

Summary of ESSA Assessment and Accountability Requirements

Brian Gong



Assessment Requirements - same

- ⌘ Annual assessments in Reading or ELA, mathematics in grades 3-8 and at least once in high school, of state's CCR content standards; science at least once in elementary, middle, and high school grade spans
- ⌘ Assessment of English language proficiency for ELs
- ⌘ At least three achievement levels
- ⌘ All students, including AA-AAS, and ELP

Assessment Requirements – New

- ⌘ High school assessments’ “college ready” must be aligned with state’s IHE
- ⌘ CAT may include out-of-grade items
- ⌘ State must consider request of districts to use a “locally selected, nationally recognized” high school assessment in lieu of state assessment
- ⌘ State may consider using multiple interim assessments to yield a single summative student score instead of a single summative assessment
- ⌘ States may apply for “Innovative Assessment and Accountability” Demonstration Authority project
- ⌘ States encouraged to use performance, portfolio, simulations, and other complex assessments of higher-order content standards (See 8/16 EAG announcement)

Accountability Requirements - same

⌘ Structurally the same as NCLB:

- State sets performance goals
- State measures performance largely based on student assessments
- State evaluates performance in relation to performance goals
- State labels schools on basis of performance evaluation
- State mandates consequences (support and intervention) for low-performing schools

Accountability Requirements – new

⌘ Biggest differences:

- Incorporates EL performance on ELP assessment into Title I accountability
- State charged to set “challenging but realistic” performance goals, but not prescribed how to do so; subject to review
- State told how to identify low-performing schools for Comprehensive and Targeted support and intervention—different than NCLB
- State charged to determine what support and intervention should consist of—no NCLB “menu”; some timing requirements
- State and/or LEAs charged with determining exit criteria for schools leaving Comprehensive/Targeted Support

Accountability Indicator Reporting

⌘ Five indicators

- Academic Achievement
- Other Academic Indicator (e.g., Progress/Growth)
- Graduation Rate
- Progress towards achieving English language proficiency
- School Quality/Student Success

⌘ State must set at least three levels of performance on each indicator, consistent with State's long-term goals and measurement of interim progress

⌘ State must define how to combine measures to produce indicator result (e.g., combine ELA & math performance across grades (probably not subgroups?) to produce Academic Achievement indicator for school)

⌘ State reports by indicator for school and by subgroup

Accountability Summative Rating Reporting

- ⌘ State must report annually an accountability result (i.e., “Summative Rating”) for each school that “meaningfully differentiates” between schools
- ⌘ “Summative Rating” is based on performance on accountability indicators in relation to State’s “measurement of interim progress” goals
- ⌘ State must define how to combine performance on indicators to produce Summative Rating
- ⌘ State must set at least three categories/levels for the single “Summative Rating”
- ⌘ Summative Rating is a school rating; disaggregated reports *not* required; but need disaggregated later

ESSA Requirements for Subgroups (from 5/31/16 draft regulations)

- ⌘ **Assessment Reporting** – report assessment results annually, disaggregated by federal accountability subgroups (racial/ethnic, economically disadvantaged, students with disabilities, English learners) *and* federal reporting subgroups (gender, children in foster care, homeless, children with military duty parent)
- ⌘ **Accountability Indicator Reporting** – report performance on accountability indicators annually, disaggregated by federal accountability subgroups
- ⌘ **Accountability Summative Rating Reporting** – report school’s overall accountability determination annually, based on accountability indicators; no disaggregation required for reporting; subgroup required for Targeted ident.?
- ⌘ **Targeted Support Identification** – identify schools annually that qualify for targeted support—based on subgroup performance
- ⌘ **Comprehensive Support Identification** – identify schools at least every three years that qualify for comprehensive support—subgroups one pathway
- ⌘ **Exit Criteria and Beyond Comprehensive Support Identification** – identify schools that did not meet performance criteria after comprehensive support—subgroups not mentioned

Subgroups in ESSA

Accountability Architecture

⌘ How subgroups function in ESSA accountability is somewhat different than in NCLB or in ESEA Waivers

	NCLB	ESEA Waivers	ESSA
How subgroup evaluated	Compared to AMOs	Compared to AMOs/targets	Compared to 5%, to AMOs, to other
Relation to school rating	Incorporated into school rating (INOI)	Optional to incorporate into school rating	Optional to/do not incorporate into school rating
Relation to consequences	Triggered consequences from rating	Optional use to help identify bottom 5%	Triggers consequences; and optional uses

Targeted Support Identification

⌘ State identifies for Targeted support and improvement all schools where either:

- School has at least one “**low-performing subgroup**,” defined as a federally defined subgroup performing at a level at or below the summative performance of all students in any of the lowest-performing 5% of Title I schools in comprehensive support. Every 3 years starting 2017-18.
 - For example: Find the “All Students” subgroup in the 5% schools with the highest Summative Performance score. Compare that performance to the Summative Performance score of every subgroup in every school. If school’s subgroup performance is lower, then school must implement targeted support for that subgroup.
- School has at least one “**consistently underperforming subgroup**,” where the state defines the subgroup and “consistently underperforming” based on no more than two years’ data. Identified annually starting 2018-19.
 - For example: Could be “super subgroup,” “lowest performing students”; low performance could be relative to an indicator, to a norm group, low participation criterion, etc.

Comprehensive Support Identification

- ⌘ State identifies for comprehensive support and improvement at least every three years starting 2017-18:
- The lowest-performing 5% of all Title I schools in the state, considering the annual Summative Rating data over 3 years;
 - Any public high school in the state failing to graduate one-third or more of its students (based on 4-year adjusted cohort graduation rate among all students);
 - School with at least one “**chronically low-performing subgroup**,” defined as a subgroup performing as poorly as “All Students” in the lowest-performing 5% of Title I schools and that has failed to improve after implementation of targeted support over no more than three years starting 2019-20.

Exit Criteria and Beyond Comprehensive Support Identification

- ⌘ State must set exit criteria for schools to qualify to leave targeted, comprehensive, and beyond-comprehensive support.
- ⌘ If a school in comprehensive support does not meet the exit criteria within a state-determined time period (not to exceed four years), the state must specify more rigorous state-determined action in the school, such as the implementation of interventions (which may address school-level operations).

Targeted Support Detail – 1

- ⌘ **Targeted Support:** School with a subgroup performance lower than All Students in schools in Bottom 5%
- ⌘ Identify schools in lowest 5%
- ⌘ Find school with the highest performing All Students subgroup on the Summative Performance rating?
- ⌘ That school's All Student performance becomes the "5% subgroup criterion"
- ⌘ Identify any school with a subgroup whose Summative Performance is lower than the "5% subgroup criterion" (Up to 8 subgroups per school) (This implies that subgroup performance is not used to calculate Summative Performance)

Targeted Support Detail – 2

⌘ State must develop two ways to use subgroups to identify schools/subgroups for Targeted Support

- “Low-Performing” = Lower than “All Students” in 5% Schools
- “Consistently Low-Performing”
 - State defines subgroups, “consistently low-performing,” data (no more than two years of data)
 - Subgroup could be “super subgroup,” “lower performing students,” one student-partial weight across member-subgroups, etc.

⌘ Conjunctive

⌘ Direct identification for Support (regardless of school’s Summative Rating)

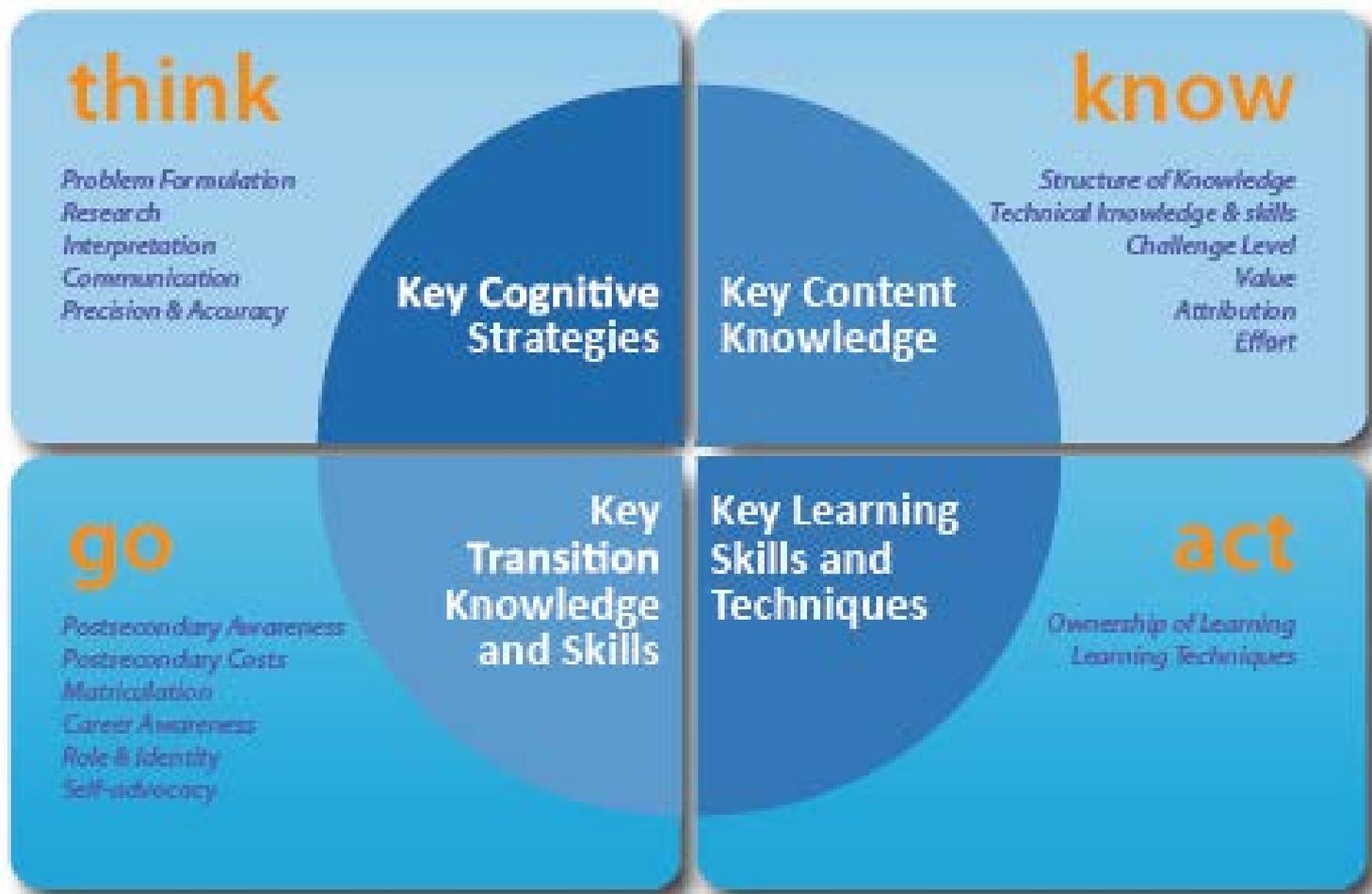
Some Key Technical Subgroup Accountability Design Decisions

- ⌘ Much is known about the factors that affect the technical quality of accountability systems' ratings, including subgroup accountability.
- ⌘ What will your target balance be between validity and reliability? In particular, how much error are you willing to make on precision when the trade-off is less inclusion? Why?
- ⌘ What Type I/Type II error will you build into your system? Is it better, from a value perspective, to err on the side of falsely identifying a school as having a low-performing-subgroup when in fact the school does not, or to say the school is fine when in fact it does have a low-performing subgroup? Do your tolerances for Type I/Type II error differ by accountability consequences or at different points in the implementation cycle of your accountability system?
- ⌘ What is the relative value you place on a subgroup compared with the performance of other subgroups or the group as a whole? Does the value differ for different accountability decisions?
- ⌘ What should be the “good enough” performance criteria for different individuals/groups? How should those criteria be “the same” and “different”?
- ⌘ What will your balance be between simplicity and validity/reliability? For example, the accountability systems that are simplest and easiest to communicate are probably less valid and less reliable than some more complex alternatives.

Some Key ESSA Decisions and Definitions for States

1. Define federal racial/ethnic subgroups for accountability
2. Define values and approaches for safeguarding validity/reliability/credibility
3. Define minimum-n for subgroups for accountability—no greater than 30, except may be up to 100 for high school graduation rate (30 may not be enough)
4. Define state accountability for indicators—that will influence identification of “low-performing” and “chronically low-performing” subgroups
5. For “consistently underperforming” subgroup determinations
 - 5.1 Define subgroup
 - 5.2 Define “consistently underperforming”
 - 5.3 Define how many years of data to consider (up to 2 years)
 - 5.4 Define how to combine data across years (if more than 1 year)
6. For “comprehensive support” determinations
 - 6.1 Define how many years of data to consider (up to 3 years)
 - 6.2 Define how to combine data across years
7. Define exit criteria (performance, time allowed) from Targeted and from Comprehensive support
8. Define what state interventions are after Comprehensive support

Defining College and Career Readiness (Conley)



Defining CCR: Setting performance standards for a CEE

- ⌘ What are the definitions of each performance level?
- ⌘ Accept the performance level descriptors of the publisher or use state-specific descriptors?
- ⌘ Coherence of performance levels descriptors, cutpoints, and impact between high school and grades 3-8
- ⌘ Coherence between assessment performances and other state systems, e.g., accountability, support
- ⌘ Communication
- ⌘ These should all inform the standard-setting plans

Definitions of performance on CEE

⌘ Publisher's benchmark definition

⌘ Readiness = entry

⌘ Success

- Retention in college, second semester/second year
- Grades in associated entry-level credit-bearing course
 - ACT: "75% probability of C or higher; 50% probability of B or higher in courses X, Y, Z"
 - SAT: "75% probability of C or higher in courses A, B, C"
- Overall first semester/year GPA

⌘ For which reference groups/institutions

- Publisher's national reference, average
- State's selective, less-selective 4-year, 2-year IHE
- Career-ready reference groups

Defining high school cutscores

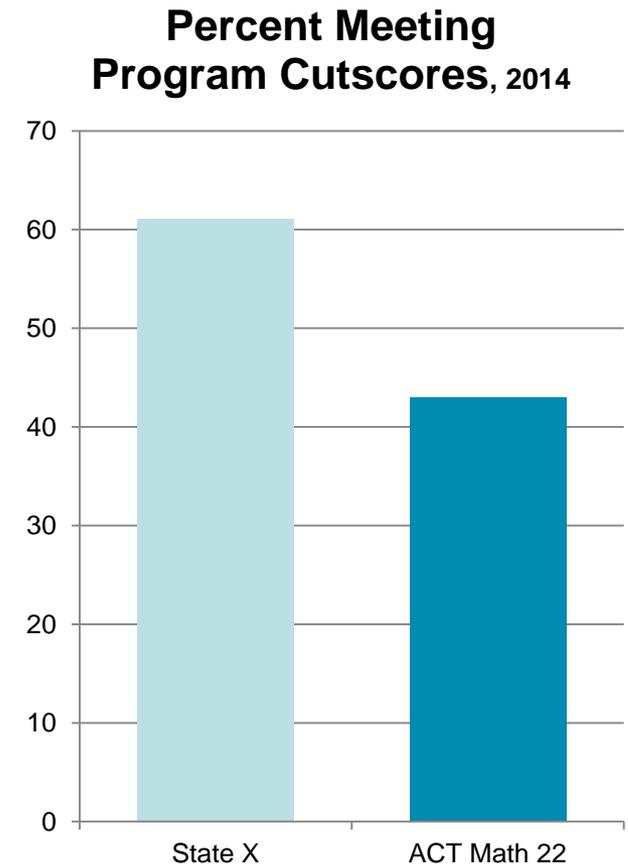
⌘ What ACT* score to use in the state for “CCR cutscore? (composite ACT score, 25%ile admitted for IHE. AFQT = Armed Forces Qualifying Test, a subset of the ASVAB. A score of 31 required for Army entrance; varies by service branch and area of interest) (Note: ELA more difficult; SAT has not produced an official way to combine Reading and Writing.)

State X's IHE/Career	Min. ACT score
State X selective 4-yr	22
Other State X 4-yr	20
AFQT (31-49)	15-16

College Course	ACT Subject-Area Test	The ACT® Benchmark
English Composition	English	18
College Algebra	Mathematics	22
Social Sciences	Reading	22
Biology	Science	23

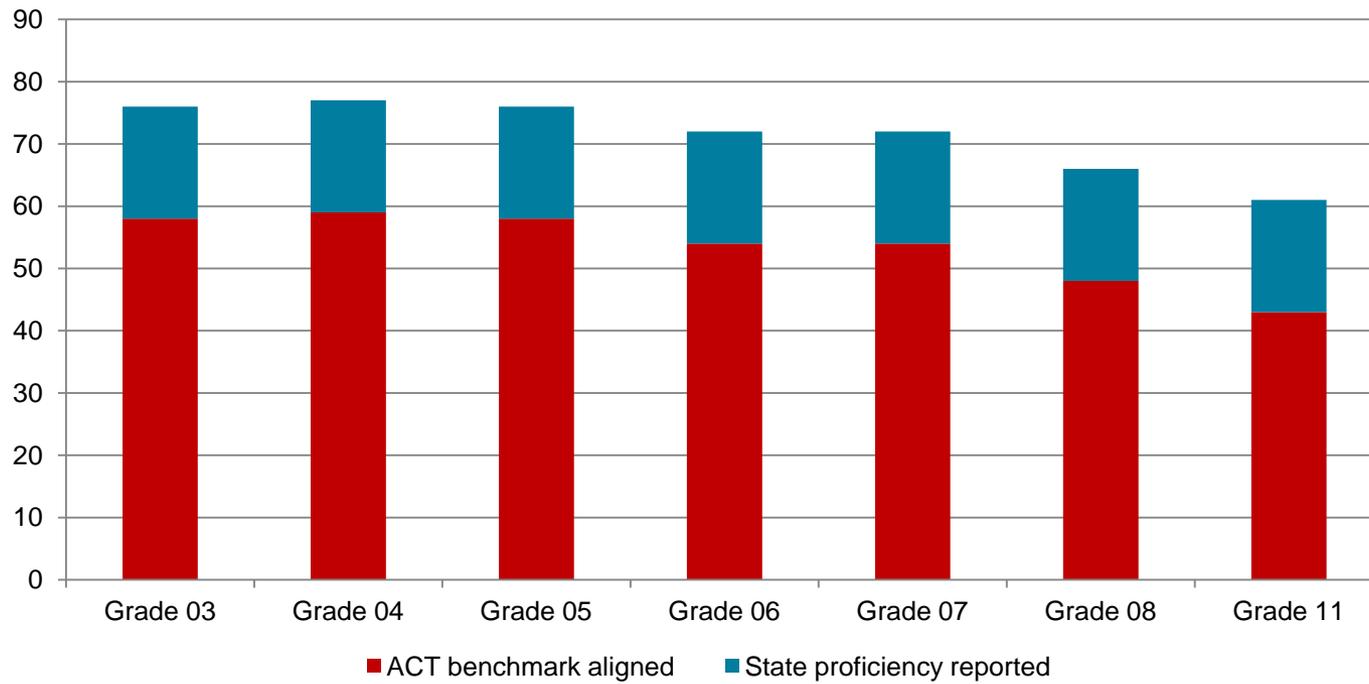
Systemic coherence: high school

- ⌘ Large difference between percentages of students in state meeting state proficiency cut and ACT national mathematics benchmark in high school (approx. 84% of graduating students had an ACT score)



Systemic coherence: grades 3-high school

⌘ Illustrative adjustments to grades 3-11 achievement to correspond with ACT “percent meeting benchmark” (mathematics, 2014)



Coherence of performance “meaning”

⌘ State must create **interpretations** of performance that are coherent across assessments

- Evaluation of soundness of respective ALDs, blueprints, standard setting, reported scores (including subscores)
- Sound content analysis of comparability of Achievement Level Descriptors, Test Blueprints, and reported scores
- Sound linking/predictive study, including considerations of populations, assessment conditions, scale properties for all cutscores; linking study design and execution; credibility
 - Comp’-rable = means the same
 - Com-par’-able = has a systematic (e.g., statistical) relationship
- Consideration of systemic coherence over grades, uses, time

Other systemic implications

⌘ Consider implications of multiple high school assessments for rest of educational system:

- Grades 3-8 ALDs and cutscores
- Accountability system
- “Career readiness” efforts
- School/district support
- Public support
- Assessment contract planning (e.g., volume pricing)
- State department of education capacity (e.g., management support and communication)

Standard-setting plan considerations

- ⌘ Consider setting performance standards starting with high school
- ⌘ Consider using separate panels with expertise in:
 - Content-based judgments
 - Policy-based judgments
 - Bring together content-based recommendations, empirical-based recommendations, and other policy considerations
 - “Policy-book” standard-setting (see Haertel; ADP Algebra 2 documented by Pearson)

Not strongly addressed in ESSA

- ⌘ Opportunity and Access: mostly subgroups
- ⌘ Educational Innovations

Issues to consider - 1

- ⌘ What are the current status and educational needs of subgroups in the state? How is data collected to identify each of the reported subgroups? How will the state determine major racial/ethnic subgroups for schools accountability?
- ⌘ How will your state incorporate subgroup performance into annual school and subgroup accountability determinations required by ESSA? Will you do more than is required by ESSA in reporting, making annual determinations, or identifying/assigning consequences, e.g., include district or state goals/accountability?
- ⌘ How will you decide how to specify the key decisions given to states by ESSA (e.g., defining “consistently underperforming,” “chronically underperforming,” and requirements to exit from targeted or comprehensive support and assistance)?

Issues to consider – 2

- ⌘ How will you decide how to balance the various tensions in ESSA accountability, such as between subgroup inclusion, unreliability, and possible over/under-identification when establishing a minimum-n? What safeguards will you build in, and how will you know if they are working appropriately?
- ⌘ Will the state disaggregate data by subgroups for indicator(s) of School Quality/Student Success (not required)?
- ⌘ How will you operationalize the measurement of and goals for indicators (e.g., will Academic Progress indicator goals, e.g., growth, be normative or criterion, conditioned on variables such as prior achievement or anchored on an end target, vary for individual (students, subgroups, schools) or be the same)? How will measurements of interim progress be set (e.g., what should be the relationships between meeting Achievement and Progress goals)?
- ⌘ How will you get evidence that your subgroup accountability does what it is intended to do and is not accompanied by undue negative consequences, especially that schools are being accurately classified in terms of (not)qualifying for high-stakes consequences, and that the supports and interventions are appropriate?

Next steps

- ⌘ Determine how the state's accountability goals and theory of action, including subgroup accountability, are similar to and different from those the state developed under NCLB and/or ESEA Waivers.
- ⌘ Determine if the state will apply for a waiver under 8401. States with rigorous, proven accountability systems in place may seek waivers of new accountability requirements that are inconsistent with their current approach to accountability and can demonstrate it will advance student achievement.
- ⌘ Decide how the state will incorporate subgroup accountability, including the specific ESSA subgroup accountability requirements, into the state's comprehensive accountability system, including supports and interventions.
- ⌘ Identify key specific decisions about subgroup accountability specific to the state's plan and context.
- ⌘ Review the draft ESSA Accountability regulations (issued 5/31/16 in Federal Register), draft ESSA assessment and ESSA Demonstration Authority regulations and analyze the final regulations.