/Fail: How test-based accountability stacks up.* Debate moderator. Harvard Graduate School of Education: Askwith debates. Cambridge, MA.

Ho, A. D. (2016, November 30). *The 12-hour degree? New measures of student engagement in online and residential courses.* Featured speaker. Harvard Graduate School of Education: Alumni Event. New York, NY.

Ho, A. D. (2016, November 10). *The 12-hour degree? Open online courses and asynchronous learning.* Keynote address. International Baccalaureate: Higher Education Symposium. Medford, MA.

Ho, A. D. (2016, August 9). *Advancing educational research and student privacy in the "big data" era.* Opening address. Workshop on big data in education: Balancing research needs and student privacy. National Academy of Education. Washington, DC.

Ho, A. D. (2016, July 2). *The mismeasurement of MOOCs: Lessons from Massive Open Online Courses about metrics for online learning.* Keynote address. Annual meeting of the International Test Commission. Vancouver, Canada.

Ho, A. D. (2015, June 2). *MOOCs.* Panelist. Advancing data-intensive research in education. National Science Foundation. Arlington, VA.

Ho, A. D. (2015, April 19). *A major assessment literacy campaign aimed at parents, policy makers, and students.* Panelist. Annual meeting of the American Educational Research Association. Chicago, IL.

Ho, A. D. (2015, April 18). *Technical investigation of SGPs/VAMs for teacher evaluation.* Discussant. Annual meeting of the National Council on Measurement in Education. Chicago, IL.

Ho, A. D. (2015, April 18). *MOOC hype vs. MOOC research: The role of cross-institutional collaborations in advancing science and equity.* Panelist. Annual meeting of the American Educational Research Association. Chicago, IL.

Ho, A. D. (2015, April 16). *Massive Open Online Course students in context: Equity, motivation, and performance.* Coordinated session discussant. Annual meeting of the American Educational Research Association. Chicago, IL.

Castellano, K. E., & Ho, A. D. (2015, April 15). *A practitioner's guide to growth models.* Workshop. Annual meeting of the National Council on Measurement in Education. Chicago, IL.

Ho, A. D. (2015, April 8). *Translating learning research into practice.* Panel moderator. Online

learning: Shaping the future of higher education on and off campus. Cambridge, MA.

Ho, A. D., & Castellano, K. E. (2015, March 4). *A practitioner's guide to growth models.* Workshop. Annual meeting of the Society for Research on Educational Effectiveness. Washington, DC.

TEACHING

Harvard Graduate School of Education, Cambridge, MA. (Fall, 2010 through Fall, 2019)
 S-061 – Statistical and psychometric methods for educational measurement
   (Fall, 2010-2013, 2015-2019)
 S-052 – Intermediate and Advanced Statistical Methods for Applied Educational
   Research (Spring, 2018-2020)
 S-052 – Applied Data Analysis (Spring, 2013, 2015-2017)
 S-030 – Applied Regression and Data Analysis (Spring 2011, 2012)

University of Iowa, Iowa City, IA. (Spring, 2006 through Spring, 2009)
 Introductory and Intermediate Statistics, Factor Analysis and Structural Equation Models

Stanford University, Stanford, CA. (Autumn, 2002 through Winter, 2003)
 Introductory Statistics and Regression Analysis (Assistant)

The Thacher School, Ojai, CA. (Autumn, 1999 to Spring, 2000)
 AP and Freshman Physics, Advising.

Punahou School, Honolulu, HI. (Summers, 1999-2001)
 Creative Writing, Physics.

Exhibit B to the Expert Report of Andrew Ho, Ph.D.

Fee for preparation of the Expert Report of Andrew Ho, Ph.D.

Andrew Ho, Ph.D.

    40 hours at $300 per hour

Graduate Student Research Assistant

    40 hours at $80 per hour