

ESSA & HB 3218 Requirements & 16-17 Transition Year Overview

Dr. Katie Dunlap, Deputy Superintendent of
Assessment & Accountability

Craig Walker, Executive Director of State Assessments

HB 3218 - Assessments

- Eliminates non-federally required tests except US History in high school.
 - Repeals EOIs, OCCTs and ACE as currently given
- Eliminates requirement for districts to administer a fine arts assessment.
- SBE to adopt rules to allow for students transferring to OK after junior year of high school to not be denied diploma.
- Eliminates the Retest/Winter/Trimester and Summer testing windows.



Transition Year in 2016-17

- Assessments include:
 - ELA & Math each year in grades 3-8, and once in high school
 - Science in grades 5 and 8, and once in high school.
 - U.S. History once in high school
- College & Career Readiness Assessment (Grade 11)
- Solicit a Request for Proposals (RFP) by May 2017 to ensure a smooth transition to the legislature approved student assessment system for 2017-18 & beyond.



OKLAHOMA SCHOOL TESTING PROGRAM

TEST BLUEPRINT **MATHEMATICS** 2016-2017 **GRADE 7**

This blueprint describes the content and structure of an assessment and defines the ideal number of test items by strand and standard of the Oklahoma Academic Standards (OAS).

IDEAL PERCENTAGE OF ITEMS	IDEAL NUMBER OF ITEMS	STRANDS AND STANDARDS
20%	10	NUMBER AND OPERATIONS 7.N.1 Representation and Comparison of Rational Numbers (4) 7.N.2 Number Operations and Absolute Value (6)
30%	15	ALGEBRAIC REASONING AND ALGEBRA 7.A.1 Proportional Relationships (4) 7.A.2 Proportions, Rates and Ratios (5) 7.A.3 Linear Equations and Inequalities (4) 7.A.4 Order of Operations (2)
30%	15	GEOMETRY AND MEASUREMENT 7.GM.1 Surface Area and Volume of Rectangular Prisms (2) 7.GM.2 Trapezoids and Composite Figures (2) 7.GM.3 Circles (5) 7.GM.4 Transformations
20%	10	DATA AND PROBABILITY 7.D.1 Data Analysis (6) 7.D.2 Probability (4)

Grade 3-8 ELA/Math/Science Standards Based Assessments

2015-16

PASS Standards
Stand-alone writing tests for 5/8
Untimed tests
Stand-alone science items

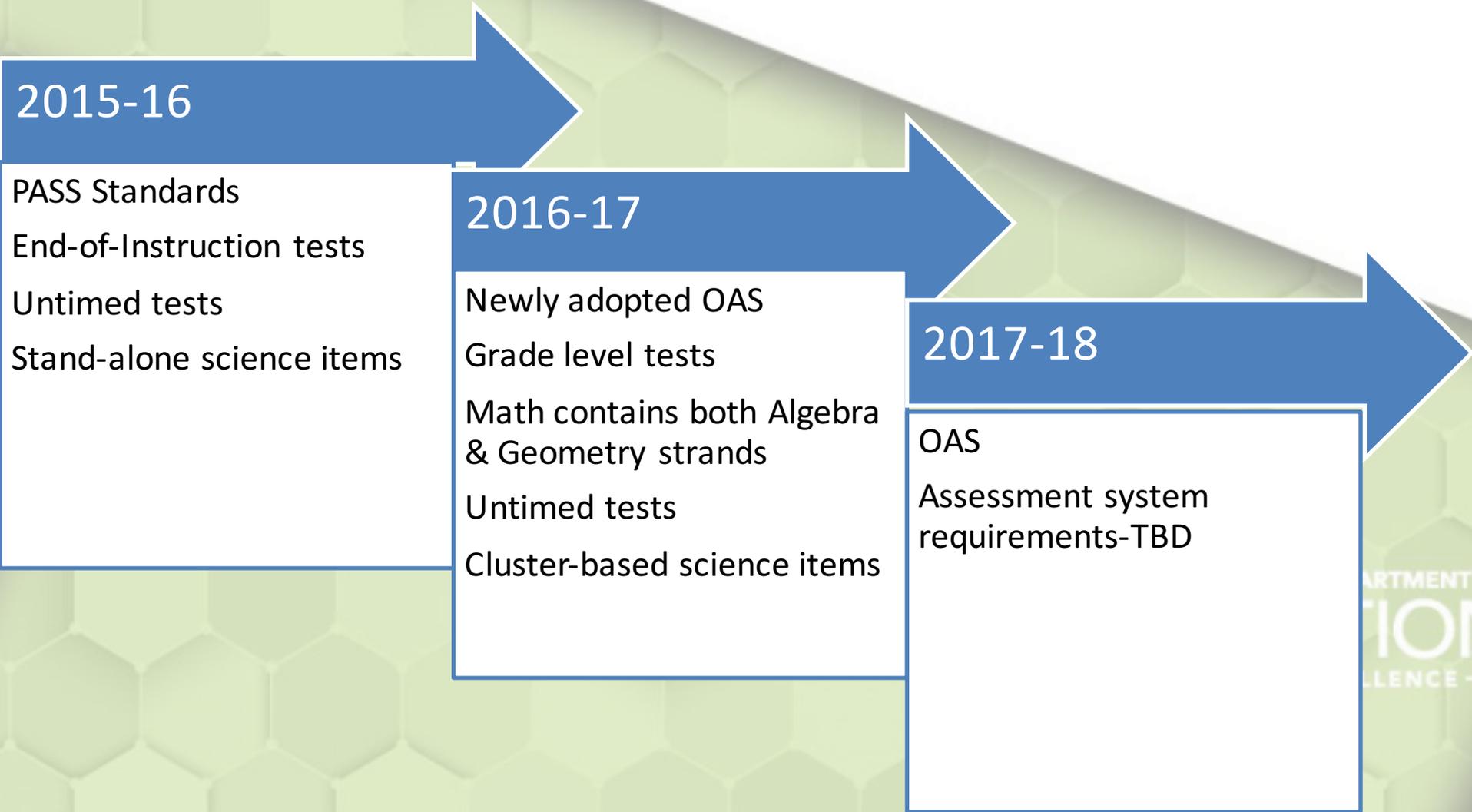
2016-17

Newly adopted OAS
Writing response included in ELA 5/8
Untimed tests
Cluster-based science items

2017-18

OAS
Assessment system requirements-TBD

Grade 10 ELA/Math/Science Standards Based Assessments



2016-17 Assessment Program

Grade	Math	ELA	Science
Grade 3	✓	✓	
Grade 4	✓	✓	
Grade 5	✓	✓	✓
Grade 6	✓	✓	
Grade 7	✓	✓	
Grade 8	✓	✓	✓
Grade 10	✓	✓	✓



- College/Career Readiness Assessment
- U.S. History once in high school



ESSA & Assessment

- States must administer high-quality annual assessments in at least reading/language arts, mathematics, and science that meet nationally recognized professional and technical standards.
 - States are required to test students in reading or language arts and math annually in grades 3-8 and once in grades 9-12, and in science once in each of the following grade spans: 3-5, 6-9 and 10-12.
- Maintains Assessment Peer Review Requirements
- Requires states to measure the full depth and breadth of their state academic content standards

HB 3218 - Assessments

- For 2017-2018, assessments include:
 - ELA & Math each year in grades 3-8, and once in high school
 - Science in grades 5 and 8, and once in high school
 - U.S. History once in high school
 - Any others adopted by SBE
 - May include college and career ready assessment
- Requires SBE to adopt rules for assessments in compliance with ESSA by December 31, 2016.
- RFP for assessments to be administered in coordination with six-year textbook adoption cycle.

HB 3218 - Assessments

- By January 1, 2017, the SBE shall adopt the assessment requirements as studied & developed:
 - After SBE adoption, submit to Legislature.
 - Legislature has 30 calendar days to approve or disapprove requirements, with or without instructions, by joint resolution (JR).
 - If Legislature fails to adopt JR, requirements are disapproved.
 - If disapproved, SBE can resubmit prior to last 30 calendar days of legislative session.



HB 3218 - Assessments

- Statewide student assessment system shall include assessments that:
 - Align to standards
 - Provide a measure of comparability to other states
 - Yield norm-referenced and criterion-referenced scores
 - Statistically reliable and accurate
 - In high school, provide measure of future academic performance



HB 3218 - Assessments

- The SBE in consultation with the OSRHE, CEQA, Career Tech State Board, Sec. of Ed., and Workforce Development shall include assessments in the statewide student assessment system that:
 - Multi-measure approach to high school graduation
 - Determination of the performance level on the assessments at which students will be provide remediation/intervention.
 - Means for ensuring student accountability
 - Ways to make the school testing program more efficient.



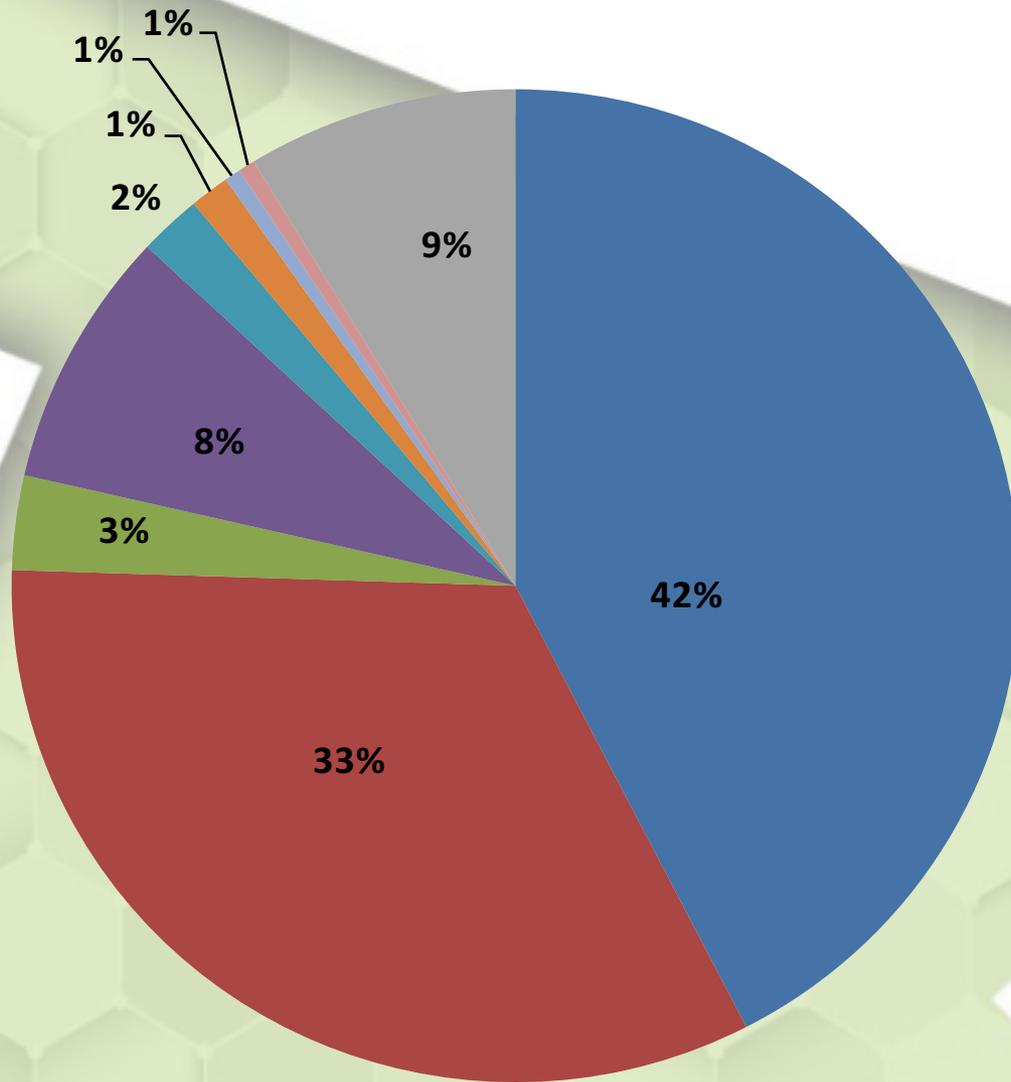
Assessment Stakeholder Poll Results

Craig Walker, Executive Director of State Assessments



OKLAHOMA STATE DEPARTMENT OF
EDUCATION
— CHAMPION EXCELLENCE —

What is your role?



■ School Administrator

■ Teacher

■ Paraprofessional

■ Specialized instruction support personnel

■ Community member

■ Parent

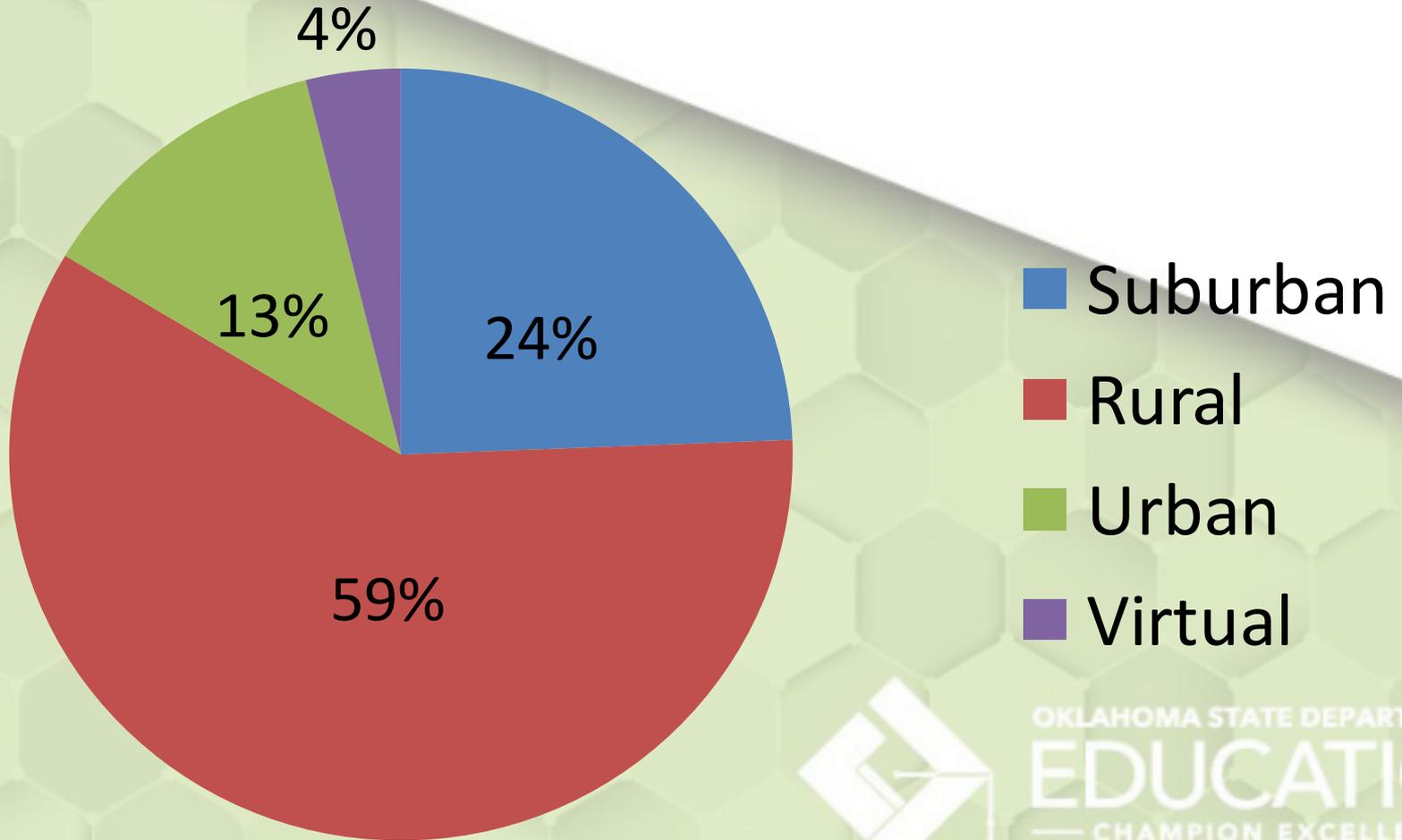
■ Business leader

■ Student

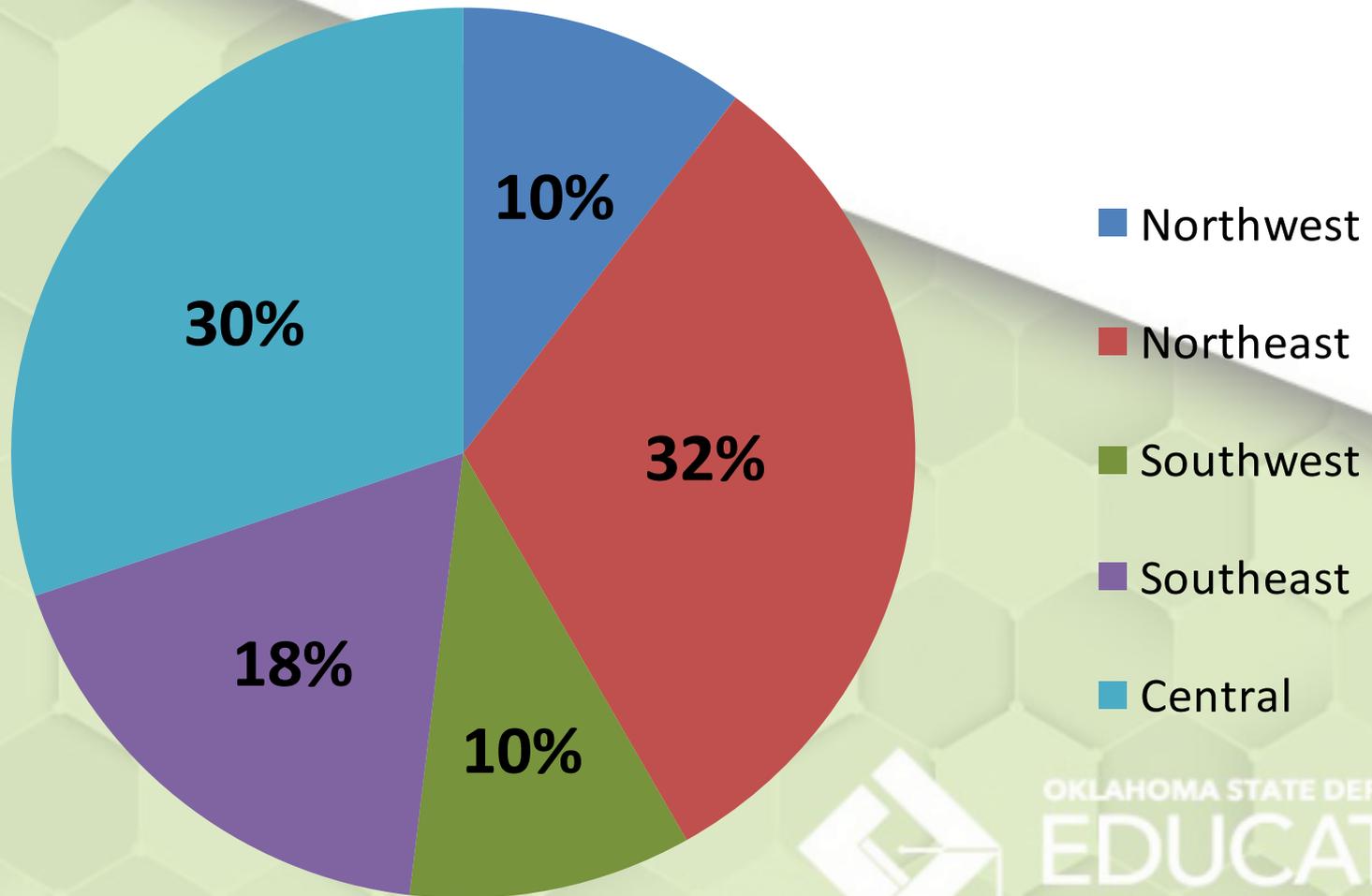
■ Other

What community or school type do you represent?

(Select all that apply.)

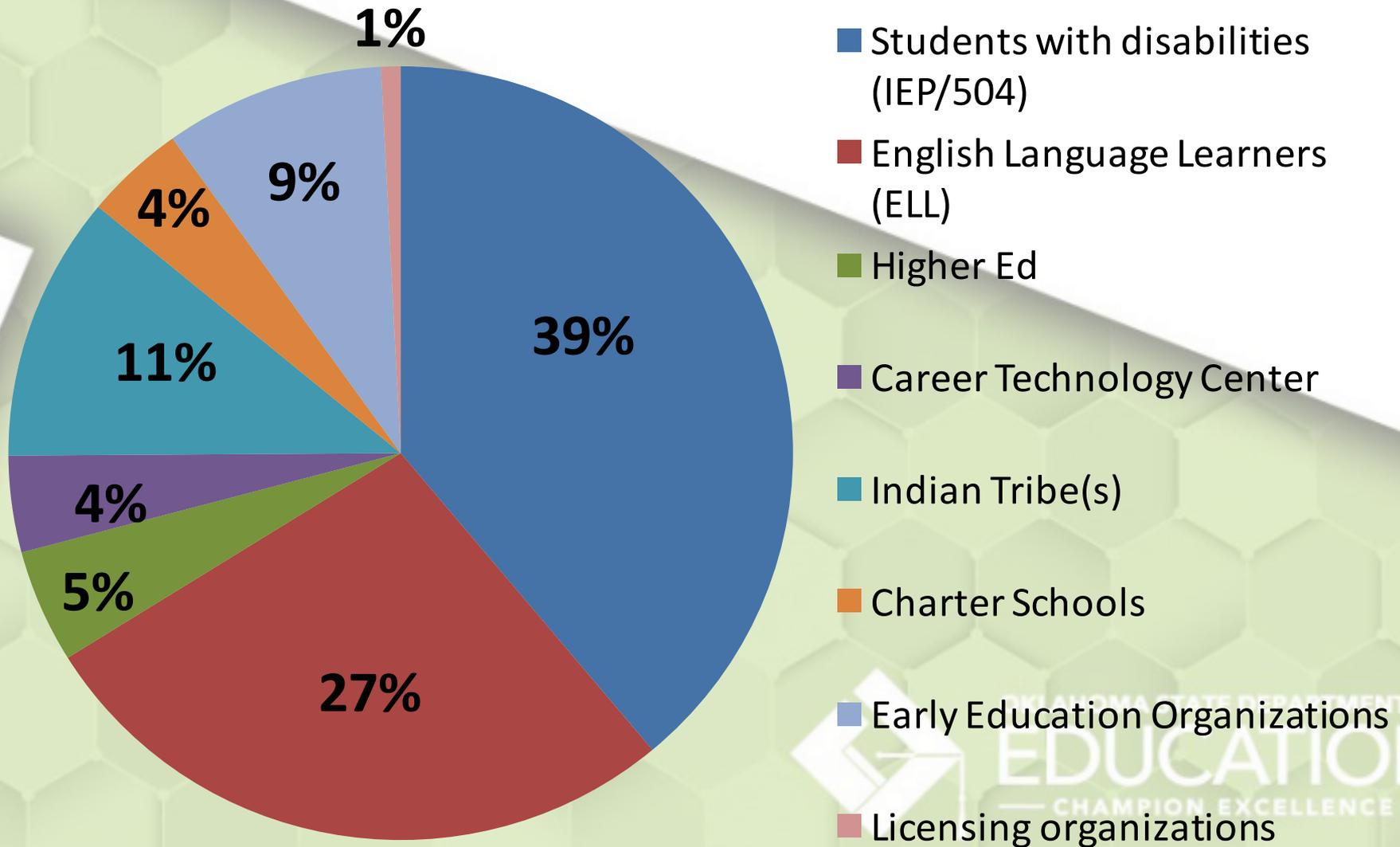


What part of Oklahoma do you represent?



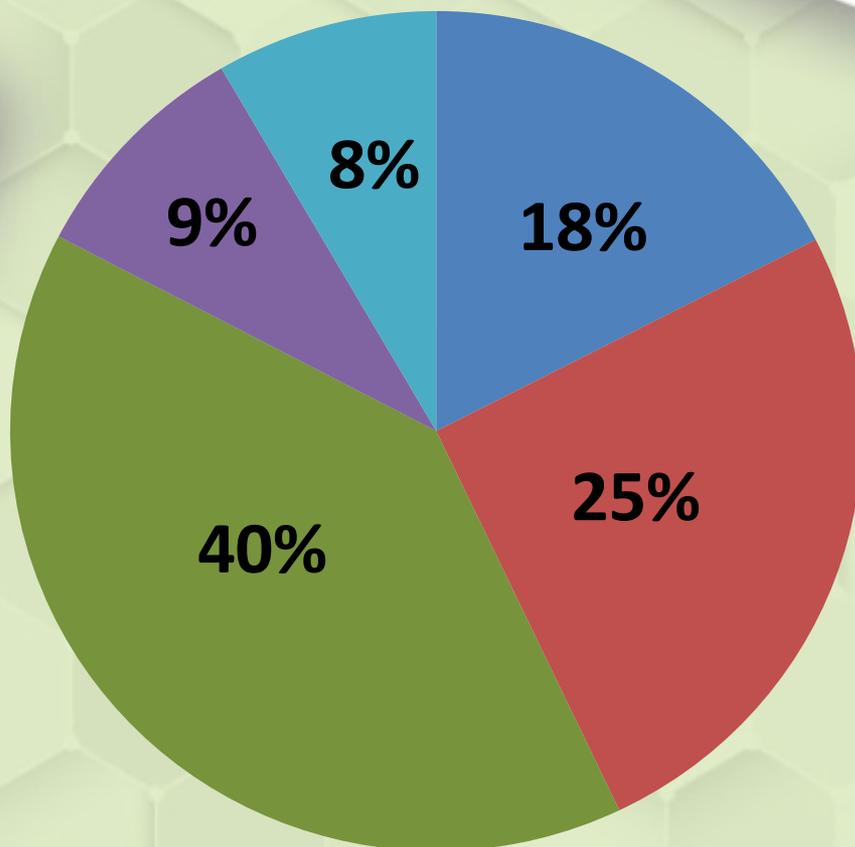
What group(s) are you a representative for?

(Select all that apply.)



Which of the following measure college readiness?

(Select all that apply.)



■ GPA/Class rank

■ Advanced coursework

■ College entrance exam

■ Oklahoma School Testing Program State Assessments

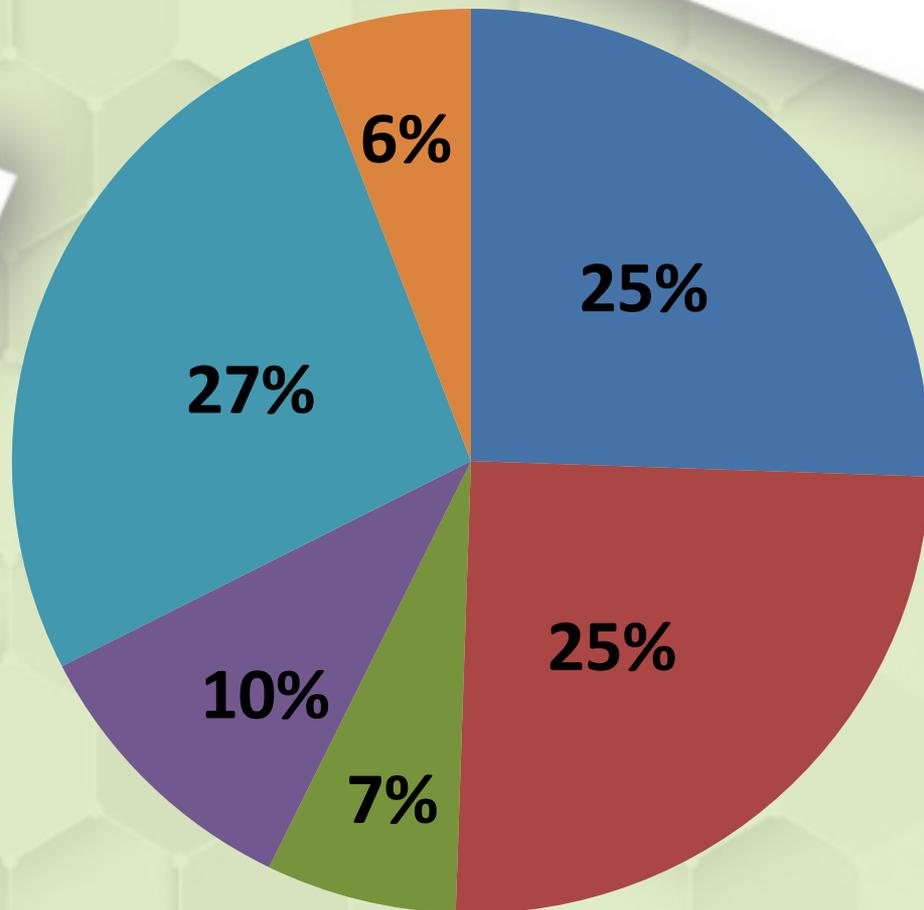
■ Other



OKLAHOMA STATE DEPARTMENT OF
EDUCATION
— CHAMPION EXCELLENCE —

Which of the following measure career readiness?

(Select all that apply.)



■ Industry certification

■ Career pathway assessment (ASVAB, WorkKeys, etc.)

■ Oklahoma School Testing Program State Assessments

■ State graduation requirement completed

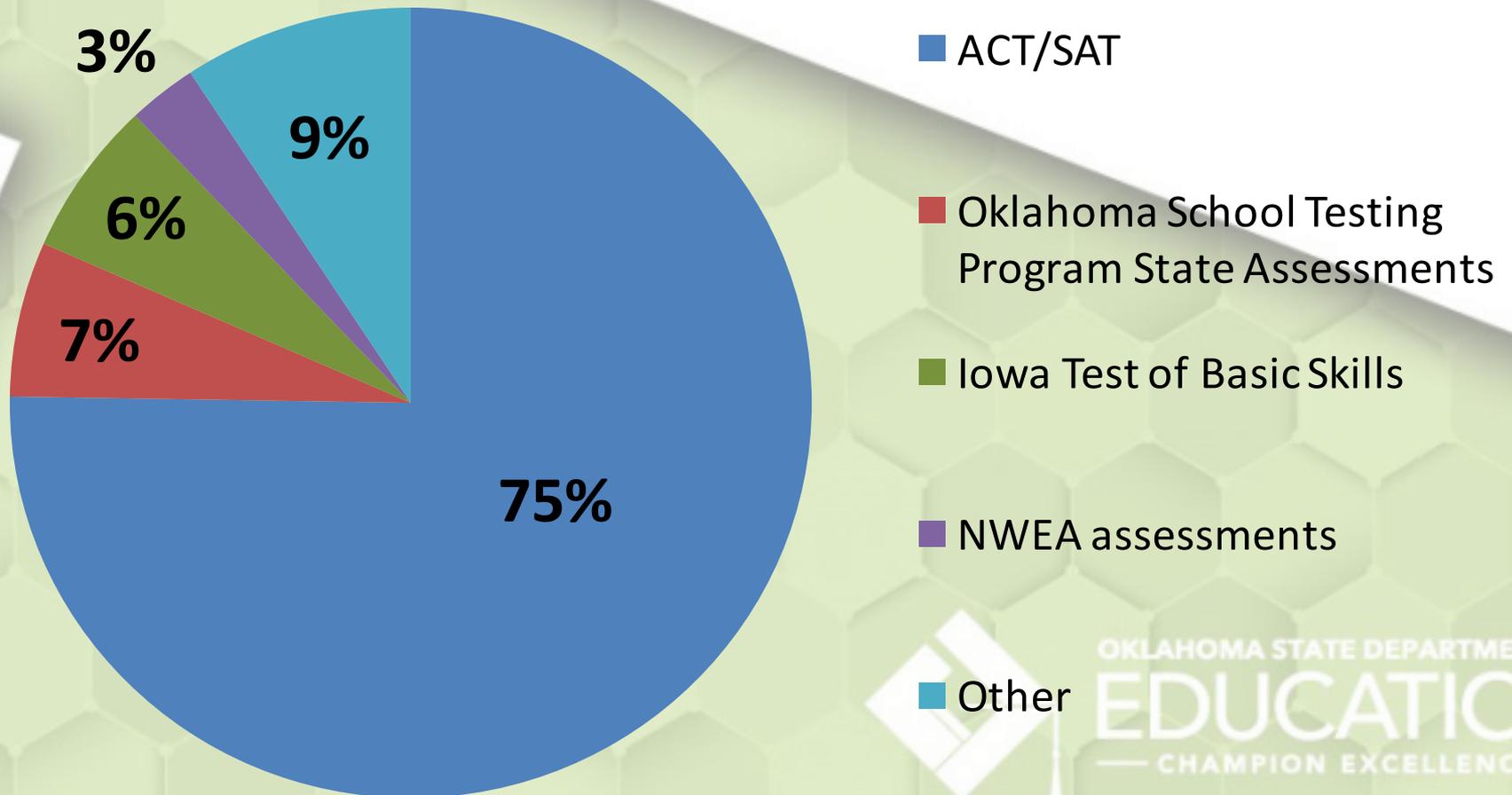
■ Internship/Apprenticeship

■ Other



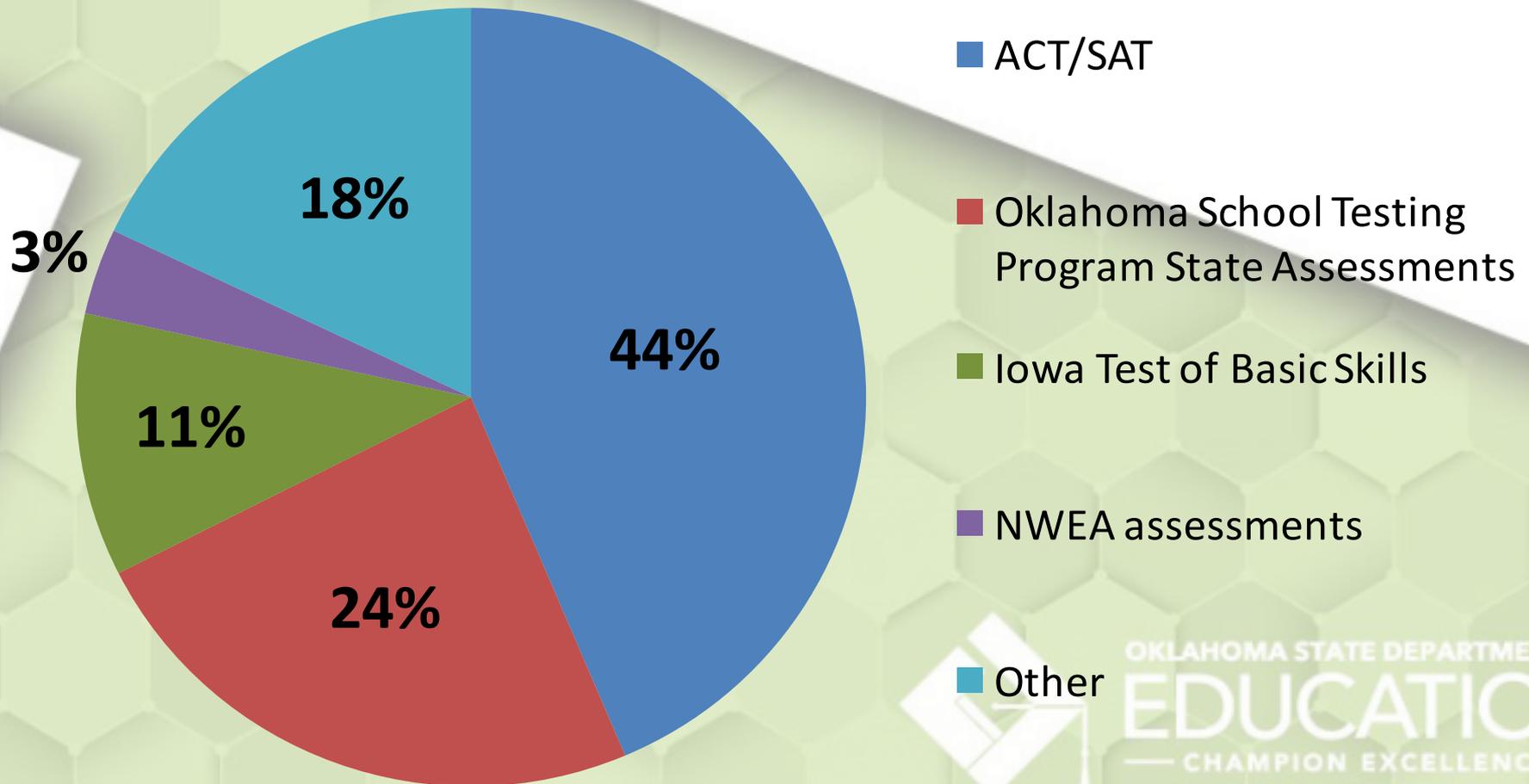
What test(s) would you prefer students to take for college- and career-readiness?

(Select all that apply.)

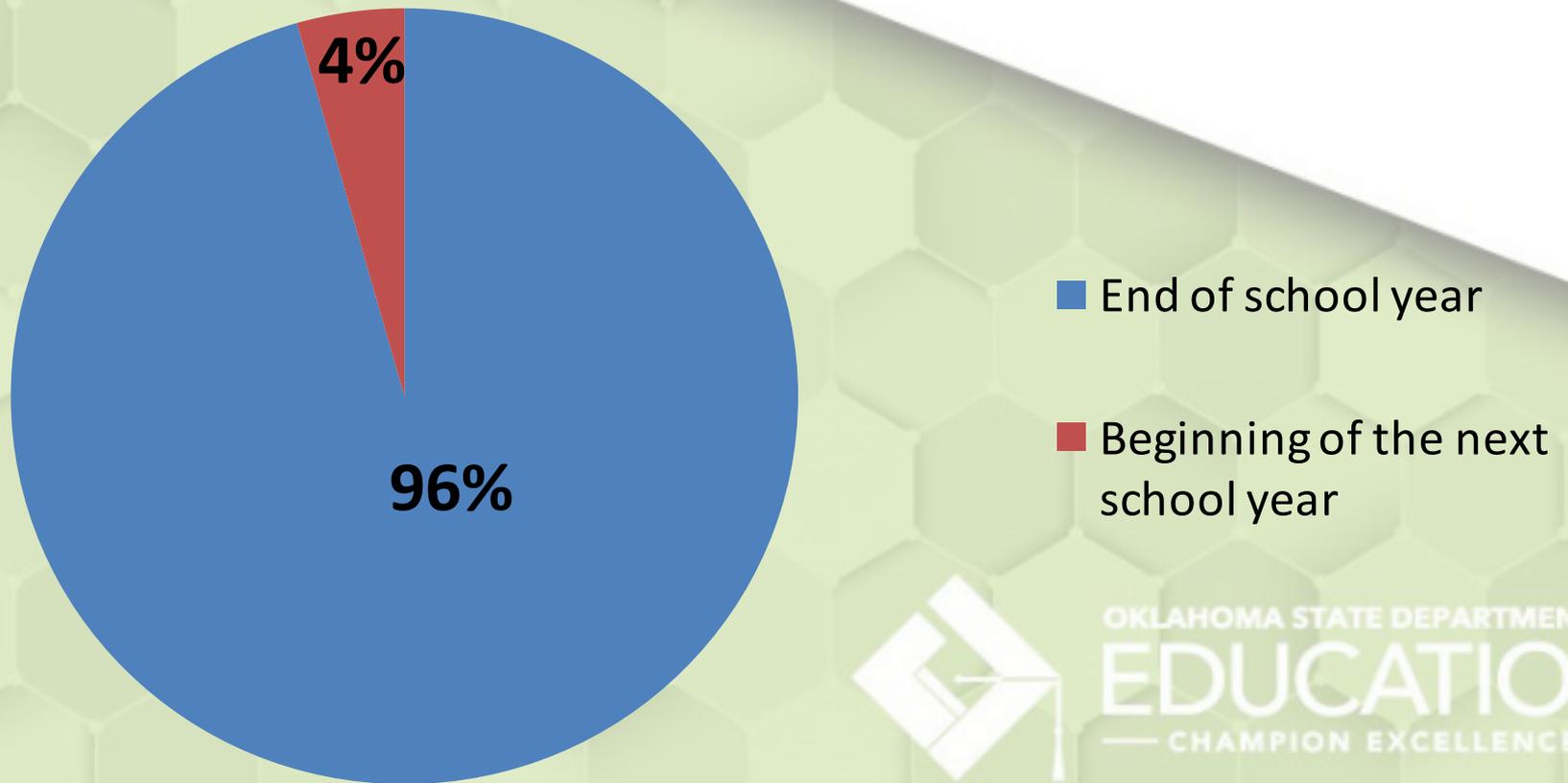


What test(s) would you prefer students to take for school accountability purposes?

(Select all that apply.)

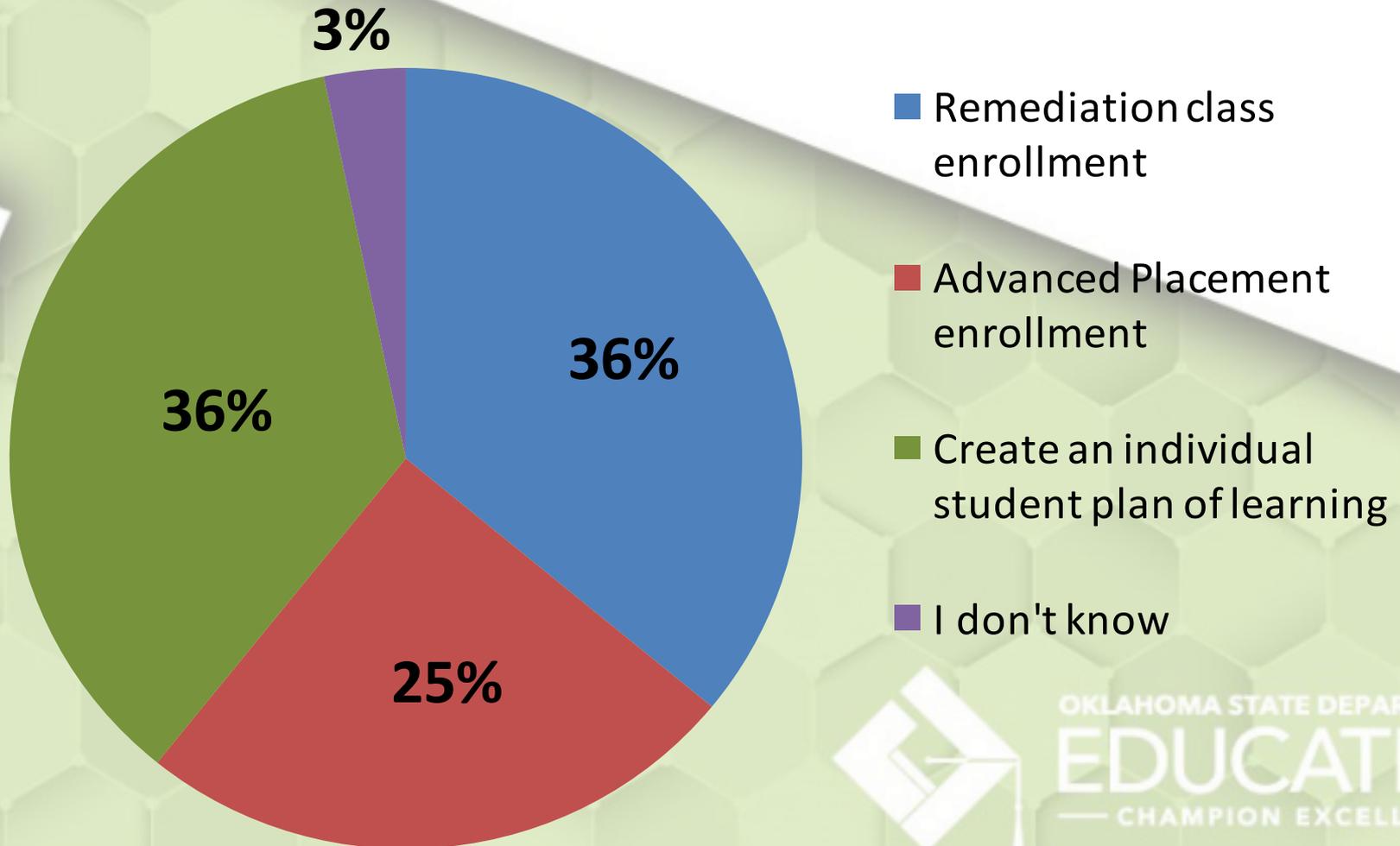


For the purpose of the summative reporting, when would be most beneficial to receive reports?



How do you think schools should use state assessment results to support student success?

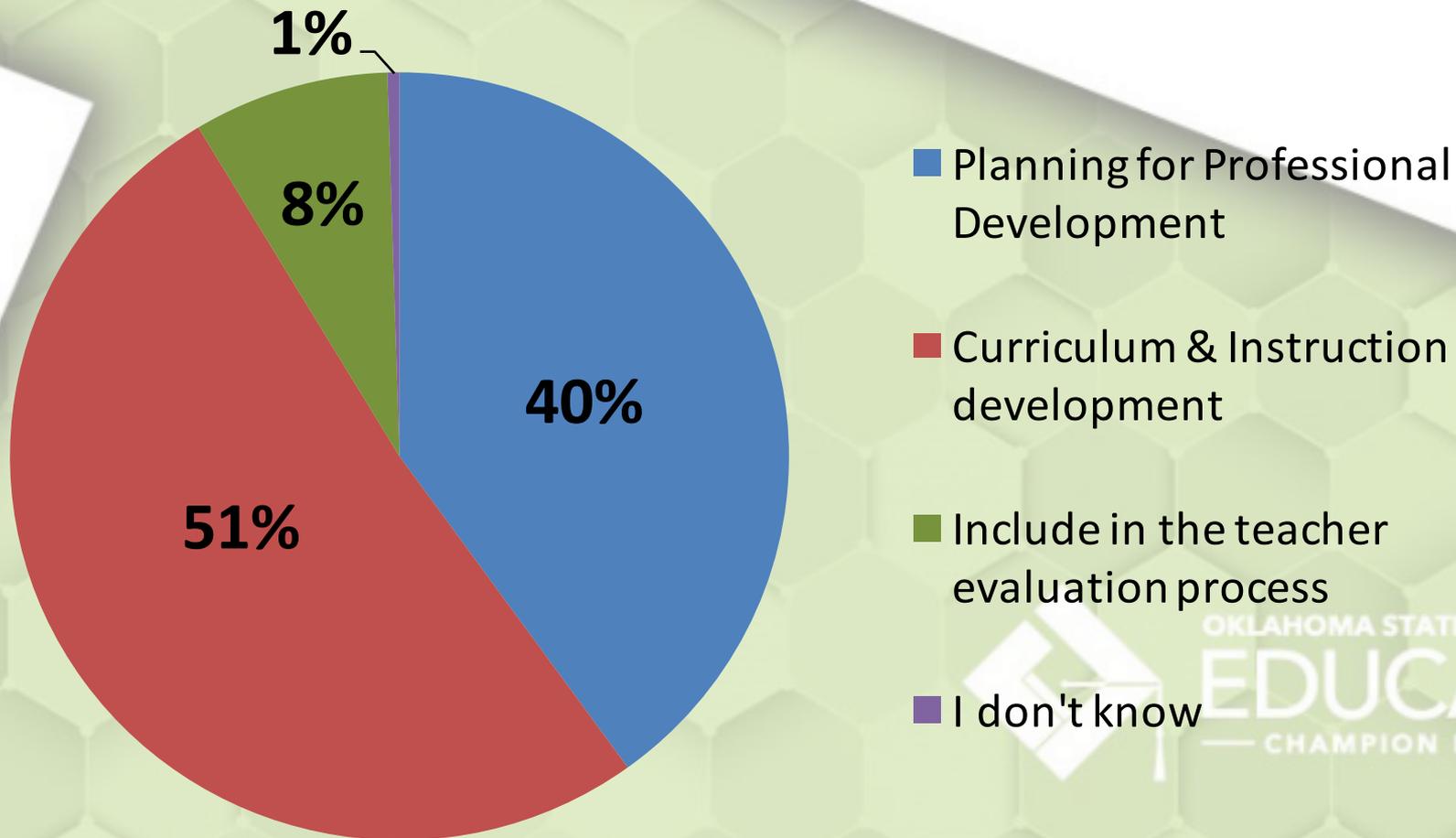
(Select all that apply.)



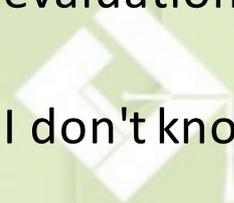
OKLAHOMA STATE DEPARTMENT OF
EDUCATION
— CHAMPION EXCELLENCE —

How do you think schools should use state assessment results to strengthen teacher quality and instruction?

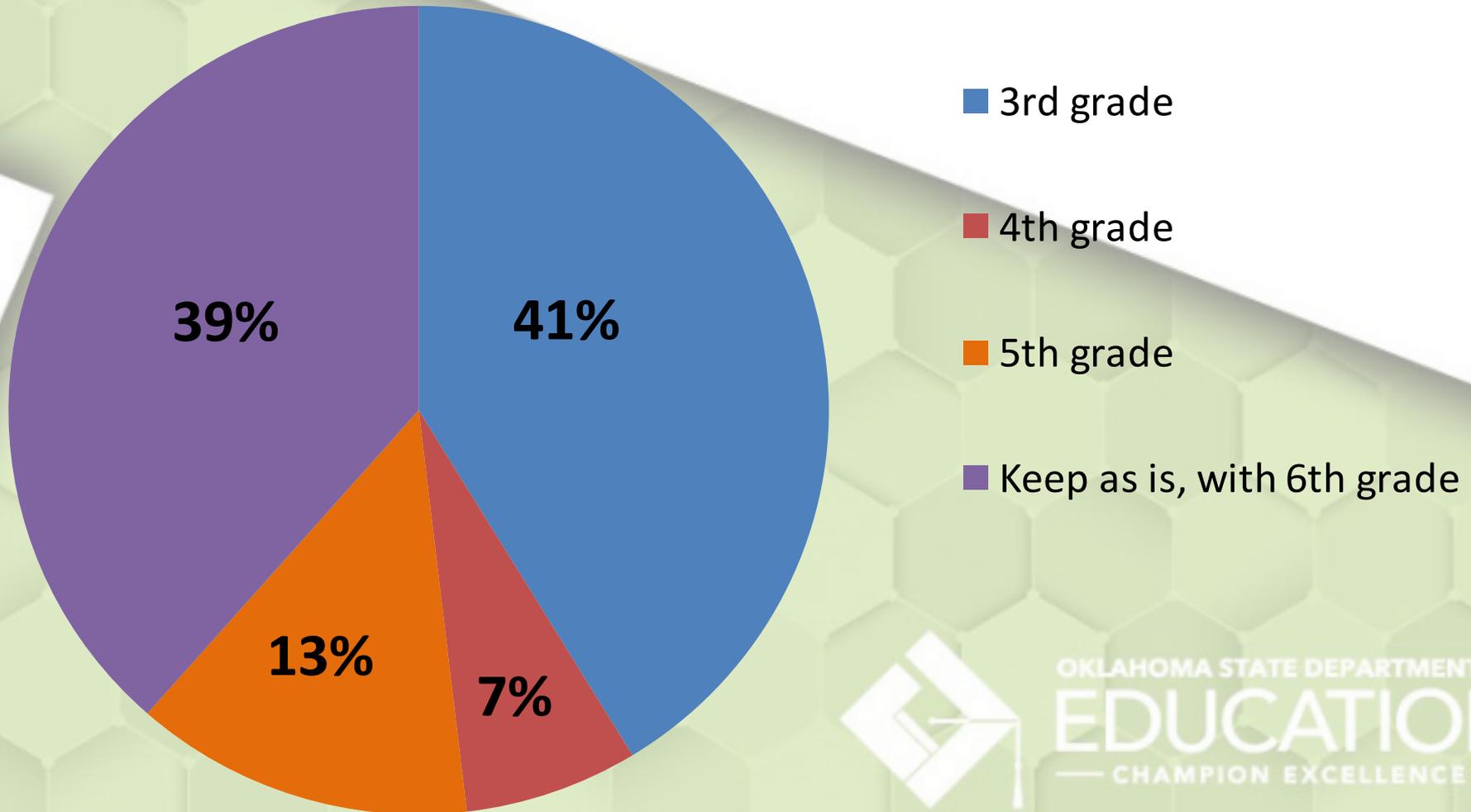
(Select all that apply.)



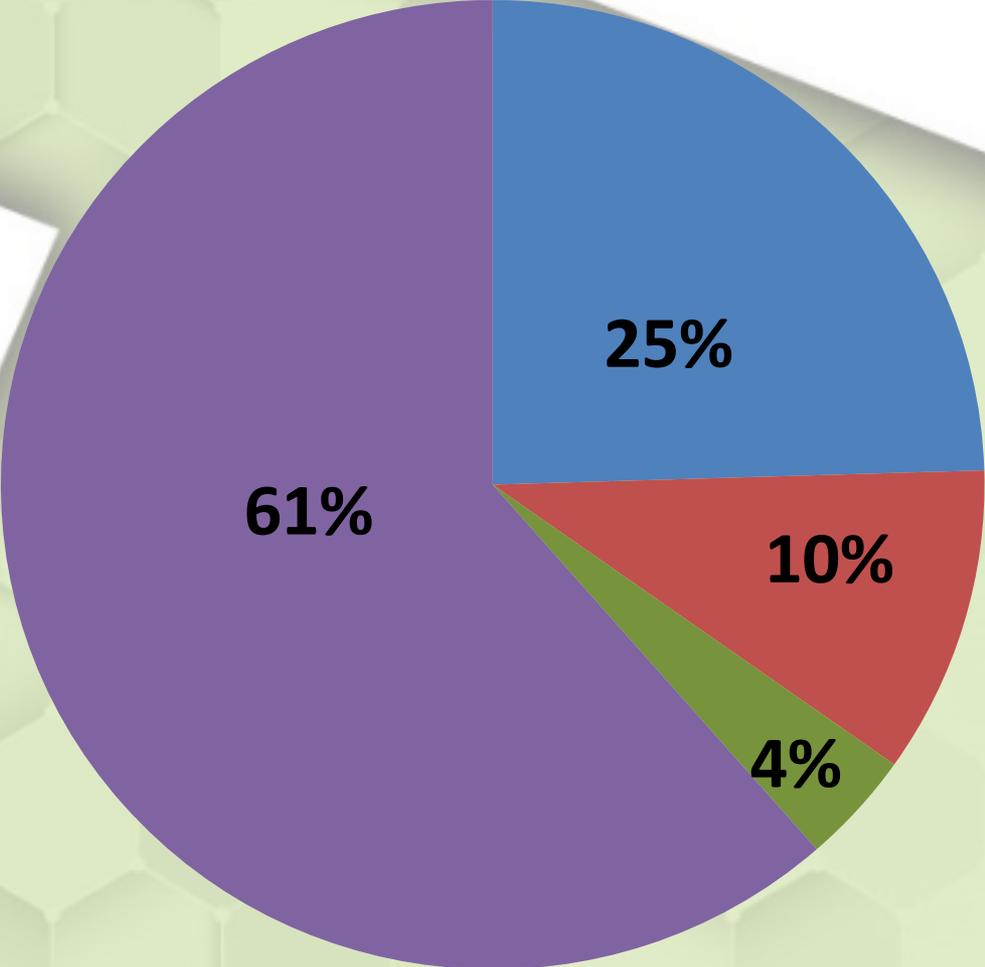
- Planning for Professional Development
- Curriculum & Instruction development
- Include in the teacher evaluation process
- I don't know



At what grade-level should students begin taking computer based assessments?



How do we best ensure student accountability on state assessments?



Put results on transcript

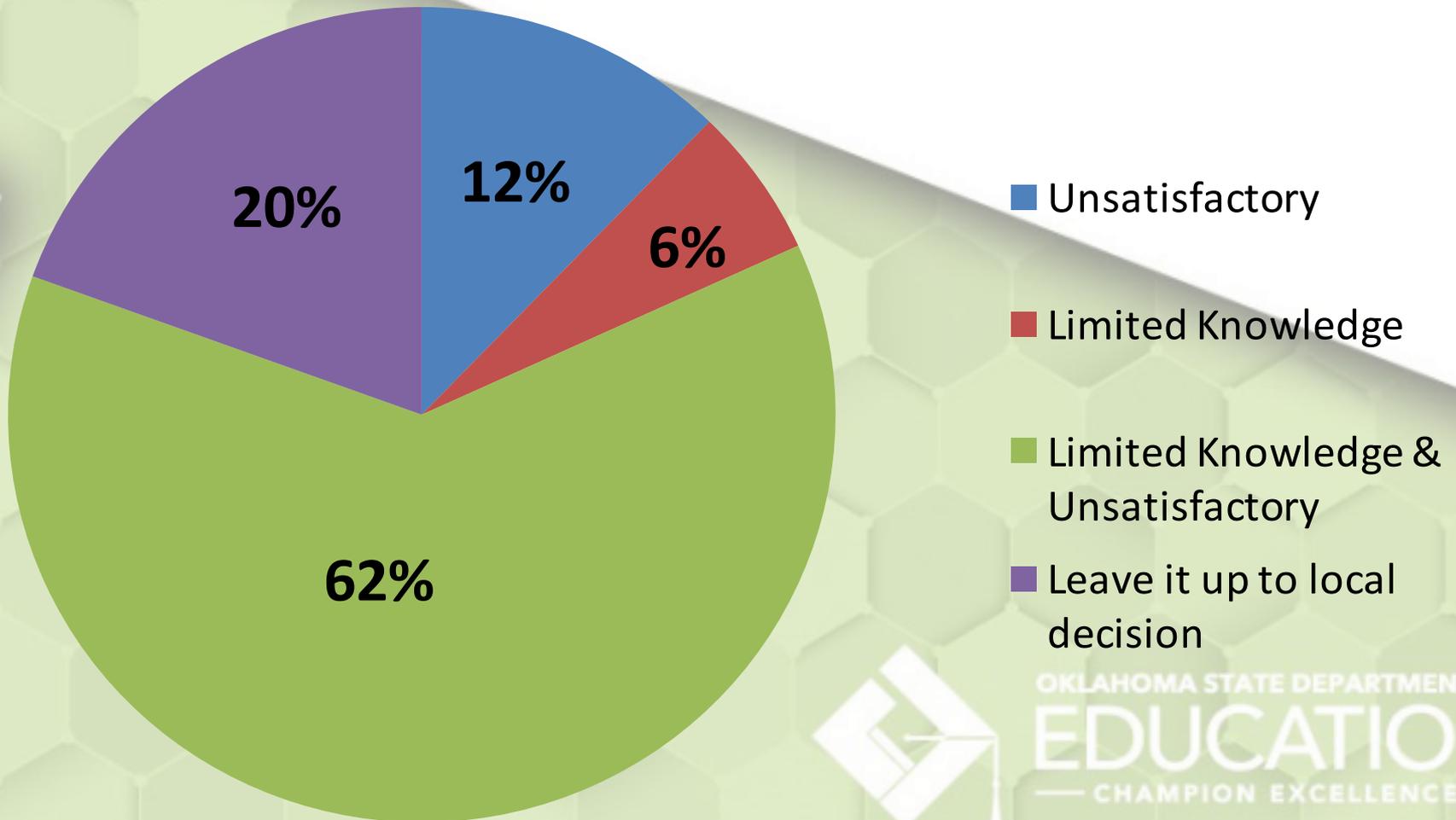
Include in student's grade

Include in student's GPA

Local Decision

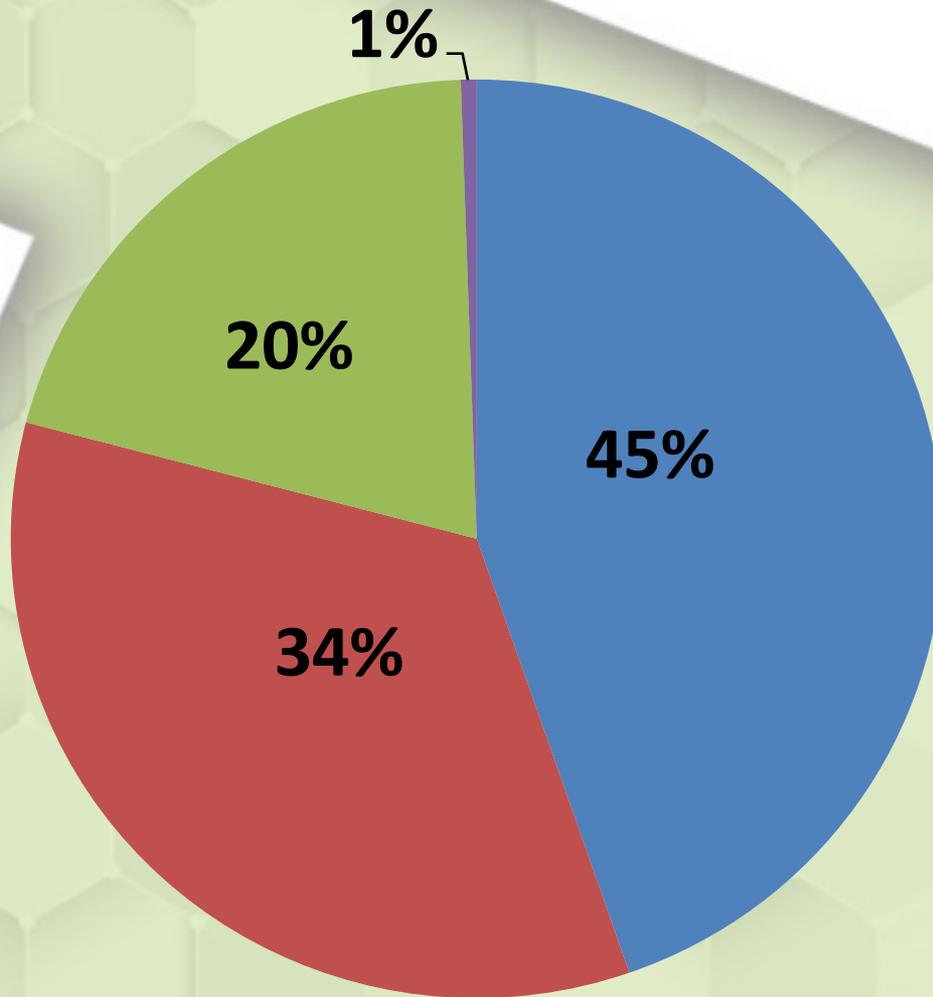


At what performance level should students be provided remediation?



What is the purpose of the Oklahoma School Testing Program state assessments?

(Select all that apply.)



■ To measure progress toward college and career readiness

■ To indicate the need for remediation

■ To determine proficiency based promotion

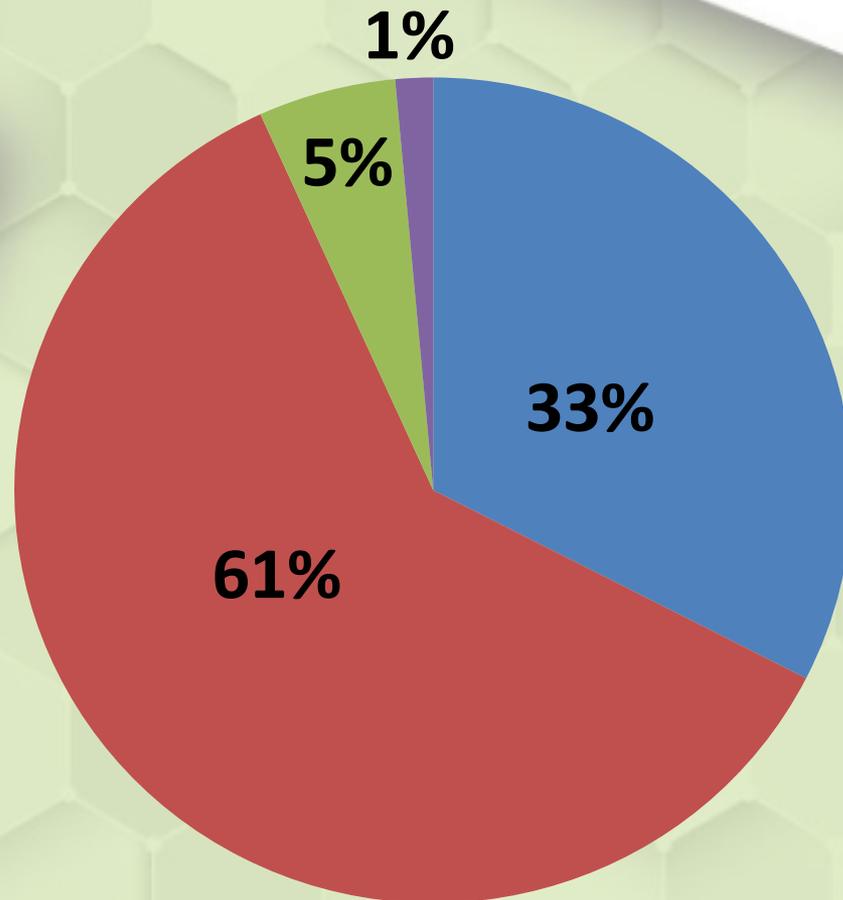
■ To be included in the student's GPA calculation



OKLAHOMA STATE DEPARTMENT OF
EDUCATION
— CHAMPION EXCELLENCE —

What recommendation(s) would you give to promote an effective and efficient testing system?

(Select all that apply.)



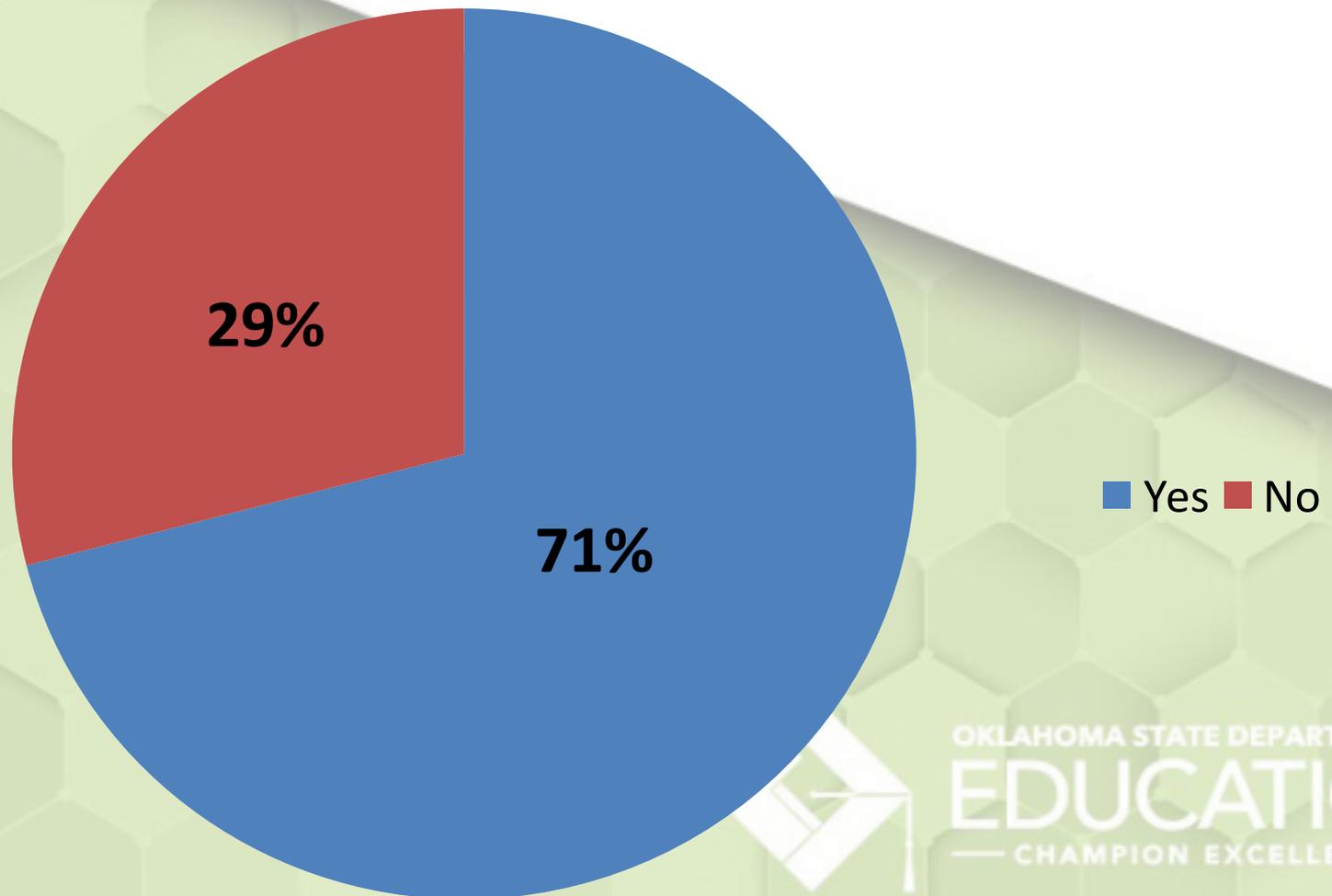
■ Eliminate writing prompts and open-ended responses from assessments

■ Combine different contents into one assessment

■ Combine different grade levels into one assessment

■ Adding additional subject area assessments

Should state assessment results show how our students compare to other states?



DESIGNING A BALANCED ASSESSMENT SYSTEM

MARIANNE PERIE, UNIVERSITY OF KANSAS

H. GARY COOK, WISCONSIN CENTER FOR EDUCATION RESEARCH

JUAN M. D'BROT, CENTER FOR ASSESSMENT

AUGUST 4, 2016

GOAL FOR TODAY

- Describe components of a balanced assessment system
 - Formative
 - Interim
 - Summative
- Discuss considerations for designing assessments
 - Purpose of assessment
 - Length versus information
 - Pros and cons of various item types

BALANCED ASSESSMENT SYSTEM

Formative Tools

- Based on learning theory
- Minute by minute between teacher and student
- Includes instructional resources to build student learning
- Not intended for aggregation or teacher/program evaluation

Interim Assessment

- Optional
- District choice
- Diagnostic information
- Tracks growth
- Predicts summative
- Can be aggregated at classroom or building level

Summative Assessment

- End of year
- Can be used as a snapshot within and across schools and districts
- ESSA eliminated punitive consequences
- Information & transparency
- Examine equity and resource allocation

All based on Oklahoma Standards and Goals for Students

CONSIDERING THE GOALS OF AN ASSESSMENT



DESIGNING AN ASSESSMENT

- First, determine goal/purpose of the assessment
 - Provide information on student performance relative to some target
 - Sort schools
 - Identify achievement gaps among student groups
 - Provide instructional feedback
 - Evaluate instruction or instructor
- Be judicious: An assessment purporting to serve multiple purposes serves no purpose well.

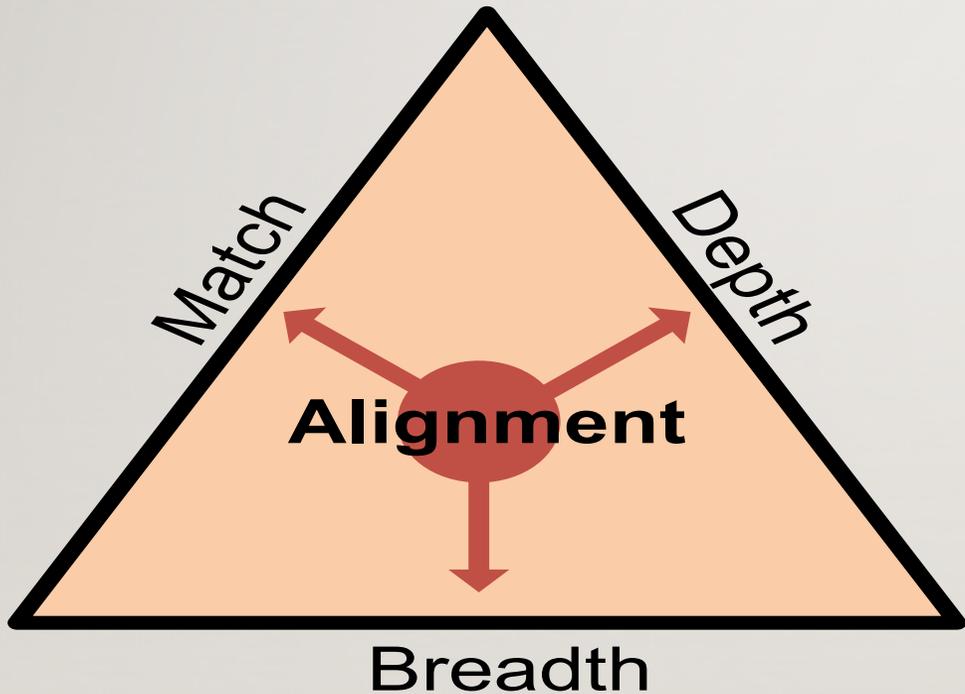
CONSIDER THESE QUESTIONS...

- What do I want to learn from this assessment?
- Who will use the information gathered from this assessment?
- What action steps will be taken as a result of this assessment?
- What professional development or support structures should be in place to ensure the action steps are taken?

GOALS ARE DEPENDENT ON CONSTRAINTS

- Alignment to the Oklahoma Academic Standards
- Objective items to eliminate subjective bias
- Standardized settings to minimize the effect of local factors
- Sufficient reliability to support decisions (more on reliability later)

ALIGNED TO STATE STANDARDS



- **Match:** The degree to which assessment items connect to standards
- **Depth:** The degree to which assessment items cover the cognitive complexity of the standards
- **Breadth:** The degree to which assessment items cover the full range of the standards

RE-CONSIDER THESE QUESTIONS

- What do I want to learn from this assessment?
- Who will use the information gathered from this assessment?
- What action steps will be taken as a result of this assessment?
- What professional development or support structures should be in place to ensure the action steps are taken?

CONSIDERING HOW RESULTS CAN BE USED

HOW CAN RESULTS BE INTERPRETED?

- Criterion-referenced interpretations
- Norm-referenced interpretations

CRITERION V. NORM-REFERENCE

- Criterion reference interpretations support inferences to external criteria(-ion) and often are associated with mastery states, e.g., state standards based assessments, graduation or certification examinations [Alignment & Mastery]
- Norm reference interpretations support inferences about ranking and the relationship between equivalent groups, e.g., college entrance examinations like ACT or SAT [Relevance & Population]

CRITERION-REFERENCED INTERPRETATIONS

- No direct comparisons to other test takers—focus on alignment and mastery
- Results are compared to some criterion or many criteria
 - A proficiency target
 - Likelihood for success in post-secondary
 - Likelihood of future success in math or reading
 - Likely proportion of correct responses
- Performance against criteria are dependent on the target of interest
 - E.g., policy requirement, practical question like success in post-secondary

NORM-REFERENCED INTERPRETATION

- Often compared to other students' scores—focus relevance and population
- Tries to support interpretations of relative performance framed as
 - Percentiles
 - Comparison against averages
 - Other normatively defined cut scores
- Comparisons are dependent on the group to which norms are defined
 - School, district, grade, state, multi-state
 - E.g., comparing my vertical jump to the NBA norm

WHAT INTERPRETATIONS MAKE THE MOST SENSE?

- Depends on the question being asked
- Tests can support
 - Only criterion-referenced interpretations
 - Only norm-referenced interpretations
 - Both norm- and criterion-referenced interpretations

HB 3218 REQUIREMENTS

- Support both norm- and criterion-referenced interpretations
- Criterion-referenced are part of our requirements
 - Performance Levels
 - Proficiency
 - Informed by Achievement Levels (descriptions of what students should know and be able to do at each grade)
- Norm-referenced interpretations warrant additional discussion

CONSIDER THESE QUESTIONS

- To whom do we want to compare students?

Samantha scored a 850 in grade 8 math and achieved a performance level of Proficient

CONSIDER THESE QUESTIONS

- To whom do we want to compare students?
Samantha scored a 850 in grade 8 math and achieved a performance level of Proficient
- What questions might we ask to better contextualize Samantha's performance?
- What other information would we need to answer those questions?

CONSIDER THESE QUESTIONS

- To whom do we want to compare students?
Samantha scored a 850 in grade 8 math and achieved a performance level of Proficient
- What questions might we ask to better contextualize Samantha's performance?
 - Prior years What does Samantha know?
 - Class comparison Is what Samantha knows adequate for the grade?
 - School comparison Is what Samantha knows enough for what comes next?
 - District comparison What should she know?
 - State comparison Is what Samantha knows enough?
 - Other state comparisons?
- What other information would we need to answer those questions?

COMPARABILITY TO DRIVE COMPARISONS



WHY DO WE CARE ABOUT COMPARABILITY?

- Comparability is the ability to make the same inferences from different observations
- Comparability is not important when examining **only** Samantha's test event, but is important when
 - Comparing her test to others
 - Aggregating her and other's tests
 - Comparing tests across time
- The more standardized the testing conditions, the more comparable.
- The more flexible the testing conditions, the less comparable

HOW COMPARABLE CAN WE MAKE IT?

- Can we make the same inferences across different observations?
- The level of comparability can range depending on the data we to which we have access
 - Interchangeable scores
 - Average scores
 - Claims across tests

HOW COMPARABLE CAN WE MAKE IT?

- How can we compare tests?
 - We need to use
 - The same items across students or
 - The same students taking multiple items
- Within-state comparisons:
 - OK tests use same items within grades (with newer items replacing poorly performing ones over time)
 - Can facilitate normative comparisons
- Across-state comparisons can be more difficult

HOW COMPARABLE CAN WE MAKE IT?

- There are many ways to compare OK student performance to other state's student performance
 - Embedding items from other tests (e.g., multi-state assessments)
 - Students take two tests (OK assessment and ACT)
 - Include nationally-normed tests like ITBS, CAT, MAP
 - Using NAEP comparisons
 - Two-stage link (e.g., Lexiles and Quantiles)
 - If $A \rightarrow B$, $B \rightarrow C$, then $A \rightarrow C$

HOW COMPARABLE CAN WE MAKE IT?

What are the limitations?

- There are many ways to compare OK student performance to other state's student performance
 - Embedding items from other tests (e.g., multi-state assessments)
 - Students take two tests (OK assessment and ACT)
 - Include nationally-normed tests like ITBS, CAT, MAP
 - Using NAEP comparisons
 - Two-stage link (e.g., Lexiles and Quantiles)
 - If $A \rightarrow B$, $B \rightarrow C$, then $A \rightarrow C$

HOW COMPARABLE CAN WE MAKE IT?

What are the limitations?

- There are many ways to compare OK student performance to other state's student performance
 - Embedding items from other tests (e.g., multi-state assessments) ← Increased Test Length
 - Students take two tests (OK assessment and ACT) ← Additional Tests
 - Include nationally-normed tests like ITBS, CAT, MAP ← Increased Test Length
 - Using NAEP comparisons ← Already completed
 - Two-stage link (e.g., Lexiles and Quantiles) ← Additional Tests
 - If $A \rightarrow B$, $B \rightarrow C$, then $A \rightarrow C$

HOW COMPARABLE CAN WE MAKE IT?

- The previous slides raise important questions about Samantha's score
 - Comparisons to other grade 8 OK students ← assumed
 - Comparisons to grade 8 students in other states ← a source of work
- Consider the claim you want to make and to whom you want to compare those claims
- Note: All of these comparisons are subject to the same quality questions of alignment to standards, technical quality, inclusion of all students, and rigorous performance standards

CONSIDER THESE QUESTIONS

- How important is it that we maintain comparisons within the state?
- How important is it to make comparisons between OK student performance to the performance of students outside of the state?
- What comparisons could you live without for OK students?
- Are we willing to increase test length or have students take more than one test to make additional comparisons (note, this would not necessarily have to take place every year)?

TECHNICAL QUALITY OF AN ASSESSMENT

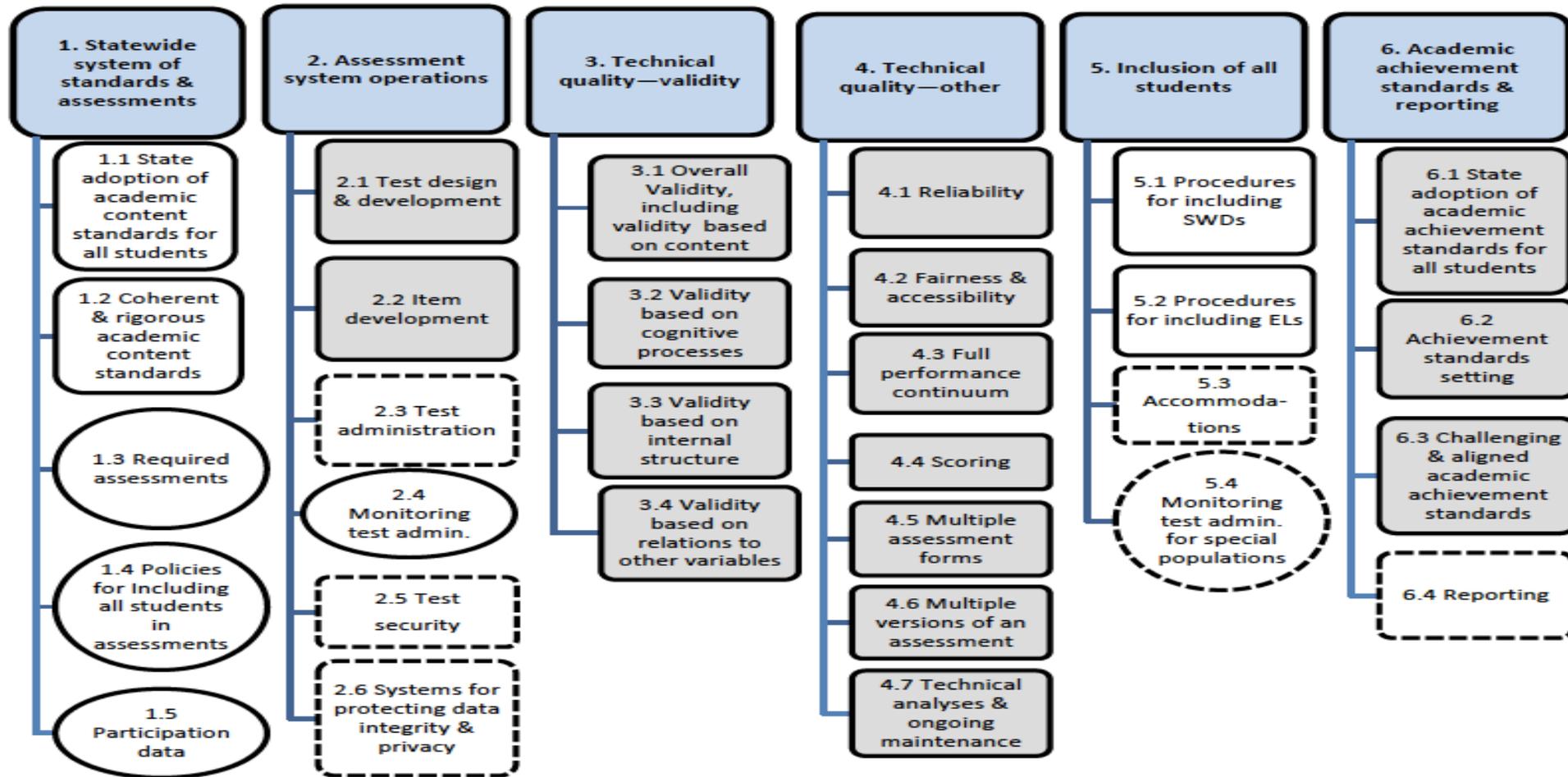
TECHNICAL QUALITY MUST ADDRESS REQUIREMENTS

- How do we ensure that the we are able to meet the necessary requirements:
 - House Bill 3218
 - Intended uses of measuring proficiency, growth, accountability
 - Supporting peer review

TECHNICAL QUALITY MUST ADDRESS REQUIREMENTS

- Peer Review helps to ensure that high stakes assessments are valid and reliable (i.e., high quality, relevant, and useful for their purposes)
 - Standards
 - Assessment system operations
 - Technical quality
 - Inclusion of all students
 - Rigorous achievement standards
 - Informative and understandable reporting

Map of the Critical Elements for the State Assessment System Peer Review



KEY

- Critical elements in ovals will be checked for completeness by Department staff; if necessary, they may also be reviewed by assessment peer reviewers (e.g., Critical Element 1.3). All other critical elements will be reviewed by assessment peer reviewers.
- Critical elements in shaded boxes likely will be addressed by coordinated evidence for all States administering the same assessments (e.g., Critical Element 2.1).
- Critical elements in clear boxes with solid outlines likely will be addressed with State-specific evidence, even if a State administers the same assessments administered by other States (e.g., Critical Element 5.1).
- Critical elements in ovals or clear boxes with dashed outlines likely will be addressed by both State-specific evidence and coordinated evidence for States administering the same assessments (e.g., Critical Element 2.3, 5.4).

TECHNICAL QUALITY: RELIABILITY

- Reliability refers to the consistency of results.
 - If you gave a student the same test 100 times (erasing their memory of the assessment in between administrations), how many times would they get the same score?
 - If you gave 100 scorers the same essay to score, how many different scores would you get?
- Accuracy (Decision Consistency) is the degree to which interpretations of scores are salient across categories and similar across test administrations (viewed as part of reliability)

TECHNICAL QUALITY: INCREASING RELIABILITY

- Assessment development is a series of compromises driven by intended uses
- Increase reliability by increasing the measures (i.e., lengthening tests)
 - That means more items
 - Consider, for example, if I asked you to add $54+79$. If you got it right, can I assume you have mastered adding two-digit numbers? If you get it wrong, can I assume you don't know how to add two-digit numbers? How many problems would you have to answer before I could comfortably say you've mastered the skill, you don't know it at all, or you have partial understanding?
- Also, increase reliability by testing the same construct multiple ways

TECHNICAL QUALITY: VALIDITY

- Validity is not a yes/no or on/off.
- Tests are not valid, score interpretations are. The interpretation is subject to enough evidence to support a validity claim.
- For a score to have a valid interpretation, it must accurately and reliably reflect a student's knowledge and skills.
- Can a test be reliable but not valid?
 - Yes, but this isn't good. Consider an archer who consistently shoots the arrow in the same spot but that spot is always a foot from the bullseye.
- Can a test be valid but not reliable?
 - No. That would be like saying if an arrow hit the bullseye once, the person is a master archer.

ITEM TYPES

- Multiple choice
- Multi-select multiple choice
- Technology enhanced
- Short constructed response
- Extended constructed response

What do we learn from each type and when would we use each?

SAMPLE MC ITEM

Consider the four diagrams shown below. In which of the following diagrams, is one quarter of the area shaded?

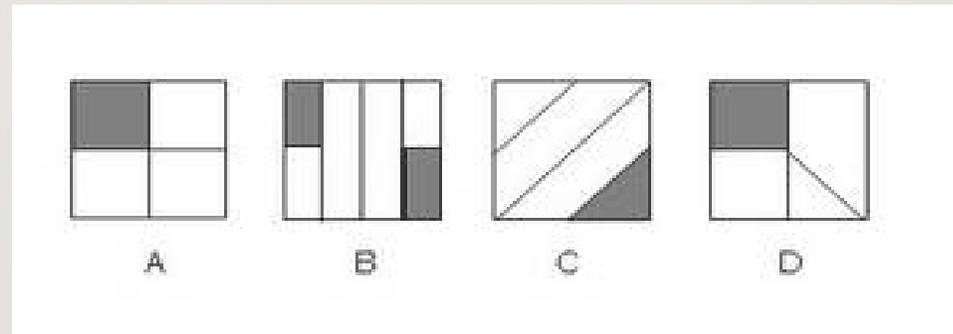


Diagram A is the obvious answer, but B is also correct. However, some students do not believe that one quarter of B is shaded because of a belief that the shaded parts have to be contiguous. Students who believe that one quarter of C is shaded have not understood that one region shaded out of four is not necessarily a quarter. Diagram D is perhaps the most interesting here. One quarter of this diagram is shaded, although the pieces are not all equal; students who rely too literally on the "equal areas" definition of fractions will say that D is not a correct response.

SAMPLE MSMC WITH SHORT CONSTRUCTED RESPONSE

Can It Reflect Light?

What types of objects or materials can reflect light? Put an X next to the things you think can reflect light.

- | | |
|---|--|
| <input type="checkbox"/> water | <input type="checkbox"/> red apple |
| <input type="checkbox"/> gray rock | <input type="checkbox"/> rough cardboard |
| <input type="checkbox"/> leaf | <input type="checkbox"/> the Moon |
| <input type="checkbox"/> mirror | <input type="checkbox"/> rusty nail |
| <input type="checkbox"/> glass | <input type="checkbox"/> clouds |
| <input type="checkbox"/> sand | <input type="checkbox"/> soil |
| <input type="checkbox"/> potato skin | <input type="checkbox"/> wood |
| <input type="checkbox"/> wax paper | <input type="checkbox"/> milk |
| <input type="checkbox"/> tomato soup | <input type="checkbox"/> bedsheet |
| <input type="checkbox"/> crumpled paper | <input type="checkbox"/> brand new penny |
| <input type="checkbox"/> shiny metal | <input type="checkbox"/> old tarnished penny |
| <input type="checkbox"/> dull metal | <input type="checkbox"/> smooth sheet of aluminum foil |

Explain your thinking. Describe the “rule” or the reasoning you used to decide if something can reflect light.



SAMPLE SHORT CONSTRUCTED RESPONSE

Mr. Ruiz is starting a marching band at his school. He first does research and finds the following data about other local marching bands.

	Band 1	Band 2	Band 3
Number of Brass Instrument Players	123	42	150
Number of Percussion Instrument Players	41	14	50

Enter your answer in the box.

Mr. Ruiz realizes there are brass instrument player(s) per percussion player

LABELING (ONE TO ONE): DRAG AND DROP

Select the x value that makes each equation true.

$x = -3$

$x = -2$

$x = -4$

$x = -1$

drop correct response here $x^2 - x - 6 = 0$

drop correct response here $3x^2 - 12x - 15 = 0$

drop correct response here $6x^2 - 6x - 72 = 0$

drop correct response here $6x^2 + 18x - 24 = 0$

CATEGORIZATION: DRAG AND DROP

Will is rewriting a report about going to the zoo. He needs to use facts instead of opinions. Read the sentences from the report and sort each sentence to show whether it is a fact or an opinion.

Sentences



There are over 300 animals at the zoo.

The zoo is the best place to go for bird-watching.

People can ride a train to tour the zoo.

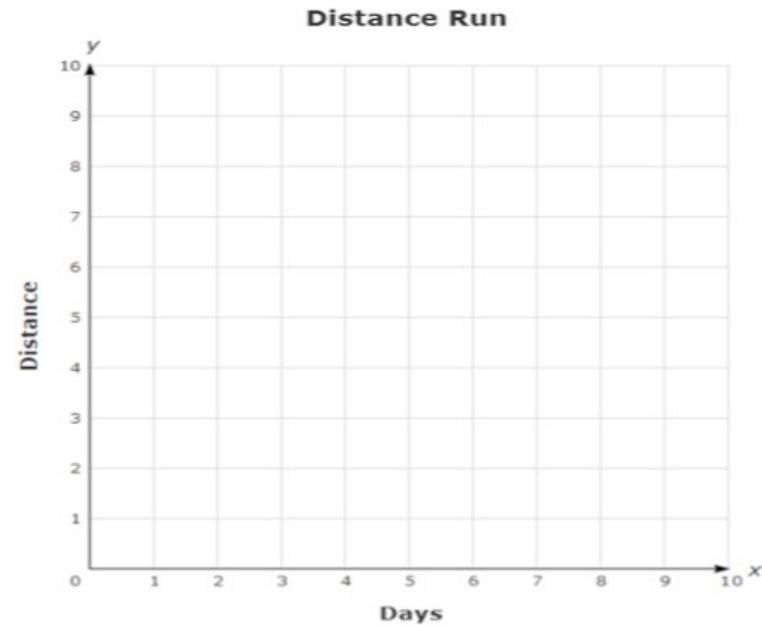
Riding the train is a fun activity for everyone.

Fact

Opinion

Graphing

Grant runs 3 miles each day. Show the graph of how far Grant runs over time.



EXTENDED CONSTRUCTED RESPONSE

You have read a website entry and an article, and viewed a video describing Amelia Earhart. All three include information that supports the claim that Earhart was a brave, courageous person.

The three titles are:

- “The Biography of Amelia Earhart”
- “Earhart’s Final Resting Place Believed Found”
- “Amelia Earhart’s Life and Disappearance” (video)

Consider the argument each author uses to demonstrate Earhart’s bravery.

Write an essay that analyzes the strength of the arguments related to Earhart’s bravery in at least two of the three supporting materials. Remember to use textual evidence to support your ideas.

CHOOSING AN ITEM TYPE

- What are you trying to measure?
 - Consider “identify” versus “create” or “interpret” versus “graph”
- What resources are available for hand scoring?
- Will the assessment be given solely on computer or split between computer and paper/pencil?

PUTTING IT ALL TOGETHER: DESIGNING AN ASSESSMENT

- Recommend backwards design
- Start with considering what you want the score report to include
 - Scale score and performance category
 - How many performance categories?
 - What are the distinctions?
 - Sub scores
 - How many?
 - What are the important reporting categories?

RULES OF THUMB

- Performance levels
 - The more levels you want, the more items you need
 - There need to be true distinctions across levels, keeping measurement error in mind
- Subscores
 - You need a minimum of six items per reporting category
 - Consider replication versus full alignment to category
- Reliability
 - More items = higher reliability
 - 50 items will typically get you close to 0.9 reliability
 - 20 items is closer to 0.6 reliability

DEVELOP A BLUEPRINT

- Cover all standards at level intended
 - Consider entry steps and extensions
- What evidence do you want to see to show a student has mastered a standard?
- What relative emphasis do you want to have on clusters of standards?
- Can standards be meaningfully combined or should they be measured individually?

SAMPLE EVIDENCE STATEMENT

Target 6: TEXT STRUCTURES & FEATURES	Relate knowledge of text structures, genre-specific features, or formats (visual/graphic/auditory effects) to obtain, interpret, explain, or connect information within text.
Evidence Required	<ol style="list-style-type: none">1. The student will determine how the overall structure of a text impacts its meaning.2. The student will analyze or interpret why the author structured elements within the text in a certain manner and the impact of that structure on meaning.
Standards	RL-5; RL-7

SAMPLE BLUEPRINT

Claim (% of Test)	Focus	Target	Goal DOK	Relative Emphasis
1. Reading (60-65%)	Literary Texts	1: Key Details	2	Low
		2: Central Ideas	2	High
		3: Word Meanings	2	Low
		4: Reasoning & Evidence	3	High
		5: Analysis Within Or Across Texts	3	Low
		6: Text Structures & Features	3	
		7: Language Use	3	
	Informational Texts	8: Key Details	2	Medium
		9: Central Ideas	2	High
		10: Word Meanings	2	Medium
		11: Reasoning & Evidence	3	High
		12: Analysis Within Or Across Texts	3	Low
		13: Text Structures & Features	3	
		14: Language Use	3	
2. Writing (25-30%)	Write / Revise	1/3/6: Write / Revise Brief Texts	2	High
	Language / Vocabulary	8: Language & Vocabulary Use	2	High
	Conventions	9: Edit	1	High
3. Listening (10-15%)	Listen	4: Listen / Interpret	3	High

WRITE PERFORMANCE LEVEL DESCRIPTORS

- How many?
- What level of rigor?
- Key components of standards within each grade

WRITE ITEMS

- Match items to blueprint
- Align items with standards
- Ensure level of rigor of items matches standards (e.g., if standard says to evaluate how an author uses specific elements to convey purpose, an item asking to identify the purpose is insufficient)
- Use universal design to ensure accessibility
- Multiple item reviews
 - Content
 - Bias/sensitivity
 - Accessibility
 - Usability
 - Editorial

ACCESSIBILITY

- Consider all item features that could impact accessibility
 - Is language necessary to measure the concept?
 - Exposure to format and English language learners
 - Drag and drop for students with orthopedic and visual disabilities
 - Video/audio presentation for students who are blind or deaf
 - Words or graphics that do not translate well into Braille
- Focus groups, item trials, and expert review help ensure an accessible test

ALLOW TIME

- Typically it takes two years to develop a strong assessment from scratch
 1. Lay out reports
 2. Develop blueprints
 3. Create item specifications
 4. Write PLDs
 5. Develop items
 6. Review items
 7. Build forms
 8. Field test items
 9. Analyze results
 10. Validate alignment with standards and score interpretation

BACK TO THE BALANCED ASSESSMENT SYSTEM

Formative Tools

Based on learning theory
Minute by minute between teacher and student
Includes instructional resources to build student learning
Not intended for aggregation or teacher/program evaluation

Interim Assessment

Optional
District choice
Diagnostic information
Tracks growth
Predicts summative
Can be aggregated at classroom or building level

Summative Assessment

End of year
Can be used as a snapshot within and across schools and districts
ESSA eliminated punitive consequences
Information & transparency
Examine equity and resource allocation

All based on Oklahoma Standards and Goals for Students

QUESTIONS?



DISCUSSION AND RECOMMENDATIONS



ESSA & HB 3218

Accountability Requirements & Transition Timeline

Dr. Katie Dunlap

Deputy Superintendent of Assessment & Accountability

Dr. Michael Tamborski

Executive Director of Accountability

OKLAHOMA STATE DEPARTMENT OF
EDUCATION
— CHAMPION EXCELLENCE —

HB 3218 - Accountability

- Directs SBE to develop a new school accountability plan by January 1, 2017 (included with assessment requirements).
- Must comply with ESSA and include the following indicators:
 - Student performance on statewide assessments
 - Graduation rates for high schools
 - An additional academic indicator for elementary and middle schools
 - English language proficiency for English learners
 - At least one indicator of school quality or student success



HB 3218 - Accountability

- HB 3218 requires that the 2016-17 report cards include an explanation of the changes to the state assessments and how the transition may impact school performance.
- No other changes to existing A – F statute.

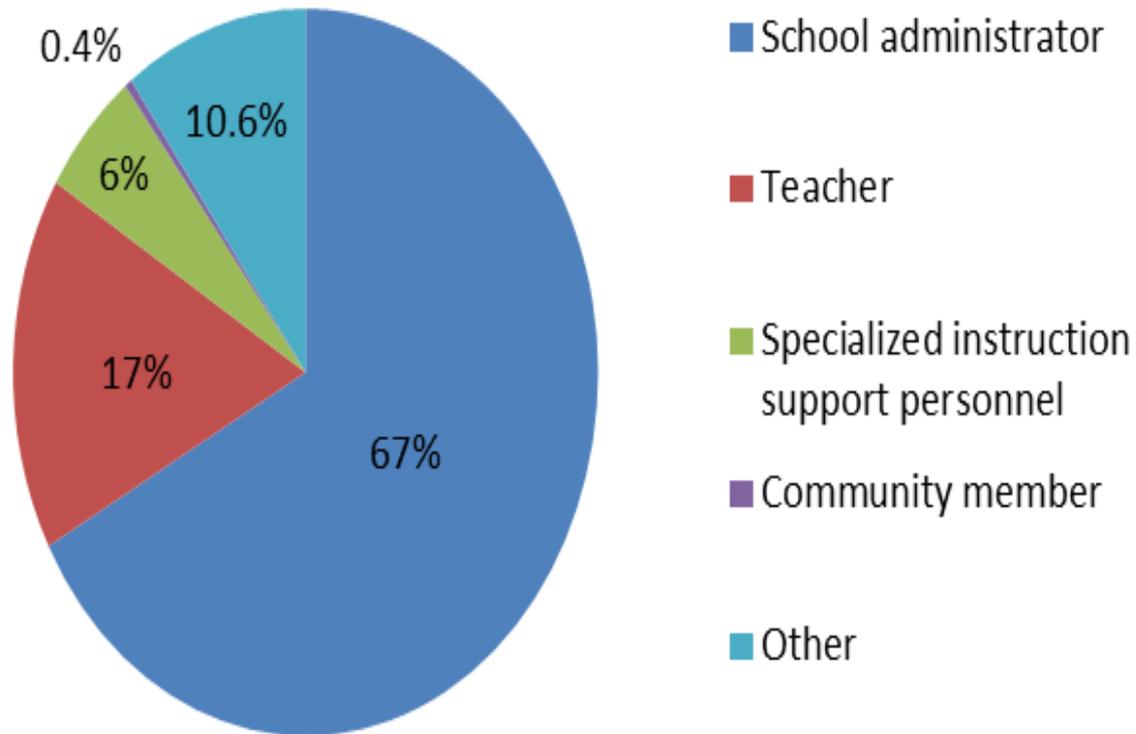


Engage OK Poll Everywhere Results



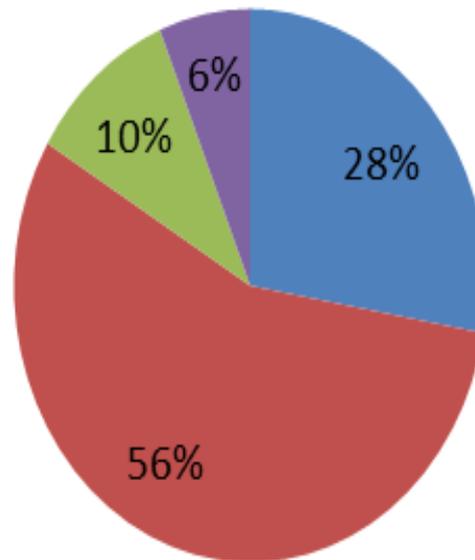
OKLAHOMA STATE DEPARTMENT OF
EDUCATION
— CHAMPION EXCELLENCE —

In what role are you attending this conference?



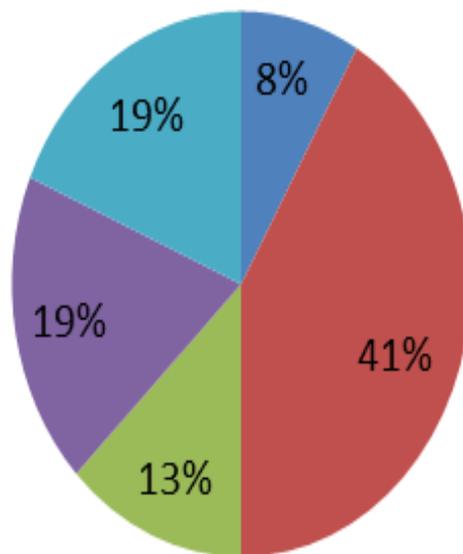
What community type are you a representative for?

■ Suburban ■ Rural ■ Urban ■ Other



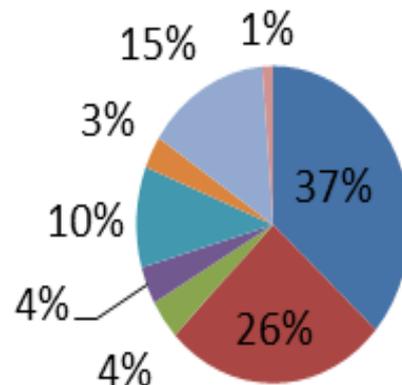
Which part of Oklahoma are you a representative for?

■ Northwest ■ Northeast ■ Southwest ■ Southeast ■ Central

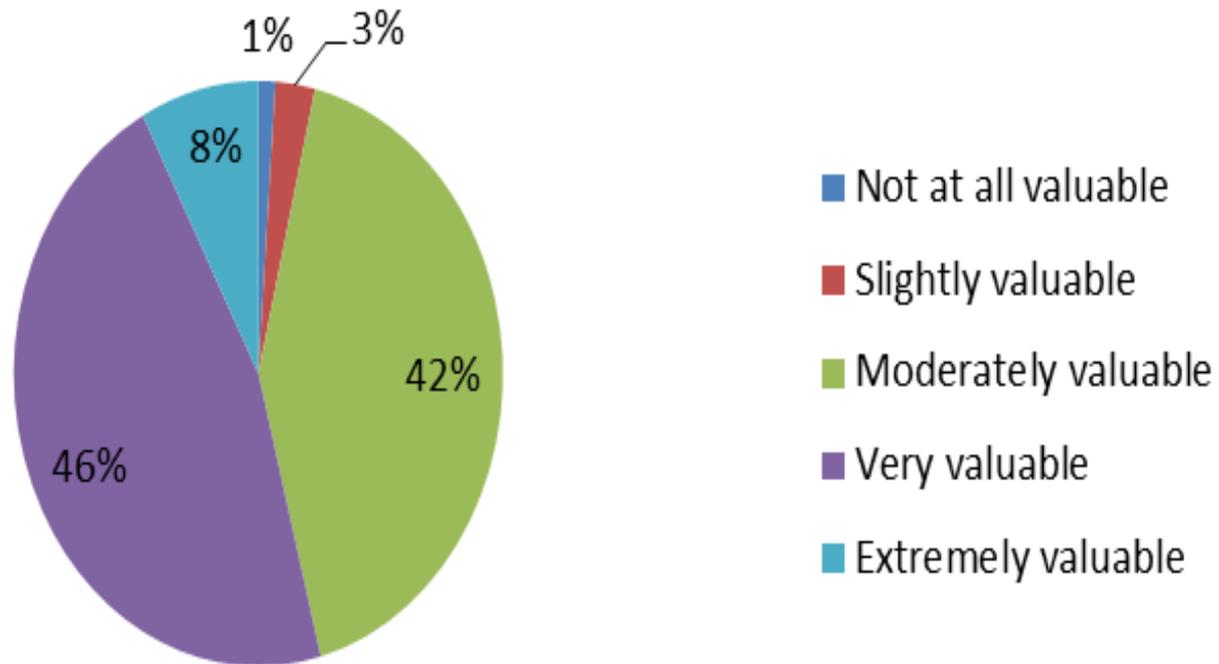


What group(s) are you a representative for? (select all that apply)

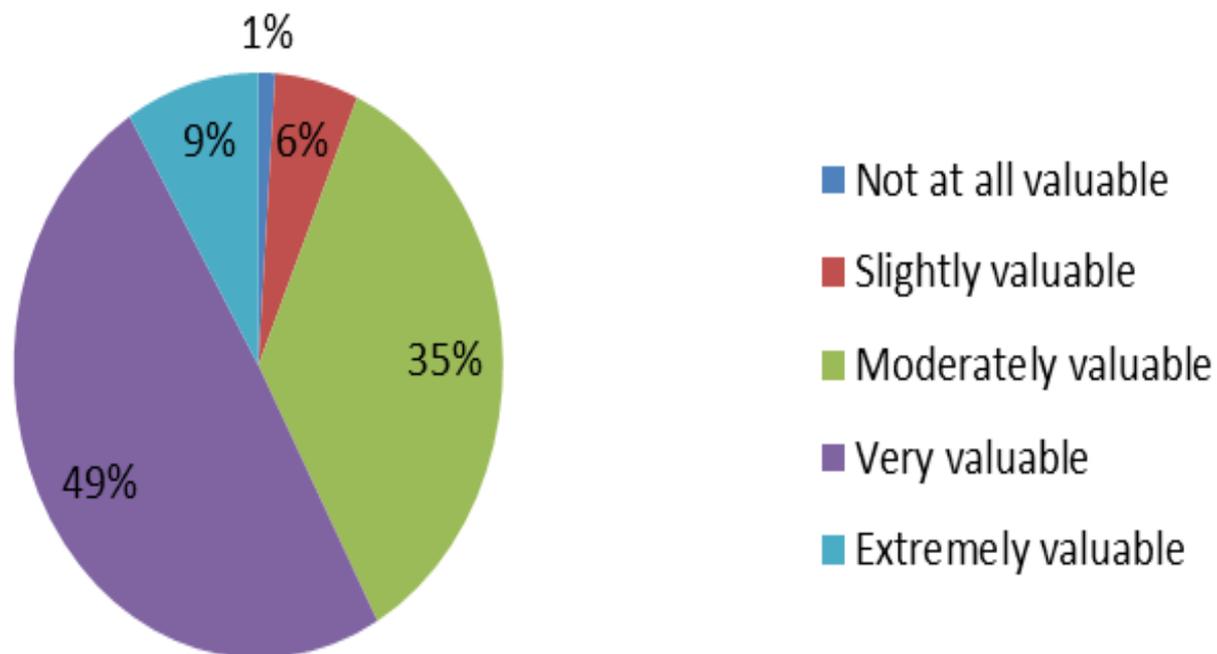
- Students with disabilities (IEP/504)
- English Language Learners (ELL)
- Higher Ed
- Career Technology Center
- Indian Tribe(s)
- Charter Schools
- Early Education Organizations
- Licensing Organizations



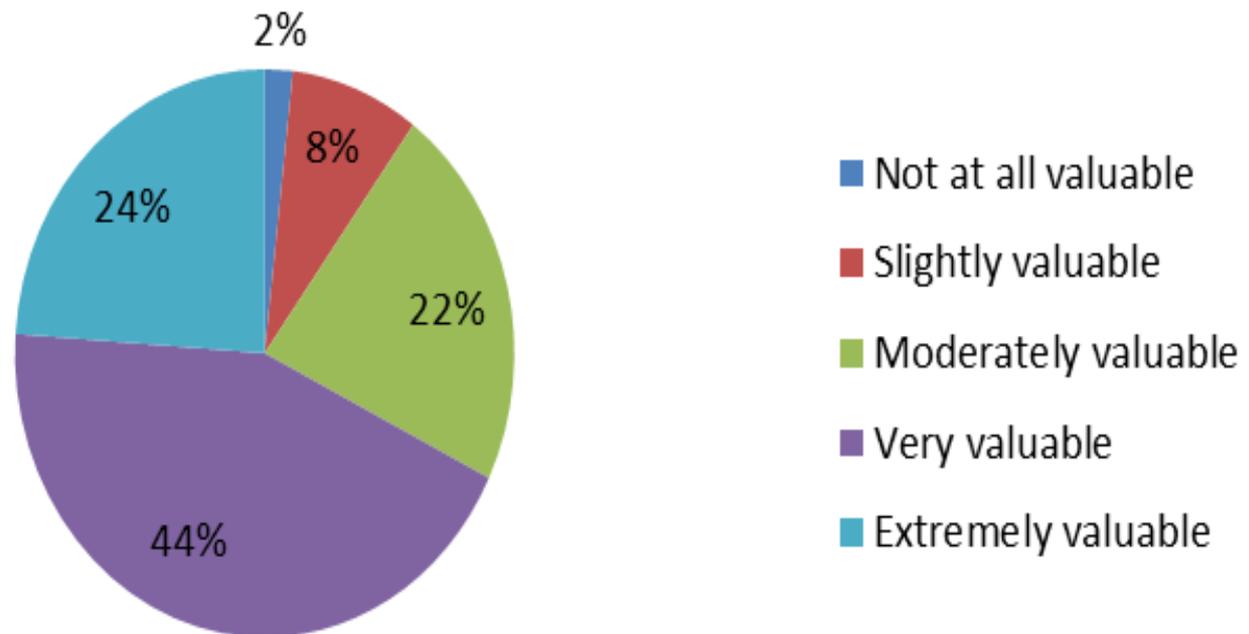
How valuable is mastery of Oklahoma content standards in indicating student progress toward college/career readiness?



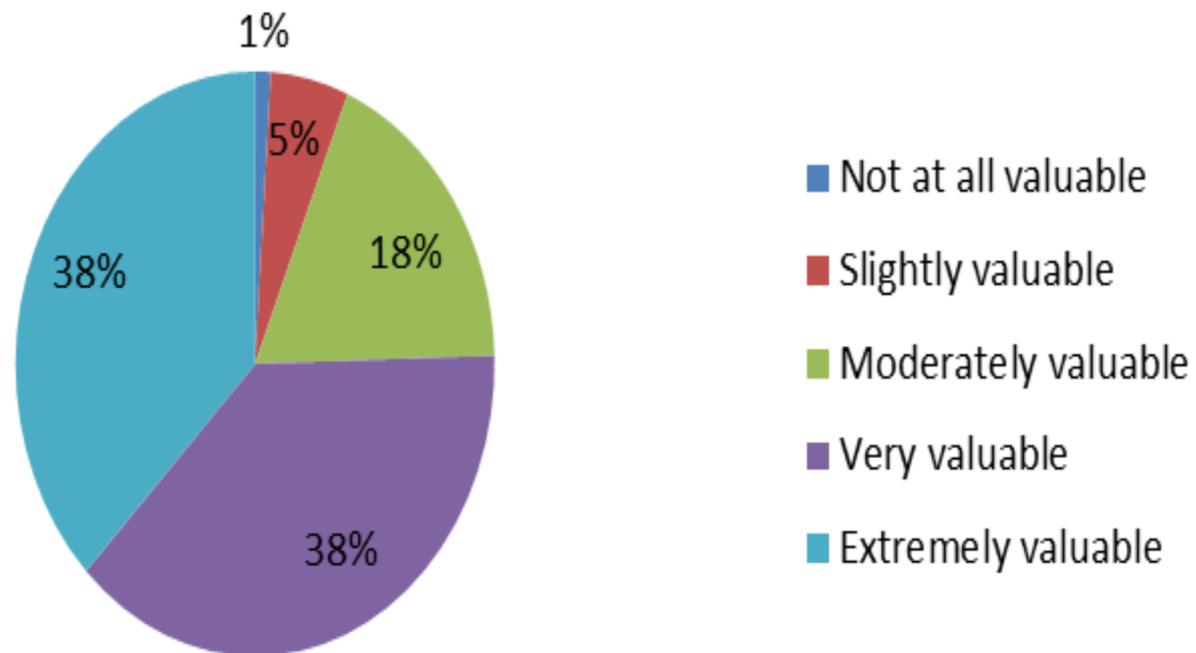
How valuable is growth in mastery of Oklahoma content standards in indicating student progress toward college/career readiness?



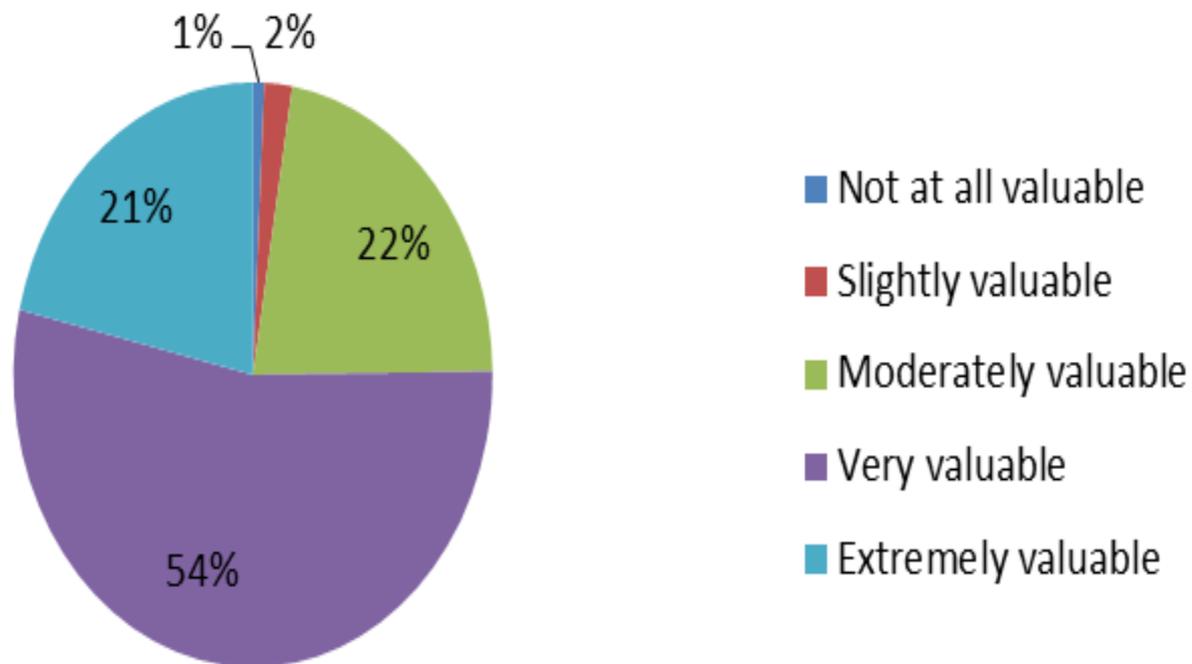
How valuable is graduation with a High School Diploma in indicating student progress toward college/career readiness?



How valuable is student attendance in indicating student progress toward college/career readiness?



How valuable is progress towards English Language Proficiency in indicating student progress toward college/career readiness?



Summary Data

1 = Not at all valuable

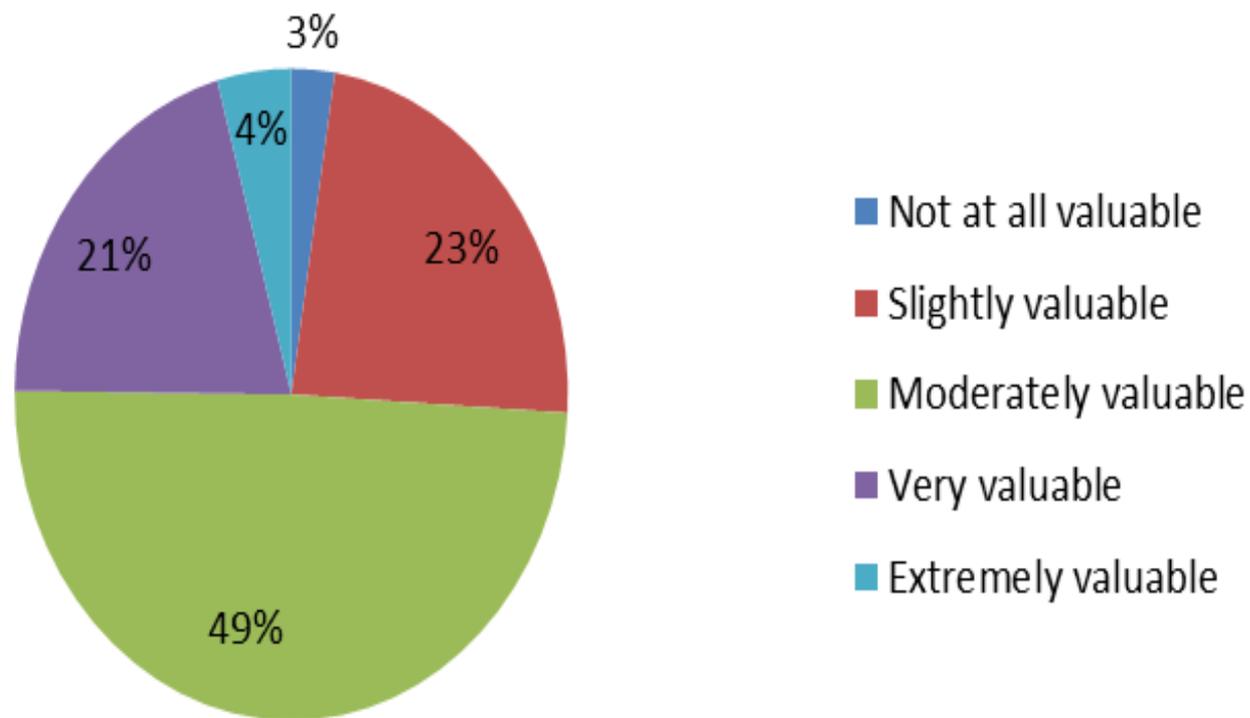
5 = Extremely valuable

	Mean	Standard Deviation
Mastery of Content Standards	3.56	.721
Growth in Mastery of Content Standards	3.58	.772
Graduation with a Diploma	3.78	.959
Attendance	4.03	.934
Progress in English Language Proficiency	3.91	.765

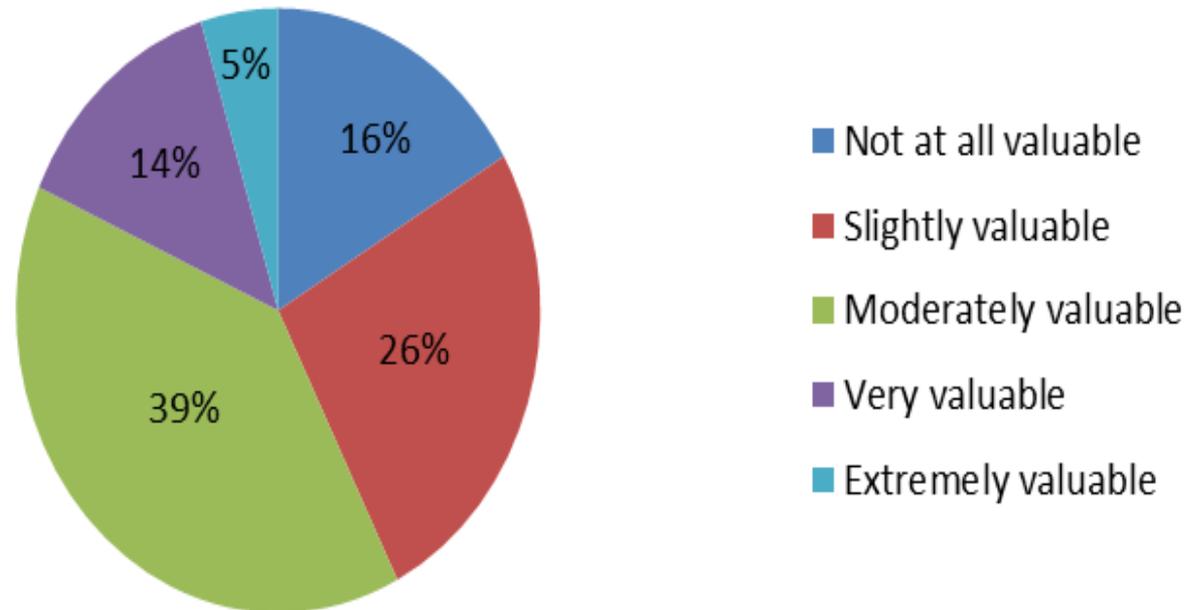


OKLAHOMA STATE DEPARTMENT OF
EDUCATION
CHAMPION EXCELLENCE

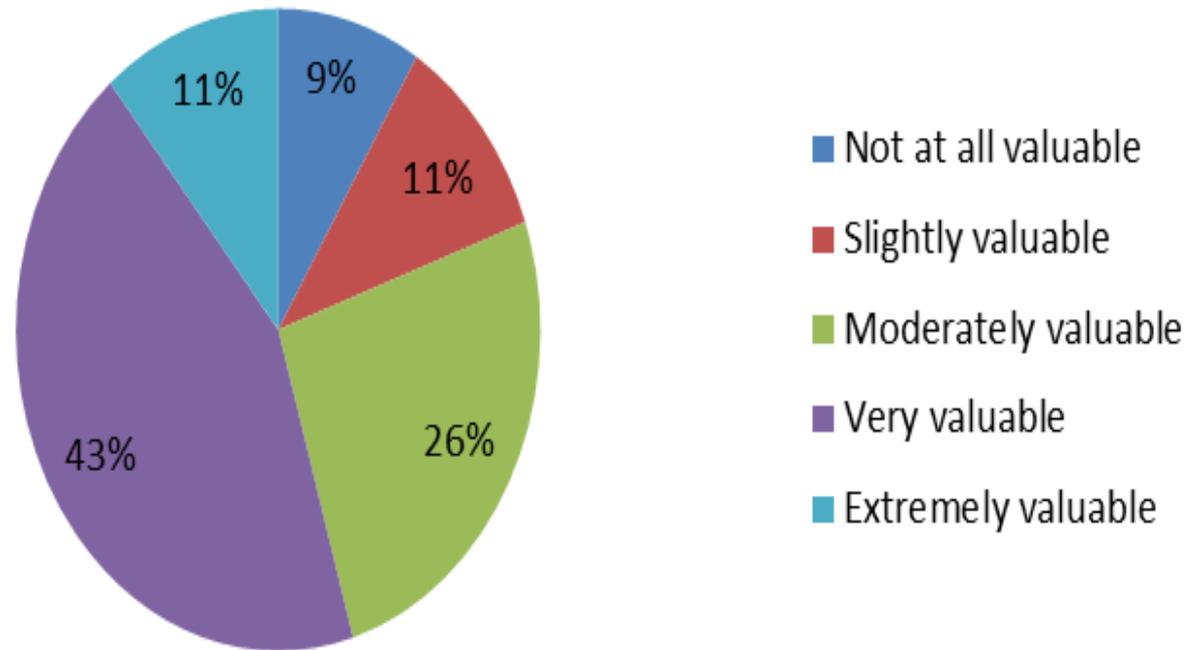
How valuable is the completion of advanced coursework in indicating school quality?



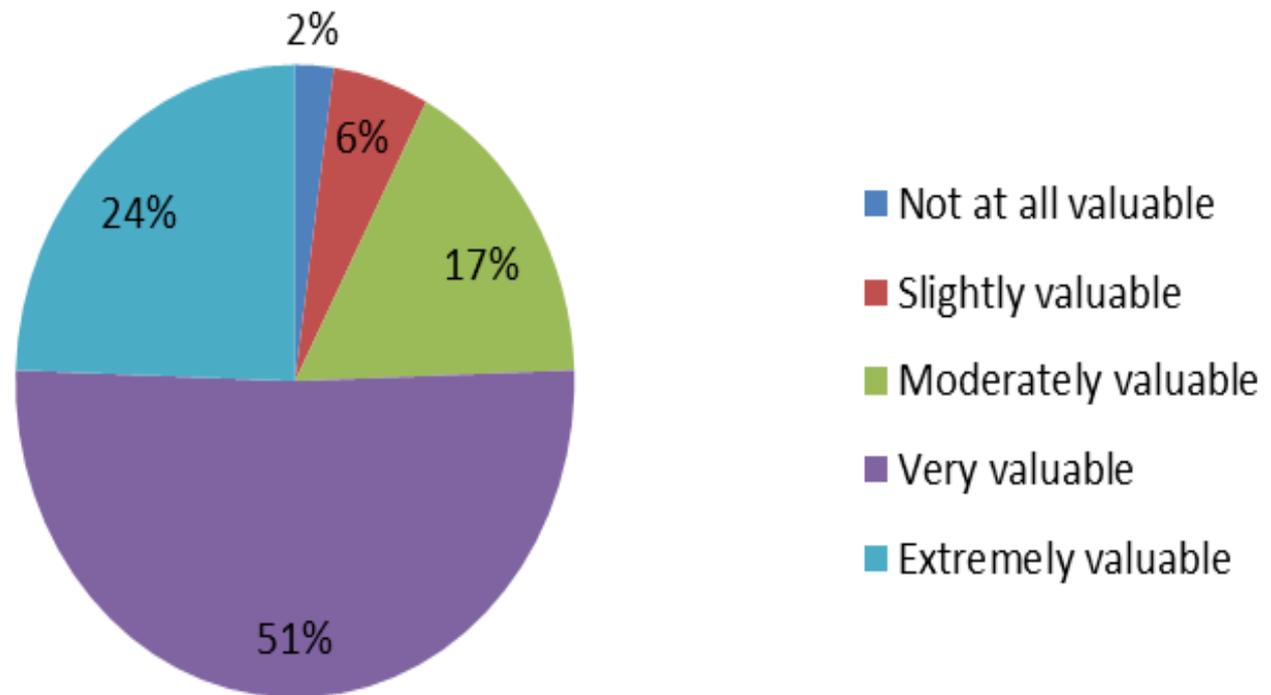
How valuable is a school's accreditation report in indicating school quality?



How valuable is chronic absenteeism in indicating school quality?



How valuable is school safety in indicating school quality?



Summary Data

1 = Not at all valuable

5 = Extremely valuable

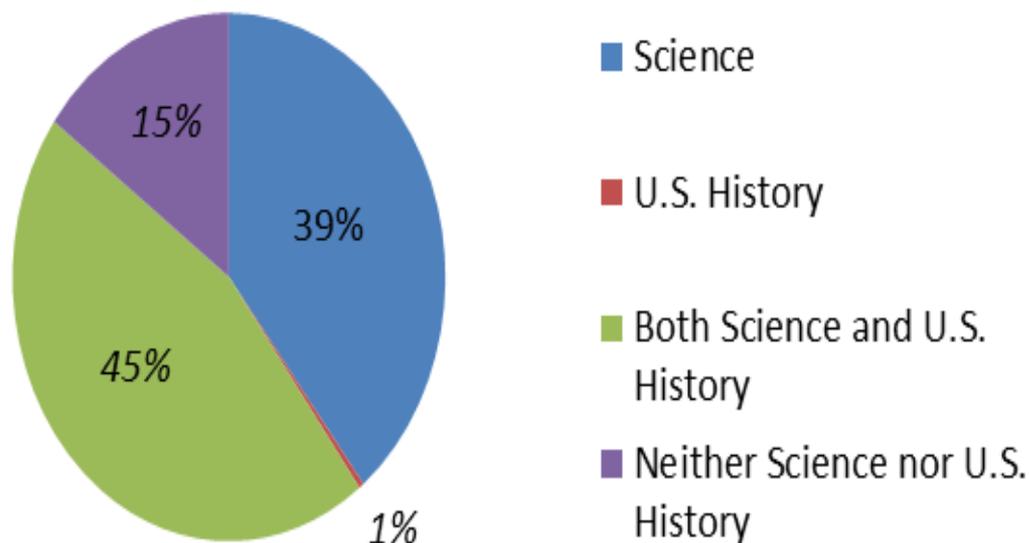
:

	Mean	Standard Deviation
Advanced Coursework	3.00	.845
Accreditation Report	2.62	1.050
Chronic Absenteeism	3.39	1.102
School Safety	3.92	.914

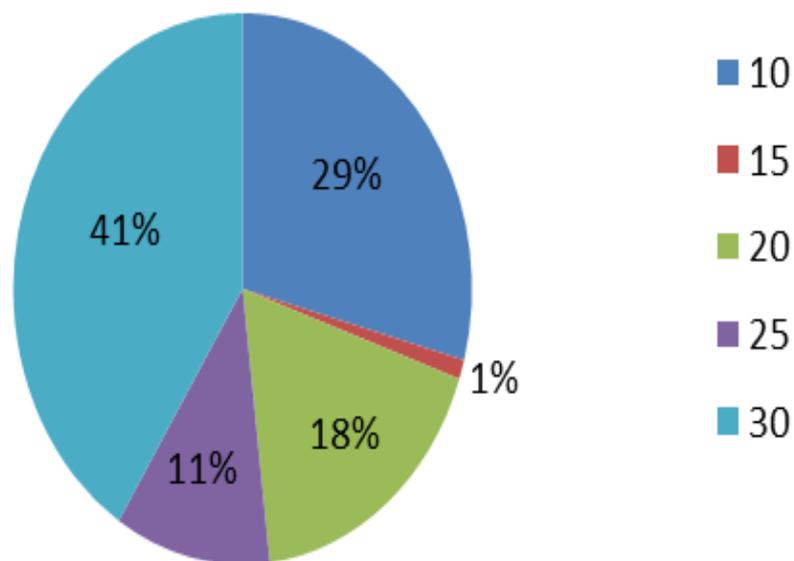


OKLAHOMA STATE DEPARTMENT OF
EDUCATION
— CHAMPION EXCELLENCE —

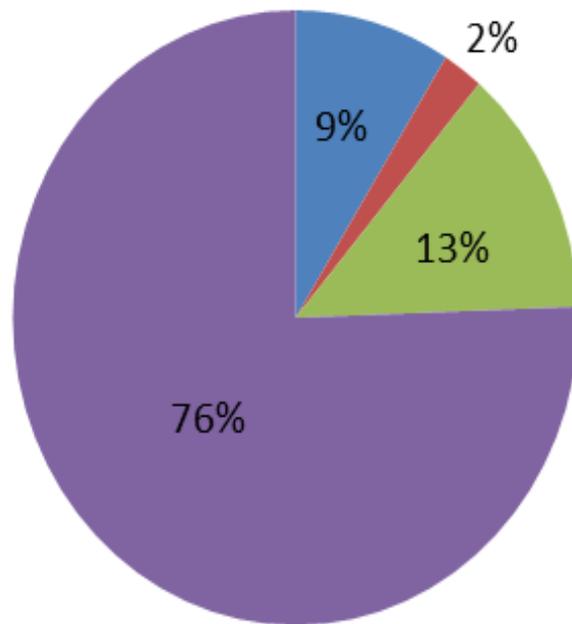
**In addition to Mathematics and Reading/Language Arts,
which of the following assessments do you think would
add value and help predict student progress toward
college/career readiness?**



What is the minimum number of students needed to generate usable data on student academic performance?

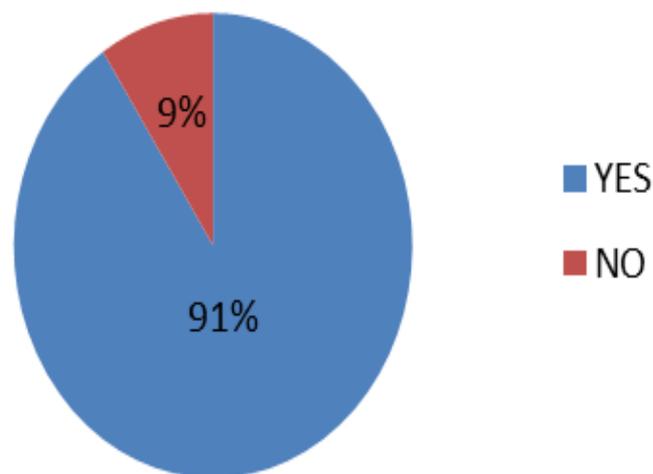


How should not meeting the 95% requirement at a school be reflected in their report card?

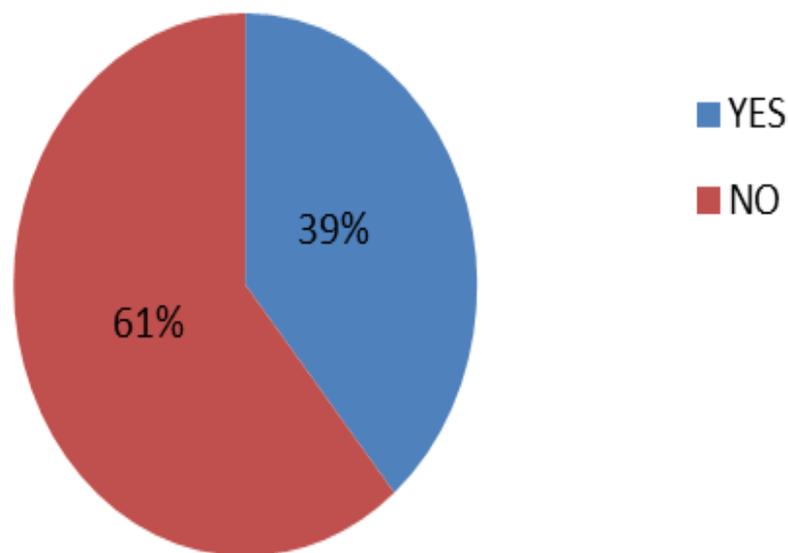


- The report card should be lowered by one letter grade
- The report card should display an F on the performance indicator
- The school should be designated as requiring Comprehensive Support
- Some other option (requires USDE approval)

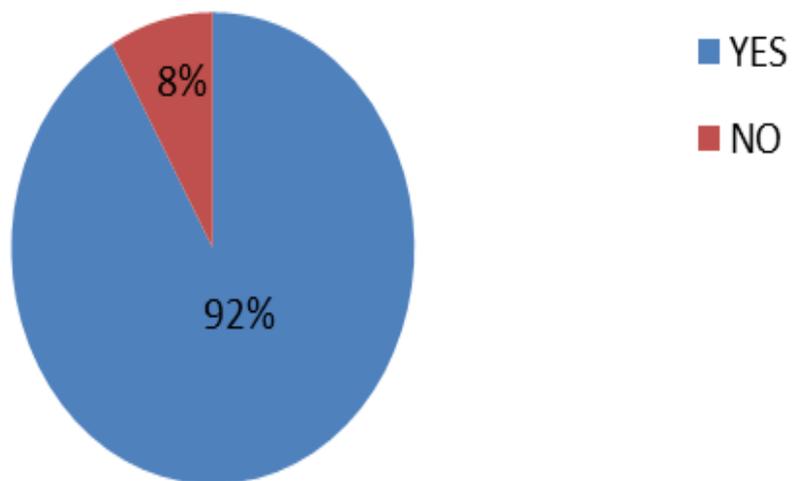
Would averaging data over up to three years for schools that do not meet the minimum number increase the reliability and accuracy of the report cards?



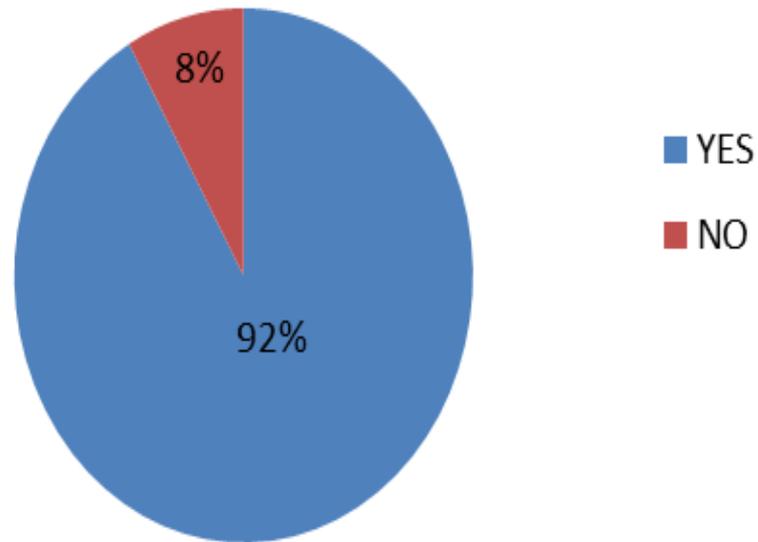
Should there be a different set of accountability requirements for schools in their first year of operation?



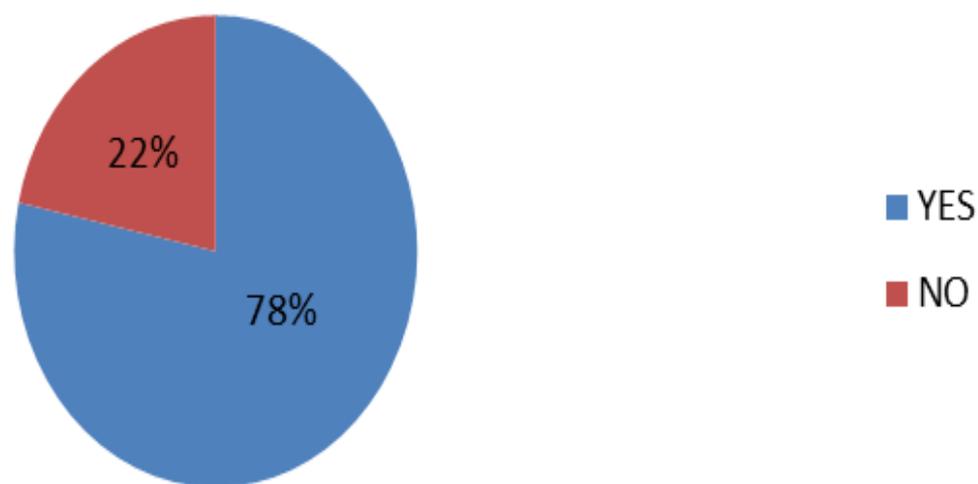
Should there be a different set of accountability requirements for schools that serve only at-risk students (e.g., Traice in Tulsa or Emerson in OKC)?



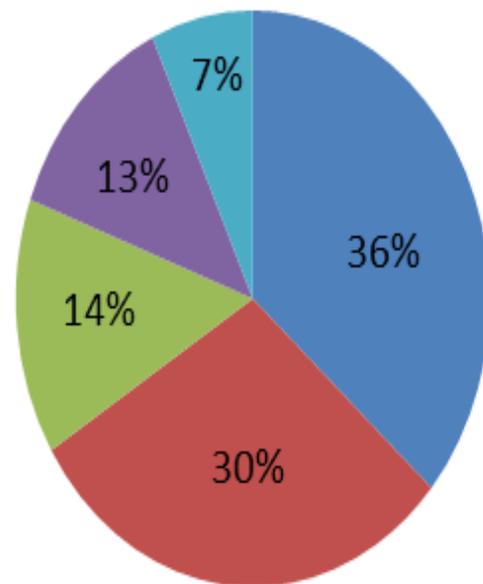
Should there be a different set of accountability requirements for schools serving grades that do not require testing (e.g., PK-2)?



Should there be a different set of accountability requirements for schools that do not meet the minimum number of students required to calculate a schools accountability report card?



Which of the following should be used to indicate school quality for elementary or middle schools (choose all that apply)?



■ Early Warning Indicators

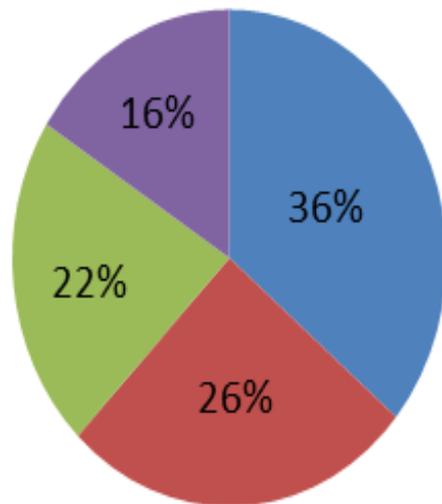
■ Chronic Absenteeism

■ Drop-out rates

■ Completion of Advanced Coursework

■ Identification of a Career Pathway

Which of the following should be used to indicate school quality for high schools (choose all that apply)?



■ College Entrance Exams

■ Industry Certification Courses

■ Career Oriented Assessments

■ 9th Grade On-Track Indicator

Accountability Timeline

- A to F report cards are produced as required by state law (Fall 2016)
- OSDE will engage with advisory and stakeholder groups (like you!) to develop a new accountability plan (Now – December 2016)
- New accountability plan approved by SBE and sent to state legislature (December 2016)
- Technical requirements of the new plan are implemented (Beginning of 2017 – Fall 2017)
- Preliminary report cards are released based on the new accountability plan (Fall 2017)
- Final report cards will be publically released (No later than December 31, 2017)

OKLAHOMA ASSESSMENT TASK FORCE MEETING

THE EVERY STUDENT SUCCEEDS ACT OVERVIEW



STANDARDS I I I I (B)(I)

- **Content Standards**
 - Must be in R/ELA, Mathematics and Science
 - Must align with entrance requirements for credit-bearing coursework at technical or higher education institutions
- **ELP Standards**
 - Must be derived from the four recognized domains
 - Must align (correspond) to content standards
- **Alternate Standards**
 - Only alternate achievement standards allowed, not alternate content or ELP standards

ASSESSMENTS I I I I (B)(2)

- Content Assessments (continued)
 - Allow for disaggregation by Race/Ethnicity, SES, Disability, English proficiency status, gender, and migrant status
 - Adaptive assessments allowable
 - Must appropriately assess EL's and SWD
 - Can count ELs subgroup 4-years after exit
- ELP Assessments
 - Provide an annual assessment of ELP for all ELs in the state
 - Must be aligned to ELP standards

ASSESSMENTS I I I I (B)(2)

- Content Assessments

- Assessed in grades 3-8, HS for Reading or ELA and Mathematics
- Assessed once in elementary, middle and high school for science
- States allowed to use annual summative or multiple interim assessments
- States given two options for recently arrived ELs in assessing Reading or ELA
 - Exclude year 1 administration and use proficiency scores for year 2 accountability
 - Include year 1 administration, use growth in year 2 accountability, and use proficiency in year 3 accountability

ACCOUNTABILITY I I I I (C)

States must

- Declare minimum cell size for all groups
- Establish long-term and interim goals for each subgroup
 - Content proficiency performance
 - 4-year cohort adjusted graduation rates
 - Increases in the percentage of EL students making progress in English proficiency

ACCOUNTABILITY I I I I (C)

States shall for all public schools

- Use these indicators in their accountability system for all students and all subgroups:
 - **Proficiency** on the academic achievement assessments
 - **Growth or another differentiating indicator** differentiating student performance in elementary, middle and high school (high school optional)
 - **Graduation rate** (i.e., 4-year cohort adjusted)
 - **Progress in achieving English proficiency**

ACCOUNTABILITY I I I I (C)

States shall for all public schools

- Use not less than one of these indicators in their accountability system for all students and all subgroups that meaningfully differentiates schools:
 - Student engagement
 - Educator engagement
 - Student completion of advanced coursework
 - Post-secondary readiness
 - School climate & safety
 - Any other indicator

ACCOUNTABILITY I I I I (C) – ADDITIONAL INDICATOR REQUIREMENTS (IN NPRM S.200.14)

- Is valid, reliable, and comparable across all LEAs
- Is calculated in the same way for all schools across the State
- Is able to be disaggregated for each subgroup of students
- Is used no more than once in its system of annual meaningful differentiation
- Is supported by research that performance or progress on such measures is likely to increase student achievement or, for measures within indicators at the high school level, graduation rates.

ACCOUNTABILITY I I I I (C)

States shall for all public schools

- An Annual Meaningful Difference (AMD) system
 - Afford substantial weights to indicators
 - Weighting the required indicators more than the optional indicators
- Based on AMDs in SY 2017-2018 **
 - Establish categories of school support & improvement
 - Create a category for the lowest 5% of schools in content & ELP performance, schools graduating less than 1/3 of students
 - Identify schools consistently not meeting interim goals for comprehensive improvement

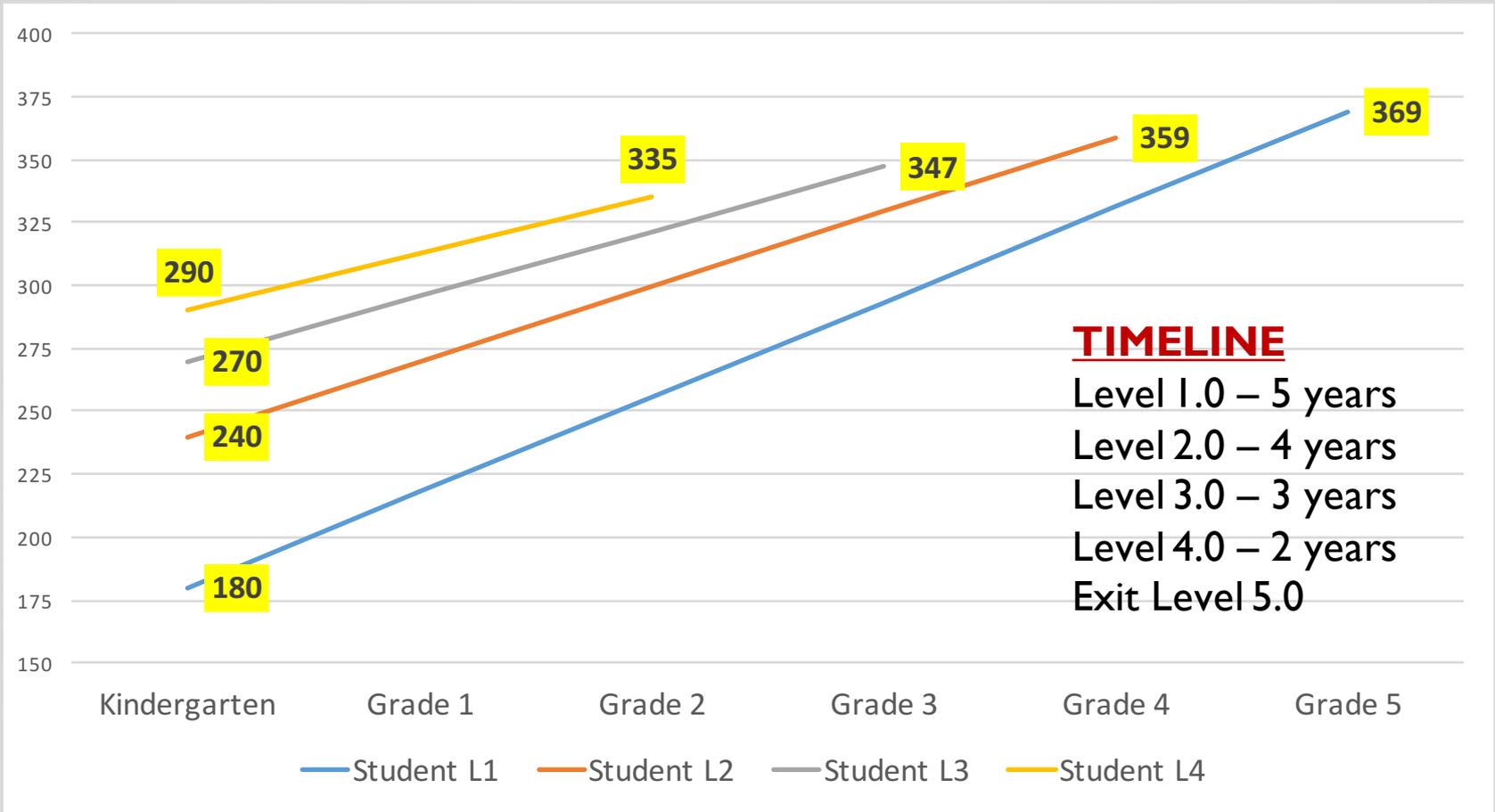
POSSIBILITIES FOR THE EL INDICATOR



GROWTH TO TARGET SCORES

- A potential method for complying with the EL accountability indicator requirements
- Growth to target scores are a calculation of where an EL student starts (initial scale score) in their language proficiency and where they should be (proficient scales score) in X number of years

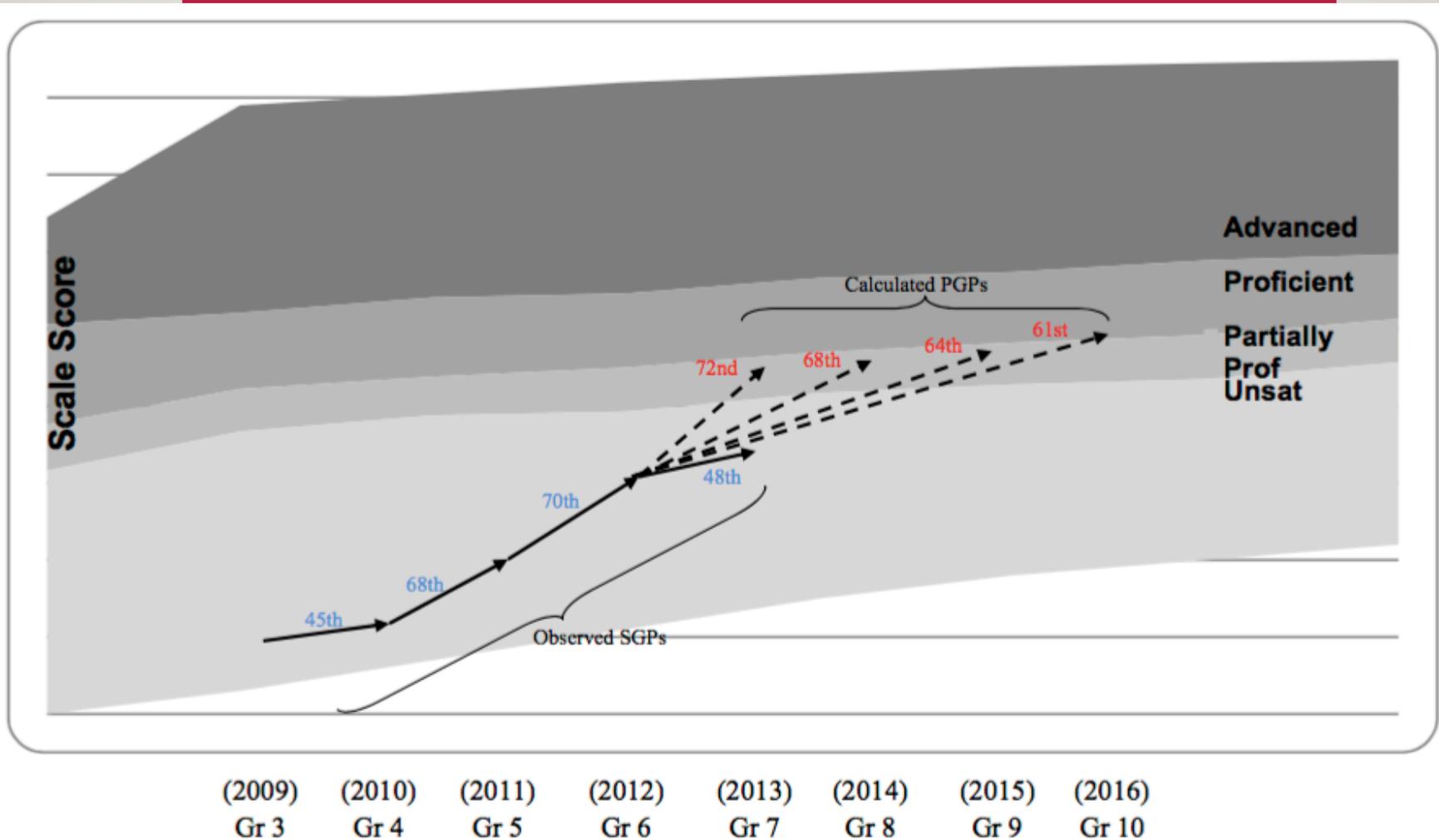
GROWTH TO TARGET 5-YEARS



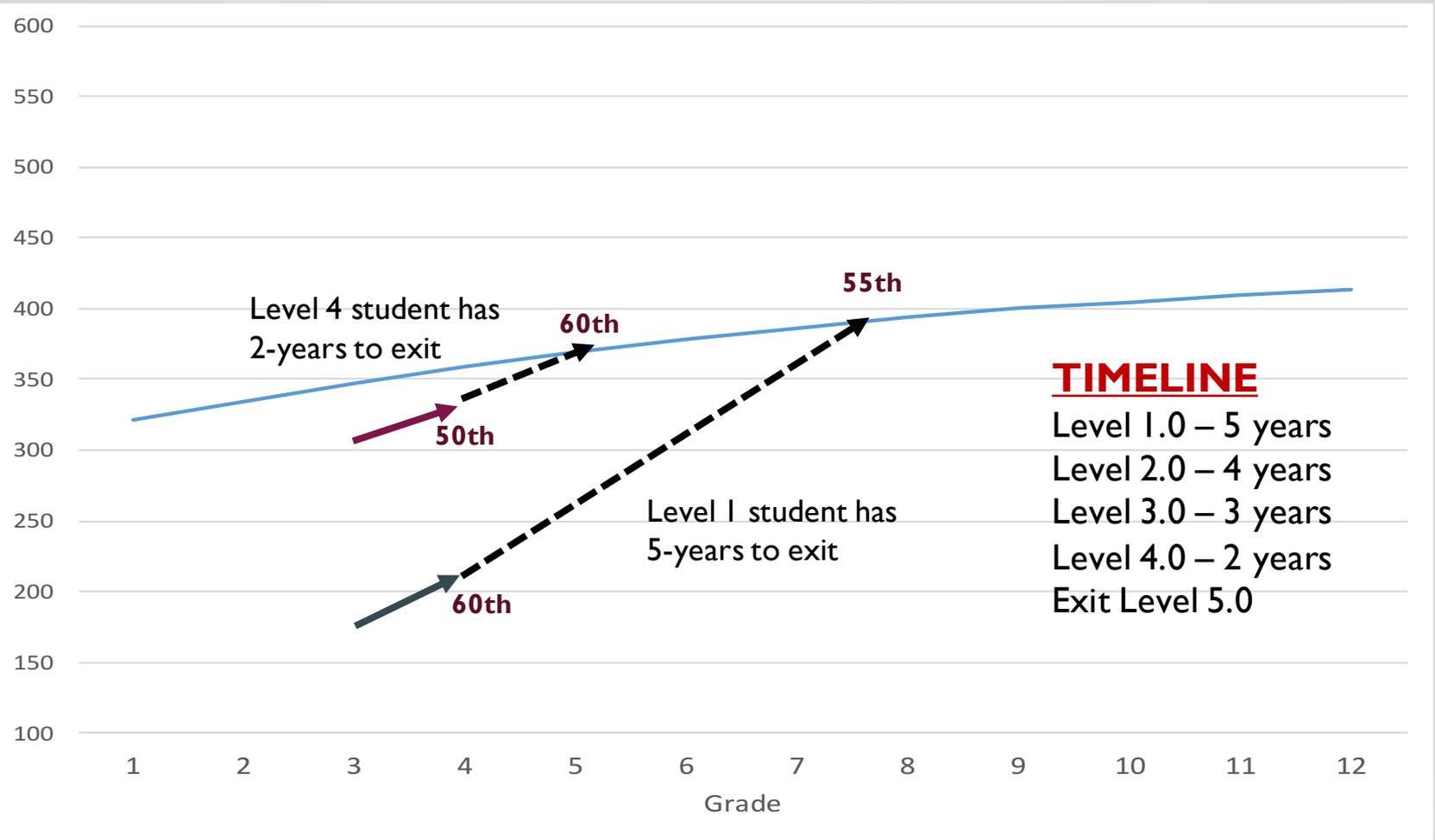
ADEQUATE GROWTH PERCENTILE

- Another potential method for complying with the EL accountability indicator requirements
- Adequate growth percentiles are based on catch-up and keep-up growth
 - catch-up growth = student scores below expectations but shows adequate SGP to attain expectations within the next 3 years
 - keep-up growth = student scores at expectations and shows adequate SGP to stay at expectations
 - For the EL growth “catch-up” growth is used exclusively

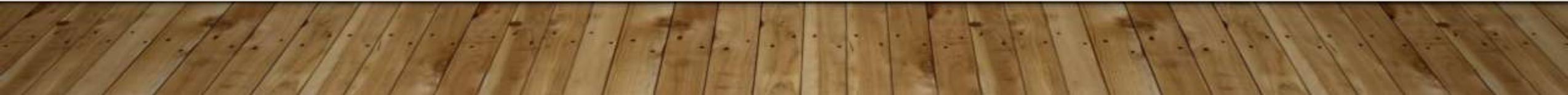
EXAMPLE OF AGP FOR CONTENT GROWTH



EXAMPLE AGP FOR EL GROWTH



LINKING GOALS TO INDICATORS AND MEASURES



ESSA REQUIRED ACTIONS/DECISIONS AND GOALS

- Evaluate school/subgroup performance on each **indicator** and report annually
- Combine performances on indicators to “meaningfully differentiate” schools and give an **overall** rating annually
- At least every three years identify schools in need of “**comprehensive** support and improvement”
- Identify schools in need of “**targeted** support and improvement”
- Identify schools that qualify to **exit** from comprehensive/targeted support and improvement
- Identify schools who go from comprehensive to more intensive support and intervention
- Each action/decision requires a goal/target

ESSA REQUIREMENTS FOR GOALS

- **Set “ambitious long-term goals” for at least five accountability system indicators** (Academic Achievement, Graduation Rate, Progress towards achieving English language proficiency; Academic Progress and/or Other Academic Indicator; School Quality/Student Success)
 - Long-term Goals should be established in terms of the Indicators [“Indicator Goals”] *and* in terms of selected Measures/Metrics [“Operational Goals”]
 - Example of Indicator Goal: *All students will be college ready, defined as qualified to be placed in a credit-bearing college-level first course in English and in mathematics*
 - Example of Operational Goal: *All high school students will score at least Level 3 on the state high school assessments in ELA and mathematics, or...*

ESSA REQUIREMENTS FOR GOALS

- Long-term goals must have same multi-year “term” for all students/subgroups
- Long-term goals must indicate significant progress in closing statewide proficiency and graduation rates gaps between subgroups

ESSA REQUIREMENTS FOR GOALS

- Set “**measurements of interim progress**” for each accountability **indicator** that reflect the long-term goals
- Must establish measurements of interim progress for each **subgroup** consistent with long-term goals
- Must **report** annually for schools and subgroups performance on each indicator in relation to measurements of interim progress
 - Draft accountability regulations stipulate that the state must establish at least performance levels for each indicator and use those performance levels in making accountability determinations
- Hardest goal problem: how to **combine**—what is **overall performance goal**; what does it mean?

CONSIDERING THE INDICATORS



ACCOUNTABILITY INDICATORS

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

ACCOUNTABILITY INDICATORS

- State must set 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 (subgroups?) to produce Academic Achievement indicator for school)
- State reports by indicator for school and by subgroup

SMALL GROUP DISCUSSION

- What is your highest priority education goal?
- For each indicator (Achievement, Growth, EL Proficiency, Graduation Rate)
 - What does success look like?
 - Would you recognize success differently individually? In the aggregate? When examining subgroups?
 - How should we inform ambitious but reasonable targets?
 - What activities should schools/districts engage in to monitor improvement over time?

SCHOOL QUALITY/STUDENT SUCCESS

Indicators of School Quality	Indicators of Student Success
<ul style="list-style-type: none">• Attendance/Student engagement• Enrollment in advanced coursework• School climate surveys• Participation in extra-curricular activities• Percentage of students enrolled in an art course• Educator quality (qualifications, experience, effectiveness)• Suspensions/expulsions• Quality of local assessments or assessment practices• Engagement in professional capacity building	<ul style="list-style-type: none">• Achievement gap indicator• Persistence/dropout rate• Data drawn from post-secondary outcomes• Social-emotional skills• Physical fitness assessment results• Credits earned by end of ninth grade• Algebra readiness by end of 8th grade• Percentage of students entering STEM field• Persistence in post-secondary education• Student performance on the RSA in 3rd grade

SCHOOL QUALITY/STUDENT SUCCESS

- Deepening the measure of academic achievement vs. broadening the measure of school quality?
- Focus on process measures or focus on outcomes?
- Questions to consider:
 - What is your highest priority education goal?
 - What data/indicator would convince you the goal (or progress toward the goal) is being met?
 - How could the inclusion of that indicator in school accountability broaden/deepen the definition of school quality or student success?

PUTTING IT ALL TOGETHER – WHAT ARE OUR OPTIONS?

Approach	Considerations
Decision matrix	A series of rules made by a separate committee
Index	Numerical aggregations that include performance standards for each component and combination rules (e.g., A to F)
Conjunctive	A series of AND statements (e.g., NCLB)
Compensatory	High performance on one indicator compensates for low performance on another
Hybrid of Conjunctive and Compensatory	A combination of compensatory and conjunctive (e.g., must reach a composite score AND a minimum score on each indicator)

PUTTING IT ALL TOGETHER – CONSIDER THESE QUESTIONS

- Should performance on one indicator compensate for another?
- Is there an indicator that should be driving the majority of the decisions?
- Should there be a minimum expectation for each indicator?