Tennessee Comptroller of the Treasury

Office of Research and Education Accountability

Student Growth Portfolios for Teacher Evaluation

February 2020

Anna Johnson
Legislative Research Analyst

Juan Napoles
Legislative Research Analyst

Linda Wesson
Assistant Director

Justin P. Wilson
Comptroller of the Treasury

4 **Key points**

5 **Introduction**

5 **Portfolios in Tennessee: Law, Policy, and History**

5 Portfolio scores as an alternative to teachers' individual TVAAS growth scores

6 Districts required to use at least one portfolio model in 2019-20

8 Teachers drive initial portfolio development

9 The future of portfolios

11 **Portfolio Process and Costs**

11 Portfolio models based on Tennessee's academic standards - changes in standards may require changes in models

11 Collection and scoring of work samples from at least three students chosen to represent different skill levels

13 Portfolio submission to an online platform

13 *GLADiS 2012-2017*

13 *Educopia 2017-18*

13 *Portfolium 2018-current*

14 Tennessee teachers as peer and expert reviewers who score portfolios

15 Converting student growth scores into a teacher portfolio score

15 *Student growth scores*

15 *Growth level ratings*

16 *Score resolution*

17 *Final teacher portfolio scores*

17 Costs associated with the portfolio process

18 **District Compliance with State Portfolio Laws**

18 Pre-k/Kindergarten portfolio model requirements for districts awarded state VPK funds

19 Districts requirements to use at least one portfolio model in 2019-20

19 **Effectiveness of Portfolio Models**

20 Expanding portfolio goals

21 Analyzing portfolio models for validity and reliability

21 *Scoring results of individual growth components*

25 *Validity*

30 *Reliability and repeatability*

34 **Implementation Challenges**

34 Time to prepare and submit portfolios

36 Rural districts' unique challenges

37 **Student Growth Portfolios in Other States**

37 Portfolios in teacher preparation and licensure

38 Student learning objectives: a portfolio-like tool

38 **Conclusion: Portfolios as a Component of Teacher Evaluation**

38 Current use

39 Changing goals

40 Other states' use

40 Portfolios' future

41 **Policy Considerations**

44 **Appendix A - District portfolio adoptions, 2018-19, 2019-20**

48 **Endnotes**

Key points

- **Nine percent of teachers in 2018-19 received a portfolio score as part of their annual teacher evaluations.** In comparison, approximately 29 percent of teachers received an individual Tennessee Value-Added Assessment System (TVAAS) score. Portfolio and individual TVAAS scores are weighted at 35 percent of a teacher's overall evaluation score, serving as the student growth component. The majority of teachers, 62 percent, do not receive individual growth scores from TVAAS or portfolios; instead, they receive a school level TVAAS score with a reduced weight for the student growth component of their teacher evaluations.
- **Portfolios are collections of student work from two points in time during the school year showing student progress in mastering a state academic standard.** Teachers select samples of work from at least three students in their class that reflect the four key standards, or blended standards, prescribed by the portfolio model. Teachers self-score the work, and peer reviewers – other teachers in the same grade or subject – also score the work, based on scoring rubrics. In cases where peer review scores are more than one rating level apart from the self-scores, additional reviews are performed.
- **Teachers with portfolios were considerably more likely to receive the top score for the student growth component of their evaluations than teachers with individual TVAAS or school level TVAAS scores.** In 2018-19, 74 percent of teachers with portfolios received the top score of 5, compared to 19 percent of teachers with individual TVAAS scores and 34 percent of teachers with school level TVAAS scores. The lack of differentiation among growth scores for portfolio teachers suggests that portfolio models may not provide as much useful information about a teacher's ability to increase student learning relative to teachers evaluated using other student growth measures.
- **Portfolio models have several design factors that make them less valid and reliable than more standardized, objective measures of student growth.** Such factors include the lack of standardized assessment tasks, the use of teachers' self-scoring, the expansion of score categories in the rubrics used to evaluate student work, and the subjective nature of purposeful sampling, with the possible use of work samples from as few as three students to judge the instruction of a whole class. OREA analysis found relatively low interrater scoring agreement, especially for some standards within some models. Scoring procedure changes over several years limit the usefulness of year-to-year portfolio score comparisons.
- **Other features of the portfolio model process may help improve teachers' instructional practice and provide more detailed information about teachers' skills and professional development needs.** The portfolio models' focus on state standards, on helping students at varying performance levels achieve growth, and on collaboration with other teachers may be a more effective way to provide teacher professional development and improve teacher instructional practice than its current use in teacher evaluations as a quantitative measure of student growth. Originally designed as a way to measure students' growth for teacher evaluations, portfolios have also been cited as a method for teachers to reflect on and improve their teaching practice, as a way to help improve pre-k program quality, and as a method to help the state meet its 3rd grade reading goals.
- **Pre-k and kindergarten teachers make up the vast majority (79 percent in 2018-19) of teachers receiving portfolio scores due to a 2016 law requiring districts that accept state Voluntary Pre-k (VPK) classroom funds to adopt the pre-k/kindergarten portfolio model.** Other portfolio models include 1st grade, 2nd grade, fine arts, physical education, and world languages. Fine arts and 1st grade teachers each make up about 8 percent of all teachers who receive portfolio scores. The other three models account for the remaining 5 percent of teachers who receive portfolio scores. The Tennessee Department of Education is considering alternative measures to the pre-k/kindergarten portfolio model for districts to pilot in 2020-21, as provided for in a 2019 law.
- **Tennessee's portfolio model is unique in both purpose and scale.** Student growth portfolios have been used in other states to assess students and teachers in other ways, but Tennessee appears to be the only state using portfolios as a quantitative measure of student growth in annual teacher evaluations. Tennessee is also the only state requiring portfolio use by all local districts.

Introduction

A portfolio is a collection of work. A student growth portfolio contains samples of student work from two points in time that attempt to demonstrate learning during that time. The two work samples, from the same student and related to specific Tennessee academic standards, are each scored, using a common set of criteria, or rubric. The second work sample, from later in the school year, should demonstrate a student's increased knowledge or skills. The difference between the scores of the two work samples is the measure of the student's progress, or growth, toward mastering selected academic standards. A teacher's submission of student work samples from multiple students is a teacher's student growth portfolio. The average of the students' growth scores is converted to become the teacher's portfolio growth score.

Teachers' portfolio growth scores are used as the required growth component for their annual teacher evaluations. Portfolio growth scores are comparable to teachers' individual TVAAS (Tennessee Value-Added Assessment System) scores – calculated from students' TN Ready test results – in that both types of scores represent the academic growth of students taught by that teacher. For teachers of grades or subjects without an applicable TN Ready test, districts can choose a portfolio model for selected grades and subjects to provide those teachers with individual growth scores. If teachers do not have a TN Ready test associated with their classes and their district has not opted to use a portfolio model in their respective classes, then teachers must use one of the school level composite TVAAS growth scores, reflecting many or all students' scores, including students not taught by the teacher.

In May 2019, Representative John Ragan requested that the Comptroller's Office of Research and Education Accountability (OREA) analyze the use and effectiveness of student growth portfolios as part of Tennessee's teacher evaluation system. The request asked OREA to specifically examine the statewide implementation of portfolio models, school districts' compliance with portfolio requirements, the reliability, validity, and repeatability of portfolio models, the extent of portfolio model use in other states and the effectiveness of such models, teacher time burdens in compiling portfolios, and implementation issues with portfolio models in rural districts.

Portfolios in Tennessee: Law, Policy, and History

Portfolio scores as an alternative to teachers' individual TVAAS growth scores

The Tennessee legislature overhauled the state's teacher evaluation system through its passage of the First to the Top Act in 2010.[1] The new evaluation system required that at least 35 percent of teachers' evaluations be based on TVAAS student growth data or some other comparable growth measure. Other elements of the teacher evaluation system include classroom observations (50 percent) and a selected student achievement measure (15 percent). All three components are combined into a single *level of effectiveness* (LOE) score to reflect the teacher's overall performance.

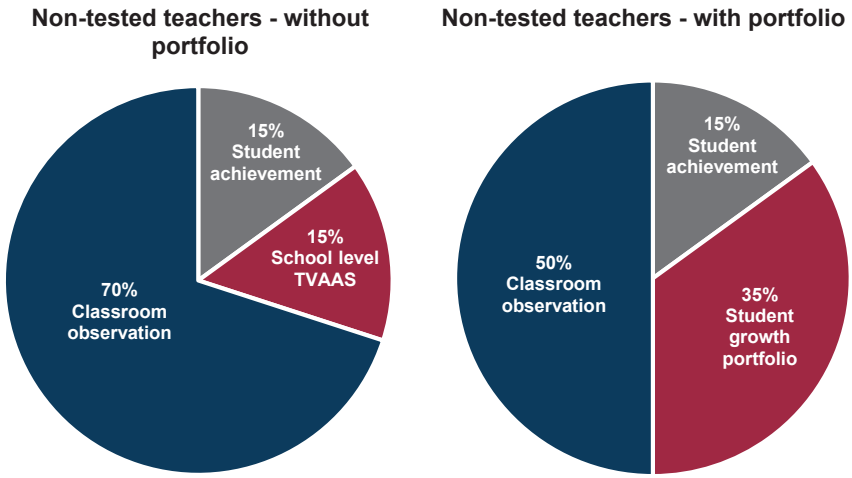
Since 2010, the teacher evaluation system has been modified several times through law, reducing the weight of the student growth component for teachers in non-tested grades and subjects from 35 percent to 15 percent. As of 2019, teachers in untested grades or subjects use TVAAS school level composite scores for 15 percent of their overall LOE evaluation score; their classroom observation component is given a correspondingly higher weight of 70 percent, rather than 50 percent.

In districts that have opted to use one or more portfolio models, a teacher's portfolio score is weighted at 35 percent, the same percentage weight as that used for individual TVAAS growth scores for teachers who have them. When the First to the Top Act restructured Tennessee's teacher evaluation system, it required a student growth data component based on TVAAS "or some other comparable measure of student growth."[2]

Later revisions to the law gave the State Board "ultimate authority" to determine, identify, and adopt measures of student growth that are comparable to TVAAS.[3] As of December 2019, the board has approved the following comparable growth measures, all of which are student growth portfolio models:

- pre-kindergarten/kindergarten portfolio model,
- 1st grade portfolio model,
- 2nd grade portfolio model,
- fine arts portfolio model (vocal and instrumental music, visual and media arts, dance, theater),
- world languages portfolio model (foreign languages such as French or Spanish), and
- physical education portfolio model (elementary grades K-5 and middle grades 6-8).

Exhibit 1: Evaluation components for teachers in non-tested grades and subjects



Source: State Board of Education, Policy 5.201.

Districts required to use at least one portfolio model in 2019-20

In 2018, Public Chapter 552 was passed, requiring local school districts to use at least one "alternative growth model" approved by the State Board of Education for teachers who do not have individual TVAAS scores. Most districts, however, were already using the pre-k/kindergarten portfolio model because of an earlier state law: the 2016 Pre-K Quality Act.[4] The Pre-K Quality Act requires school districts that receive state funding for pre-k classrooms under the Voluntary Pre-K (VPK) program to use the state-approved portfolio model for its pre-k and kindergarten teachers' annual evaluations. In 2018-19, 137 school districts (96 percent) received state funding for pre-k classrooms, thus requiring those districts to use the related portfolio model. (See more at "District Compliance with State Portfolio Laws" section, page 18.)

Before these laws were passed, districts had the option to adopt one or more portfolio models. When a district adopts a model, every school in that district must comply. For instance, when Wilson County Schools adopted the physical education model, each physical education teacher in its 22 schools was required to use the physical education portfolio model for teacher evaluation.

The Pre-k Quality Act requirements, implemented in 2017-18, have effectively compelled districts to adopt a pre-k/k portfolio model. Since almost all districts receive VPK state funding, they have to choose the pre-k/k portfolio model unless they are willing to discontinue their pre-k programs or forgo state funds, as one district has done for the past two years. The handful of districts not receiving VPK funds must adopt at least one portfolio model and all districts are free to implement multiple portfolio models.

The VPK requirements make the pre-k/kindergarten portfolio model by far the most common portfolio model in use. Other portfolio models are used by significantly fewer districts. Some portfolio models have declined in usage. For example, the fine arts portfolio model was in use by 20 districts in 2015-16, but as of 2018-19, only seven districts used that model.[5] The number of teachers submitting portfolios depends on the size and number of districts adopting a particular portfolio model. (See Exhibit 2 for a summary of 2018-19 portfolio use in districts and Appendix A for a listing of each district's portfolio use.)

Exhibit 2: 2018-19 Student growth portfolio use

Portfolio model	Number of districts that adopted model	Teachers who submitted portfolios[6]	
		Number	Percent
Pre-K/Kindergarten**	137	Pre-K: 1,105 K: 3,698	18.2% 61.0%
1st grade	35	480	7.9%
2nd grade	3	51	0.8%
Fine arts	6	508	8.4%
Physical education	3	140	2.3%
World languages	1	77	1.3%
Total	**-**	**6,059**	**100.0%**
Total unique districts	**137**		

**Pre-K/Kindergarten portfolio model is required for districts who receive state funding under the Voluntary Pre-K program. In 2018-19, 137 districts accepted VPK funding.
Source: Tennessee Department of Education portfolio score data, 2018-19.

About 30 percent of districts used more than one portfolio model in 2018-19. The most common choice of districts after the pre-k/kindergarten model was the 1st grade model. Two districts were notably high users of portfolio models in 2018-19: Shelby and Cumberland County schools each used four.

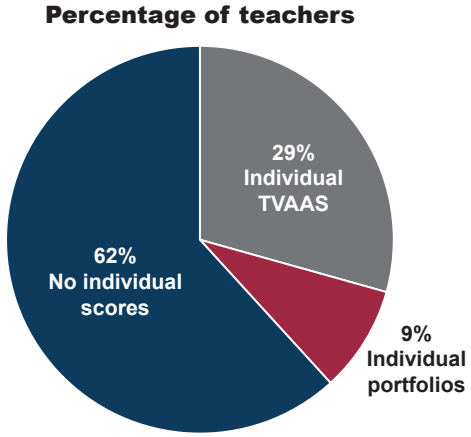
Half of all teachers who submitted portfolios for the 2018-19 school year were from 11 local districts. These districts generally have larger student enrollments and, thus, larger numbers of teachers. Because most districts are required to use the pre-k/kindergarten portfolio model, districts with the largest kindergarten enrollments tend to be the districts with the most teachers submitting portfolios. The one outlier is Cumberland County; although it has a relatively small kindergarten enrollment, its use of multiple portfolio models puts it among the top districts with participating teachers. (See Appendix A.)

Fast facts

Tennessee educated about 974,000 students in 2018-19 across 143 public school districts. Of approximately 66,000 classroom teachers who received annual evaluations that year, about 29 percent, or 19,000, teach grades or subjects that have a TN Ready test that generates an individual TVAAS growth score for their teacher evaluation.

Sources: Tennessee Department of Education, 2018-19 State Report Card and 2018-19 teacher evaluation score distribution data provided by TDOE.

Exhibit 3: 2018-19 Teacher evaluation growth scores by type of score



Source: Tennessee Department of Education, 2018-19 score data.

Despite the recent state requirements that have increased portfolio use in districts, the majority of teachers do not receive individual growth scores, from either portfolio models or TVAAS, on their annual evaluations. As Exhibit 3 shows, 62 percent of teachers were assigned a school level composite TVAAS score in 2018-19 in lieu of an individual growth score. The school level composite is weighted at 15 percent of a teacher's total evaluation and can be one of 21 different types of approved school level scores. The remaining 38 percent of teachers received a student growth portfolio score or an individual TVAAS score weighted at 35 percent of their total evaluation.

Teachers drive initial portfolio development

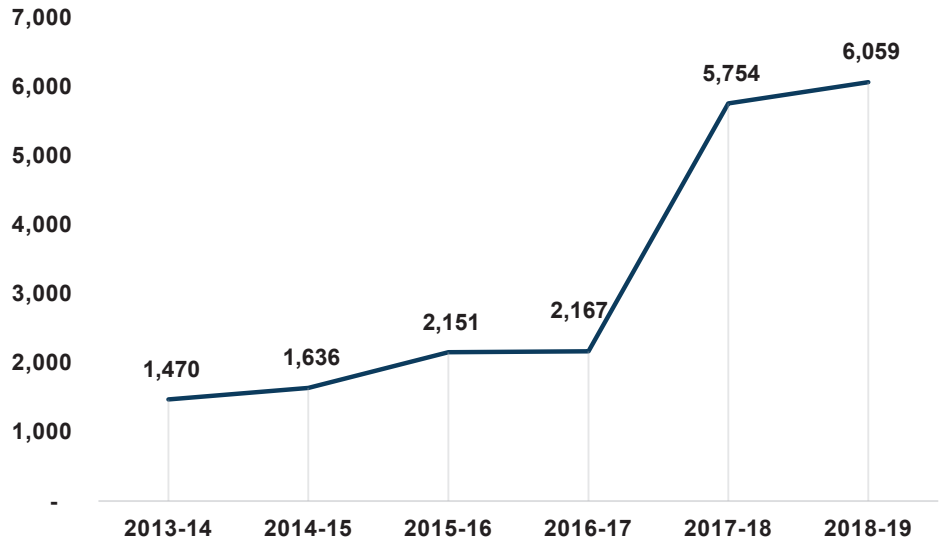
Tennessee's portfolio models for teacher evaluation were first developed by local teachers who were dissatisfied with being dependent on a school composite growth score for their annual evaluations. The fine arts portfolio model was developed and used by a group of teachers in the former Memphis City School District during 2011-12. After receiving State Board of Education (SBE or State Board) approval for use as an alternate growth measure, the fine arts portfolio model was fully implemented by three districts in the 2012-13 school year.

In 2012-13, teachers from Memphis City Schools again developed and implemented a portfolio model, this one for world languages. Three districts, including Memphis City Schools, piloted the model. More educators became involved in the development and pilot testing of subsequent portfolio models; current and retired teachers, principals, academic coaches, district leaders, university staff, and TDOE staff worked to develop other portfolio models.

As the first portfolio model approved, fine arts accounted for the bulk of portfolio use earlier on: 95 percent in 2013-14, and declining to 66 percent in 2015-16 as the use of other models grew.[7]

In 2017-18 when portfolios were first required for pre-k and kindergarten teachers in districts accepting VPK funds, portfolio usage jumped 166 percent from the previous year. (See Exhibit 4.)

Exhibit 4: Number of teachers submitting portfolios



Sources: Tennessee Department of Education score data, 2017-2019; Division of Data and Research, Tennessee Department of Education, *The Rise of Student Growth Portfolio Models in Tennessee*, Jan. 2017, p. 5, (2014-2016).

While the use of pre-k/kindergarten portfolios increased due to the related statutory requirement, the number of districts using the other portfolio models has generally declined since 2017-18. (See Exhibit 5.)

Exhibit 5: Portfolio models adopted by districts, 2017-18 through 2019-20

2017-18

- Pre-K/Kindergarten: **138**
- 1st grade: **39**
- Physical education: **16**
- Fine arts: **26**
- World languages: **9**

2018-19

- Pre-K/Kindergarten: **137**
- 1st grade: **35**
- 2nd grade: **3**
- Physical education: **3**
- Fine arts: **6**
- World languages: **1**

2019-20

- Pre-K only: **1**
- Pre-K/Kindergarten: **137**
- 1st grade: **35**
- 2nd grade: **5**
- Physical education: **4**
- Fine arts: **6**
- World languages: **2**

Sources: Tennessee Department of Education, District Flexibility Surveys, which report districts' plans to use various alternatives related to teacher evaluations (such as portfolio models for the individual student growth component), and 2018-19 portfolio score data.

The future of portfolios

In 2019, the General Assembly passed Public Chapter 376, which revised the Pre-K Quality Act requirement that districts receiving VPK funds for preschool classrooms had to use the pre-k/kindergarten growth portfolio model for teacher evaluations. The new law allows districts to to use a "comparable alternative measure of student growth approved by the State Board of Education" as an alternative to the portfolio model. The Tennessee Department of Education (TDOE or the department) has requested recommendations from districts of possible alternatives that the department could consider presenting to the State Board for approval. In November 2019, the department presented to the State Board the proposed guidelines for approving alternative growth measures, including that the measures be nationally normed, evidence based, and able to evaluate all students in pre-k and kindergarten on math and English language arts. The State Board is expected to consider final approval of the guidelines in February 2020. Following one year of pilot testing by districts, alternative growth measures must be reviewed by the Commissioner of Education before being submitted to the State Board for final approval statewide, according to State Board rules.

Some of the early alternative growth measures proposed by districts are benchmark tests that could be given to pre-k and kindergarten students at the beginning of the school year and again at the end of the year in order to calculate growth. Some benchmark tests have already been approved by the State Board for use as measures for the 15 percent of a teacher's evaluation based on student achievement scores. The department is considering alternatives only for the pre-k/kindergarten portfolio model, as the law prescribes.

The 2018 law, Public Chapter 552, required TDOE to develop "valid and reliable alternative student growth models" for the grade levels and subjects that do not yet have such models. Department staff state that they are working to first ensure smooth implementation of current portfolio models and pre-k/kindergarten alternatives before further expansion of portfolio models. The low rate of voluntary adoption of existing portfolio models by districts is another factor influencing TDOE's timeline for the development of additional models. The department has indicated a desire to revise the 2nd grade model and broaden the physical education model to include all grades, K through 12. See Exhibit 6 for subjects and grades covered by standardized tests for which an individual teacher's TVAAS growth score is calculated and for which a portfolio growth score may be available, depending on district adoption.

Exhibit 6 : Grades and subjects for which individual teacher growth scores are calculated (TVAAS) or may be available (portfolio), depending on district adoption, and those for which no individual growth scores are available

Grades	TVAAS growth scores available		Portfolio growth scores available		No individual growth scores available
Pre-K	-	-	ELA & Math		Special Ed.
K-2	-	-	ELA & Math	Fine Arts, Phys. Ed., Languages	Special Ed.
3	ELA & Math*	-	-	Fine Arts, Phys. Ed., Languages	Special Ed.
4-8	ELA & Math	Science & Social Studies	-	Fine Arts, Phys. Ed., Languages	Special Ed.
9-12	English I, II	Biology	-	Fine Arts, Languages	Special Ed. English III, IV Calculus, Statistics, and other math Chemistry, Physics, and other science World History, Geography, and other social studies CTE/vocational courses Phys. Ed.
	Algebra I, II	U.S. History			
	Geometry				
	Integrated Math I, II, III				

* Note: Grade 3 teachers in districts not using the optional 2nd grade assessment cannot receive an individual TVAAS score.
Source: Tennessee Department of Education.

Exhibit 7: Timeline of major events related to student growth portfolios

2011-12
*Fine Arts Portfolio pilot tested

2012-13
***SBE approves Fine Arts model**

*World Languages pilot tested

*SBE adopts revised early learning standards for 4-year-olds (pre-k)

2013-14
***SBE approves World Languages model**

*Phys Ed K-5 pilot tested

2014-15
***SBE approves Phys Ed K-5 model**

*PreK/K pilot tested

2015-16
***SBE approves Pre-K/K model**

*1st Grade pilot tested

*Pre-K Quality Act passed requiring use of Pre-K/K model

*SBE adopts revised math and ELA standards for K-12

2016-17
*TDOE revises early grades portfolio models to align with new math and ELA standards

***SBE approves 1st Grade model**

2017-18
***Phys Ed 6-8 model implemented**

*Required portfolio use for pre-k/k teachers begins in relevant districts

*TDOE changes online platform from GLADiS to Educopia

*SBE adopts revised pre-k standards

*New law passed requiring all districts to use at least one portfolio model by 2019-20

*Problems with new platform resulted in law prohibiting negative consequences based only on teachers' portfolio scores

2018-19
*TDOE changed electronic platform from Educopia to Portfolium

***SBE approves 2nd Grade model**

*Law revised to allow affected districts to use an SBE-approved alternative to the Pre-K/K portfolio model

*Portfolio Review Committee created by law

2019-20
*Portfolio Review Committee meets and generates recommendations

Key:
Laws in red type
Portfolio platform system changes in green type
Testing and adoption of new portfolio models in purple type
All other actions in black type

Portfolio Process and Costs

Portfolio models based on Tennessee's academic standards – changes in standards may require changes in models

The student work samples collected by teachers for their portfolios must be aligned to a relevant academic standard, or a group of standards, specified by the portfolio model. Standards are grade-level expectations of the knowledge or ability levels that students should obtain at the end of a grade/course. Teachers assign students tasks at two points in time during the school year, with such tasks providing students the opportunity to demonstrate all that they know about a standard. After students show what they know about a standard, the portfolio scoring rubric is used to score the performance level at both points of time (point A and point B).

The portfolio models specify the selected standards for which teachers must collect representative student work, but the number of standards assessed in portfolios vary from model to model, and some models provide standards options from which teachers can choose. For example, the physical education K-5 model includes one required standard and six other standards from which teachers must choose three to include in their portfolios. The early grades models (pre-k/kindergarten, 1st, and 2nd grade) require two English/language arts (ELA) assessments of a standard set – three individual standards blended together – and offers three options for each of the two required assessments. The world languages portfolio model, and the state standards it is based on, are another variation. This portfolio model is organized by student proficiency levels (novice, intermediate, and advanced) and sublevels (low, medium, and high), with expected skills detailed for each level. This allows one set of standards, and one portfolio model, to be applied to all grade levels, which is fitting for foreign language studies where students may begin the classes at various grade levels.

Because portfolios are based on academic standards, when Tennessee revises its standards, the portfolio models may also be subject to revision, depending on the extent of the standards changes. The State Board of Education updates standards at least every six years through the multi-step statutory process. In 2016-17, for example, revised math and ELA standards were approved by the State Board for all grades, K-12. Pre-kindergarten standards were then revised to align with the new math and ELA kindergarten standards.

The 2016-17 revisions to the state ELA standards resulted in significant changes to the pre-k/kindergarten portfolio model. Three standards were blended into one portfolio standard to reflect the integrated nature of early literacy standards. Although teachers have a choice among several of these blended ELA standard sets, each choice counts as one portfolio standard but measures three individual academic standards. Teachers in several districts, who had helped develop and pilot test the original portfolio model, found the revised pre-k/kindergarten portfolio model very different from the original they had worked with.

Collection and scoring of work samples from at least three students chosen to represent different skill levels

For each standard (or standard set) measured by a portfolio model, teachers must collect point A and B work samples from three students at different performance levels: emerging, proficient, and advanced. (There are exceptions: the physical education portfolio model requires six students' work samples for each standard, and the fine arts model provides for submission of group samples, such as ensemble or class performances, for certain standards.) The work documenting all three students' progress on the selected standards is referred to as a portfolio collection. With four collections making up each portfolio, there is a potential for 12 students' work to be included in a teacher's portfolio, but if the same three students' work is chosen for all four collections, then a teacher's portfolio is limited to three students. (See a summary of how portfolios are created at Exhibit 8.)

The Department of Education states that best practice is to collect point A work samples from all students at the beginning of the year, rather than only those students teachers believe would be good representatives for portfolio purposes. Teachers can use these early samples and other information, such as universal screening data, to determine which students fit in the three performance level groups.

When point B work samples are collected later in the year, teachers score them and subtract point A work sample scores in order to determine:

- the growth score for each student, and
- the growth score typical for each of the three performance groups: emerging, proficient, and advanced.

Teachers score the work samples using their portfolio's scoring criteria, or rubric, provided as part of each model. This teacher evaluation of their students' growth (the self-evaluation) is the first step in the portfolio scoring process.

Exhibit 8: How teachers build a student growth portfolio – summary of state guidelines

Portfolio Model

A portfolio model typically specifies ***four state standards*** for a subject and grade level(s) to be assessed or allows teachers to select some standards from specified options.

Standards 1 2 3 4

Teachers collect a ***point A*** work sample from students early in the school year. (TDOE encourages, but does not require, that teachers collect ***point A*** samples from all students.) Teachers score the samples using a rubric detailing what student work for that standard looks like at each score level and score them from 1 to 7. (Early grades models can score work from 0 to 7.)

Point A

Teachers are to sort their students into three *performance groups* – ***emerging, proficient, and advanced*** – based on their ***point A*** work and other performance information. While the student performance groups may vary for each standard, they can also remain the same. In other words, a student considered emerging for one standard may not be emerging on other standards.

Performance groups

emerging | proficient | advanced emerging | proficient | advanced emerging | proficient | advanced emerging | proficient | advanced

Near the end of the school year (or the end of a curriculum unit), teachers are to collect a ***point B*** work sample from students and use the same rubric from point A work samples to score them.

Point B

After teachers collect ***point B*** work samples from their students near the end of the year, and growth scores are calculated (point B score minus point A score), they are to select the student's point A and B *work sample pairs* that represent the typical amount of growth made by each of the three performance groups (emerging, proficient, advanced) for each of the four standards.

The three pairs of work samples for one standard is a collection.

Work sample pairs

Each *portfolio* contains four collections, with a total of 12 pairs of work samples. The 12 pairs of work may represent as few as three or as many as 12 individual students.

Portfolio

Source: OREA summary of state guidelines for a typical student growth portfolio.

Teachers then select a set of point A and B work samples from one student in each performance group that is representative of the most common growth of that group. This selection of representative work samples by performance group is called *purposeful sampling*. (See more about purposeful sampling in "Effectiveness of Portfolio Models" on page 19.)

A growth score for one portfolio collection is calculated by averaging the growth scores of each student whose work is selected for the portfolio standard being measured. A growth score of at least 1.0 but less than 2.0 is set as the expected amount of student growth for one school year and translates to a "growth at expectations" growth level 3 rating. (See more about portfolio scoring on pages 15-17.)

Portfolio submission to an online platform

Teachers submit their portfolios electronically to the state using a private online platform, contracted and administered by the department. Teachers input the scores for each work sample, and the platform calculates student growth.

In addition to the student work samples in their various formats (word, picture, video, or audio files) and students' scores, teachers can also submit explanations about the context of the students' assignments to help portfolio reviewers understand the work they are scoring. (See more about portfolio reviewers in "Tennessee teachers as peer and expert reviewers who score portfolios" on page 14.)

Since 2012-13, the state has contracted with three different vendors to provide the online portfolio platform.

GLADiS 2012-2017

Portfolios were initially implemented through the GLADiS Project online system during a time when Tennessee's portfolio use was primarily limited to the fine arts model. With this platform, teachers and reviewers did not record scores for the point A and B work samples, but simply recorded the growth scores.

As the portfolio program grew, and more models were added, the department sought a vendor capable of handling the increasing load of submissions and able to provide more reliability and accuracy in scoring.

Educopia 2017-18

After the General Assembly began requiring all districts accepting VPK funds to adopt the pre-k/kindergarten portfolio model, the department sought a new platform vendor that could serve a significant increase in portfolio submissions and ensure consistency statewide in the submission process as well as the scoring process. For the 2017-18 school year, the department contracted with Educopia, a vendor that the state had already worked with to test a new scoring process.

Multiple issues with the Educopia platform and scoring process resulted in the department allowing teachers affected by uploading and scoring problems to have their portfolio scores removed from their overall evaluation (LOE) scores. These issues with Educopia contributed to the state choosing a different portfolio platform vendor, though TDOE had already planned to issue a request for proposal (RFP) for the following year's (2018-19) portfolio platform in order to seek a platform that could align with a related TDOE system.

Portfolium 2018-current

In 2018-19, the state entered into a five-year contract with Portfolium. Like previous platforms, Portfolium also experienced capacity-related problems. On the last day to submit portfolios for the 2018-19 school year, Portfolium experienced a blackout and teachers were unable to access the platform. Additionally, at a meeting for peer reviewers to work on the first round of scoring, the heavy site activity overwhelmed the platform. Despite the technical issues, feedback from teachers suggests that the Portfolium platform provided a smoother and easier method for uploading and labeling student work samples compared with the Educopia

platform. The Portfolium platform reduced the number of certain kinds of mismatched student work samples. Teachers must still check to prevent other kinds of mismatches – for example, when student work samples that are either not matched to the correct academic standard or incorrectly link one student's point A or point B work samples with those of a different student. Portfolio collections with mismatched student work samples are assigned the lowest score possible by reviewers as specified in the department's portfolio scoring guidelines.

The department chose to continue using Portfolium for the 2019-20 school year, in part because the future use of pre-k/kindergarten portfolio models is unknown. A 2019 law authorized the State Board of Education to approve alternatives to the pre-k/kindergarten portfolio model, and the department presented guidelines for approving alternative growth measures to the State Board in November 2019. The State Board is expected to consider final approval of the guidelines in February 2020, and districts may begin pilot testing of alternatives in school year 2020-21 with the conditional approval of the Commissioner of Education. With pre-k and kindergarten teachers totaling 80 percent of all teacher portfolio submissions, a move by many districts to pilot portfolio alternatives could result in significant changes to the capacity requirements of an online portfolio platform.

Tennessee teachers as peer and expert reviewers who score portfolios

Fellow teachers, who teach the same portfolio grade level or subject, score the portfolios using the same rubrics as the teachers submitting portfolios, scoring each collection of student work at point A and point B. In the 2018-19 school year, there were 739 total reviewers trained and certified by the department.[8] Of those, 712 reviewers are peer teachers who also submitted their own portfolios. Almost all (99 percent) of reviewers who submitted portfolios received an "at expectations" growth level rating of 3 or higher on their own portfolios. (See Exhibit 9.)

Exhibit 9: Portfolio reviewers' own portfolio scores, 2018-19

	1	2	3	4	5
Frequency	3	4	25	103	577
Percent	0.4%	0.6%	3.5%	14.5%	81.0%

Source: Tennessee Department of Education score data, 2018-19.

The state has required local school districts to designate one teacher as a peer reviewer for every 10 portfolios the district submits per content area.[9] This policy attempts to ensure that the state has enough portfolio reviewers and to create district involvement and a shared workload across all participating districts. The department indicates it has been difficult to accurately gauge the number of reviewers needed. Each district identifies its portfolio peer reviewers in the spring, prior to teachers' portfolio submission deadline.

All peer reviewers complete online training, which covers how to apply the scoring rubric and how to access student work samples and record scores in the platform. Reviewers must pass a certification test in their content area or grade level. In 2019, certified peer reviewers were also required to attend the state scoring meeting for their portfolio model during the summer. These meetings (also called convenings) brought together all the peer reviewers in one area to score portfolios and have department staff on hand to answer questions.

For the last two years, peer reviewers have been paid stipends through the department. In 2018-19, peer reviewers received $300 for attending the one-day convening and for scoring at least five portfolio collections. (A collection is typically the point A and B student work samples from three students for assessing one portfolio model standard.) Reviewers of kindergarten portfolios were required to attend a two-day convening (which was later reduced to one day due to technical difficulties) and score at least 10 collections. Reviewers who traveled more than 50 miles to the convenings in Nashville also received a travel stipend.

Reviewers could earn an additional $450 by scoring 30 to 35 more collections on their own, bringing their total scored to 40 collections, the equivalent of 10 portfolios. If needed, reviewers could also earn an additional $25 per collection beyond the original 40 assigned.

After reviewers complete the first round of scoring, certain score results will trigger a second peer review. (See details of the scoring process in the next section, "Converting student growth scores into a teacher portfolio score.") First and second peer reviewers are drawn from the same pool of potential reviewers. In 2018-19, 8.6 percent (2,089) of portfolio collections scored were assigned to a second peer reviewer.[10]

Certain score results from the second peer review can trigger a final, expert review. Expert reviewers are recommended for the role by their district and are certified by the department after completing the reviewer training and passing the relevant certification tests. Expert reviewers can be fellow peer reviewers, district supervisors, or other subject matter experts. Of the 2,089 portfolio collections assigned to a second peer review in 2018-19, about 4 percent (90 collections) required an expert review.

Converting student growth scores into a teacher portfolio score

Peer and expert reviewers score student work samples using the portfolio scoring rubrics for each standard, which detail performance scores of 1 through 7 for student work. (Scoring for the world languages model is an exception because it is designed for use with all grade levels; scores of 1 through 9 are possible in this model.) A performance score of 3 reflects student work that meets the grade level standard based on the scoring rubrics. Early grades rubrics, as of the 2019-20 school year, now include a score of 0 for point A work samples, a suggestion from the 2019 Portfolio Review Committee discussions.[A]

Student growth scores

The portfolio platform automatically calculates students' growth, which is the difference between the point A and point B performance scores on their work samples. The department expects that one year of a typical teacher's instruction will help students achieve a growth score of at least 1.0. This would be the difference between a point B performance score of 4 and a point A performance score of 3, or a point B score of 2 and a point A score of 1, for example. If a student received a performance score of 4 on both point A and point B work samples, the difference would be a growth score of zero.

The platform also calculates the average student growth for each collection, which for most portfolios includes the growth scores of the three students representing emerging, proficient, and advanced groups. If, for example, the three students' growth scores in one portfolio were 1.0, 1.0, and 2.0, the resulting average growth score for the collection would be 1.3.

Growth level ratings

The platform then converts the average growth scores for each collection to a *growth level rating* using the department's student growth indicator chart. (See Exhibit 10.)

[A] The Portfolio Review Committee was created by Public Chapter 376 (2019) and met July 23, 2019 to consider improvements to the pre-k/kindergarten student growth portfolio model.

Exhibit 10: TDOE's Student Growth Indicator Chart used to calculate portfolio student growth level ratings

Portfolio Collection Scoring	
Growth Level Rating for Collections (Assigned to the average calculated from all student growth scores in a collection)	**Student Growth Scores** (Difference between Point B and Point A performance scores on student work samples, averaged for all students in a collection)
Student Growth Indicator Chart	
Level 5 Significantly Above Expectations	Students demonstrate, on average, **three or more levels of student growth (= or >3 levels of growth)**
Level 4 Above Expectations	Students demonstrate, on average, **two levels of student growth, but less than three levels of student growth (=2 levels of growth, but <3 levels of growth)**
Level 3 At Expectations	Students demonstrate, on average, **one, but less than two levels of student growth (=1 level of growth but <2 levels of growth)**
Level 2 Below Expectations	Students demonstrate, on average, **less than one level of student growth (>0 levels of growth but <1 level of growth)**
Level 1 Significantly Below Expectations	Students demonstrated, on average, **no growth or negative growth**

Source: OREA adaptation of TDOE Student Growth Indicator Chart.

- A 1.3 average based on three students' growth scores in a collection, for example, translates into a level 3 student growth rating, or student growth "at expectations."
- If the average growth score for a collection is 4.3, that would translate to a level 5 student growth rating because it shows a growth score of more than three and is thus "significantly above expectations."
- A collection average growth score of zero translates to a level 1 growth rating.

Score resolution

The growth level ratings for each portfolio collection are used to determine when additional peer reviews are needed for a collection.

If a collection's growth level rating resulting from a peer reviewer's scores is within one growth level of the rating that results from a teacher's self-evaluation scores, the scores are considered "in consensus" and the reviewer's score becomes final for that collection. For example, if a teacher self-scores student work that translates into a collection growth level of 2, and the peer reviewer's scores of that same work translates into growth level of 1, the peer reviewer's rating is within one level of the teacher's rating and is therefore "in consensus" and the portfolio growth level rating of 1 becomes final. An exact matching score is also "in consensus." If both the teacher and the reviewer score a collection so that the resulting growth levels are both a 2, that score becomes final.

If, however, the reviewer's rating differs from the teacher's rating by more than one growth level – for example, the reviewer's scores result in a growth level of 1 and the teacher's scores result in a growth level of 3 – the portfolio collection will be assigned to a second reviewer. If the second reviewer's scores result in a growth level rating that matches or is within one growth level of the first reviewer's rating or the teacher's rating, then the second reviewer's growth level rating becomes final. If the second reviewer's rating differs by more than one level from both the teacher's and the first reviewer's ratings, then the collection is assigned to an expert reviewer. Expert reviewers' ratings are final.

Final teacher portfolio scores

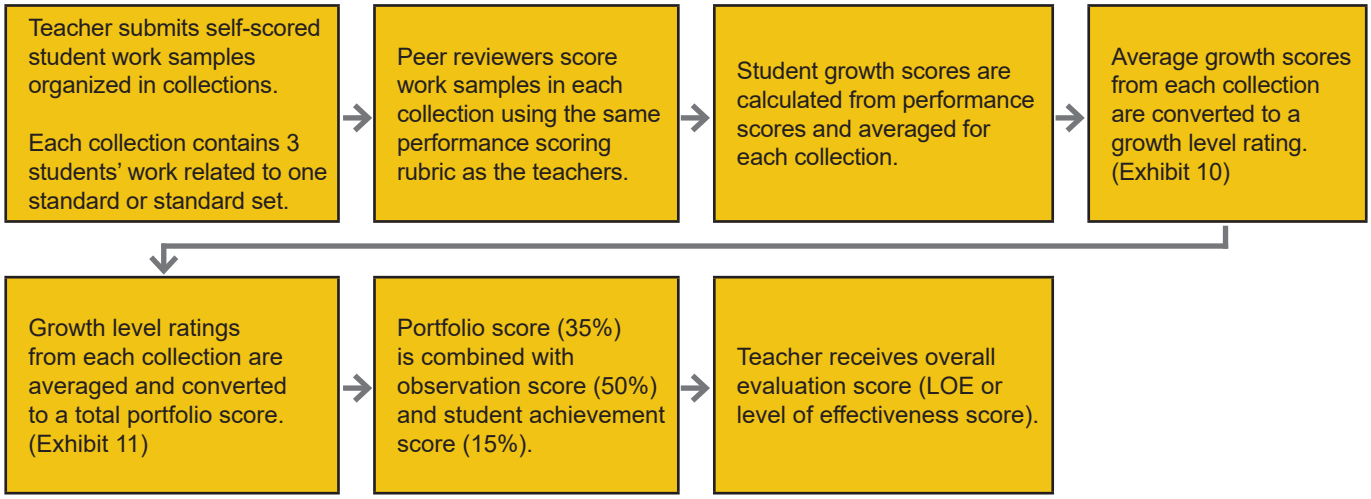
Once a final growth level rating for each portfolio collection is determined, the portfolio platform calculates the average growth level rating of teachers' four collections to produce an overall portfolio score, using the department's Teacher Effectiveness Indicator chart. (See Exhibit 11.) The total portfolio score becomes the student growth component of a teacher's overall evaluation (LOE) score and is weighted at 35 percent.

Exhibit 11: TDOE's Teacher Effectiveness Indicator Chart to calculate portfolio score

Total portfolio score	
Teacher Effectiveness Indicator	Average of growth level ratings from each portfolio collection
Level 1	1.00 – 1.79
Level 2	1.80 – 2.59
Level 3	2.60 – 3.39
Level 4	3.40 – 4.19
Level 5	4.20 – 5.00

Source: OREA adaptation of TDOE Teacher Effectiveness Indicator Chart.

Exhibit 12: Portfolio and overall evaluation scoring process



Costs associated with the portfolio process

Costs of the portfolio process include the contract for the online platform where teachers submit their portfolios and reviewers score the work, stipends paid to portfolio peer reviewers (2017-18 and 2018-19), as well as regional meetings convened for reviewers (2018-19). Costs are difficult to compare from year to year due to changes in portfolio requirements and implementation over time.

In the early years of the portfolio process, the Department of Education used the GLADiS Project platform and paid for the service through subscription fees. During the four-year period of 2013-14 through 2016-17, the department paid a total of $153,000 for the total 7,424 portfolios submitted during that period.[B] The department did not pay reviewers prior to 2017-2018; instead, districts recruited teachers to be reviewers and any compensation received by reviewers was determined at the local level.

[B] The department did not provide a breakdown of GLADiS payments by year, but across the four years GLADis was in use, a one-year average would have been $38,300 for approximately 1,900 portfolio submissions.

For the 2017-18 school year, the state approved a sole source contract with Educopia, a vendor that the state had already worked with to test a new scoring process. The initial contract with Educopia was amended twice to increase the state's financial liability, plus a subsequent short-term contract was approved. Increases to the state's contract costs resulted from higher district and teacher participation than expected, as well as from additional vendor support required to address several problems with the platform's implementation.

State payments to Educopia ultimately totaled $706,051 for work on the portfolio process for the 2017-18 school year, the same year that saw the number of teachers submitting portfolios rise from 2,170 to over 5,750.[11] Adopting a new platform administered by a new vendor the same year as this large-scale increase in portfolio submissions likely increased the amount and complexity of the problems encountered and, by extension, the amount paid by the state. When the $677,000 in stipends paid to portfolio reviewers is added to the platform contract costs, the resulting total of $1.38 million makes 2017-18 the most expensive year for portfolio implementation to date.[12]

The department released a request for proposal (RFP) for the 2018-19 school year, as it had planned, and awarded a contract to Portfolium, the only vendor other than Educopia that submitted a bid. The state signed a five-year, $2.1 million contract with Portfolium. In 2018-19, $216,496 was charged to the contract for the online platform, and $607, 282 was spent on portfolio review costs, primarily reviewer stipends.[13] One additional cost of $26,100 was paid in 2018-19 for stipends for portfolio consultants, teachers, and other educators contracted to provide feedback on revisions made to scoring rubrics for clarity.[14]

With 6,059 teachers submitting portfolios, the 2018-19 average state cost per portfolio was $140, not including compensation paid to three full-time department staff.[15] This figure also does not capture local district costs. Some districts, for example, pay for classroom substitutes so that teachers have time to complete their portfolios during the school day.

For comparison purposes, in 2018-19, Tennessee spent $2,008,720 for the state's contracted vendor to provide TVAAS scores for all public schools and districts, as well as for teachers whose students completed TN Ready tests. With 19,276 teachers receiving individual TVAAS growth scores that year, the cost per teacher was roughly $104.[16] This rough estimate overstates the cost per teacher since the state TVAAS payment covers the costs for all school and district growth data, not just individual teacher TVAAS scores.

District Compliance with State Portfolio Laws

Summary: School districts are currently in compliance with both state laws related to student growth portfolios – Public Chapter 703 (2016) and Public Chapter 552 (2018) – based on OREA's review of data available from the department.

Pre-k/kindergarten portfolio model requirements for districts awarded state VPK funds

In 2016, the state legislature passed Public Chapter 703, requiring all districts that receive approval for state funding of their Voluntary Pre-kindergarten (VPK) programs to use the pre-k/kindergarten portfolio model for their pre-k and kindergarten teachers' evaluation growth scores.[C] The intent of this requirement was to help improve the quality of pre-k programs following the 2015 release of a Vanderbilt University study that found initial benefits for children enrolled in pre-k programs were not detectable by the time the children had reached 3rd grade.[D]

[C] In 2017-18, the state funded about 935 VPK classrooms at a cost of $85 million. VPK state awards to districts ranged from $89,000 for one classroom in the Collierville municipal district to $10 million for 112 classrooms in Shelby County Schools. The median awards were $423,000 to $426,000 for four classrooms.
[D] The results of the study released in 2015 can be found at https://www.sciencedirect.com/science/article/pii/S0885200618300279. A follow-up study, released in summer 2019, found that academic gains in pre-k programs were sustained for students who had highly effective teachers and attended high quality schools in subsequent school years. For more information, see https://www.edworkingpapers.com/sites/default/files/ai19-85.pdf.

Under Tennessee's (VPK) program, local school districts apply for competitive state awards to offer pre-k programs in their districts. Districts were required to begin using the pre-k/kindergarten portfolio model in school year 2017-18, and the VPK grant applications began including criteria for teachers to implement portfolio models, among other quality improvements.

- In 2017-18, 138 districts received VPK state funding. All 138 were confirmed by the department, through its evaluation score data, to have used portfolio scores for their pre-k and kindergarten teacher evaluations.
- Portfolio score files were available for 2018-19 and OREA confirmed that all 137 districts receiving VPK funds in that year had pre-k and kindergarten teachers submit portfolios.[E]
- For the current 2019-20 school year, the 137 districts that have accepted VPK funds have indicated to the department that they plan to use the portfolio model for pre-k and kindergarten teachers as required.[F]

Districts required to use at least one portfolio model in 2019-20

All public school districts in the state must use at least one portfolio model for teacher evaluations beginning in the 2019-20 school year.[G] Most districts already use the pre-k/kindergarten portfolio model and thus meet this requirement. The five districts not already using the pre-k/kindergarten model to meet VPK requirements must adopt a portfolio model. Of these five districts, Sevier County plans to use the pre-k portfolio model; Arlington City, the fine arts model; Germantown City, the physical education (K-5) model; Lakeland City, the world languages model. The fifth school district, Carroll County, is exempt from the requirement because of its unique nature.

Effectiveness of Portfolio Models

Summary: Any judgment of a tool's effectiveness must account for the tool's purpose. An effective tool for one purpose may not be equally effective when applied to another. Originally designed as a way to assess students' growth for teacher evaluations, student growth portfolios have also been cited as a tool teachers can use to reflect on and improve their teaching practice, as a way to help improve pre-k quality, and as a method to help the state meet its 3rd grade reading goals. This report, as requested, focuses primarily on the effectiveness of using portfolios in the state's teacher evaluation system as a measure of student growth.

Portfolio models have several design factors that make them less valid and reliable measures of student growth than more standardized measures, such as TVAAS scores calculated from TN Ready tests administered to all students in a teacher's classroom.[H] Yet other features of the portfolio model process may help improve teachers' instructional practice and provide more detailed information about teachers' skills and professional development needs.

Teachers with portfolios were considerably more likely in 2018-19 to score in the top student growth level than teachers with individual or school level TVAAS growth scores. These score level results, together with the lack of standardized student tasks to assess growth, the continued expansion of rubric scoring levels, and the possible use of work samples from as few as three students to judge the instruction of an entire class reduce the validity of portfolios as an effective quantitative measure of student growth. OREA analysis found relatively low interrater scoring agreement, especially for some standards within some models. Scoring procedure changes over several years limit the usefulness of year-to-year portfolio score comparisons.

[E] In 2018-19 five districts did not receive VPK awards, including Arlington, Germantown, and Lakeland municipal districts in Shelby County and the Carroll County school district (which only provides vocational and special education programs for the other districts in Carroll County). Although Sevier County was approved for two VPK classrooms, it chose not to accept the VPK funds and thus was not required to implement the pre-k/kindergarten portfolio model.
[F] Sevier County school district chose to forgo state funding for VPK again in 2019-20 and thus is not required to adopt the pre-k/kindergarten portfolio model under Public Chapter 703.
[G] Public Chapter 552 (2018) requires each district to use at least one "alternative growth model" beginning in 2019-20. The only alternative growth models approved by the State Board of Education as of December 2019 are portfolio models so all districts must adopt at least one portfolio model.
[H] TVAAS is a statistical model that calculates student learning growth from one school year to the next, based on scores from standardized tests given to almost all the students of a category of teachers (primarily 3-8 grade classroom teachers and teachers in selected high school subjects).

Expanding portfolio goals

The goals behind portfolio models have expanded since their original introduction. Starting with providing more teachers an individual growth score for their teacher evaluation to helping improve pre-k program quality to providing support for the state's 3rd grade reading goals, the expectations around the purpose of portfolios have grown. Judging the effectiveness of portfolio models depends on the goals they are expected to fulfill.

Development of portfolio models was spearheaded in Tennessee by teachers in non-tested subjects to obtain an individual growth score for their teacher evaluations. These teachers saw portfolios as a more accurate assessment of their instruction because they are based on teachers' individual contributions to student learning, in contrast to school level, composite growth scores, which are based on academic growth of whole groups of students, some or many of whom an individual teacher may never have taught.

The department also recognized early on that portfolios could serve as a professional learning opportunity for "purposeful reflection on instructional practice," as well as a chance for teachers to collaborate and share best practices. The department reports that educators cite professional learning as an added benefit to using portfolio models.

Portfolios were optional for districts until the legislature's decision to require districts to evaluate their pre-k and kindergarten teachers with the portfolio model as part of the Pre-K Quality Act's provisions to increase accountability in the VPK program and improve teacher and program quality.[1] (For more about what prompted the Pre-K Quality Act, see the previous report section, "District Compliance with State Portfolio Laws.") The application of this law in the VPK program emphasized portfolio use as a professional development tool, rather than a measure of student growth. In the VPK grant application, districts are asked to describe their "comprehensive plan for ongoing training and support of pre-K teachers in best practices to support successful portfolio implementation."[17]

One month before the legislature passed the Pre-K Quality Act, the department launched its Read to be Ready initiative to increase the percentage of 3rd graders reading on grade level to 75 percent by 2025. The focus on improving 3rd grade reading proficiency involved multiple programs, one of which sought to improve how literacy was taught in the early primary grades. Both Read to be Ready and VPK quality improvements were part of the department's goal of creating an early learning continuum of success for students in pre-k through 3rd grade.

Portfolio models now exist for 1st grade and 2nd grade teachers, in addition to pre-k and kindergarten teachers. Portfolios have been cited as a tool to help schools and districts meet 3rd grade reading goals, through increasing the use of good instructional practices.

A department survey of teachers who submitted portfolios in 2017-18 asked to what extent teachers saw connections between their focus on student work for portfolios and their Read to be Ready work. About 39 percent of responding pre-k and kindergarten teachers reported seeing "some" or "many" connections, and 52 percent of 1st grade teachers reported such connections. Some responding teachers commented that they were not familiar with Read to be Ready, either because they were pre-k teachers or because the program had not been implemented at their school.

The effectiveness of portfolio models as a quantitative growth measure in teacher evaluations is discussed in depth below. The effectiveness of portfolio models in improving pre-k quality and reaching 3rd grade reading goals is not addressed in this report. The links between teachers using portfolio models and achievement of these broader goals are based on the expectation that the steps to successfully complete a portfolio will increase

[1] Because the portfolio model was already established as a combined pre-k and kindergarten model, kindergarten teachers were automatically included.

teachers' knowledge of key academic standards, teachers' focus on helping students at different proficiency levels achieve growth, and teachers' collaboration with colleagues, all of which will result in more effective teachers and more student learning. Using portfolios as a required evaluation component places high-stakes consequences on portfolio models, which can reduce their validity as an objective measure of student growth, and may also reduce their usefulness as a professional development tool.[18]

Analyzing portfolio models for validity and reliability

Student growth portfolios are a two-part measure when used in teacher evaluation because first, they have to assess student learning growth, and second, they have to convert student growth into a teacher growth level score included in the overall evaluation. OREA focused its analysis using two sets of questions:

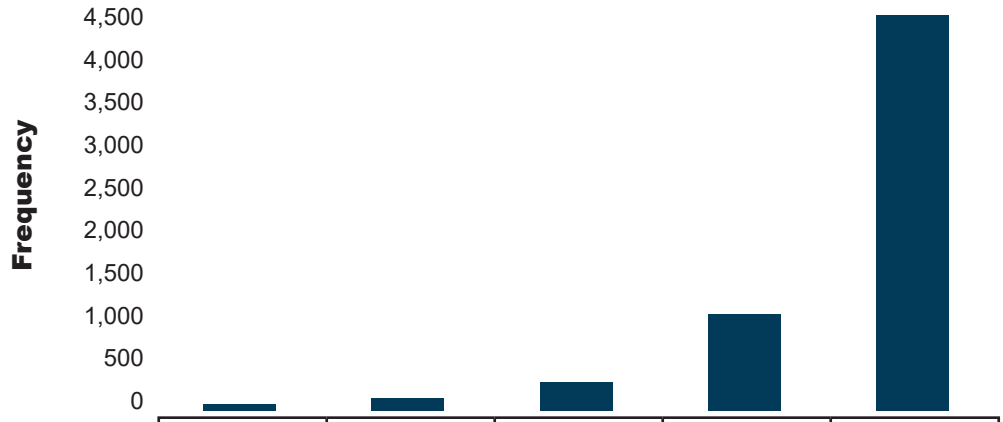
1. (a) Are portfolio models a valid way to measure student growth?
   (b) Are the scores generated from the portfolio model a valid student growth component for teacher evaluations?
2. (a) Are portfolios a reliable measure of student growth?
   (b) Are the scores generated from the portfolio model reliably translated into a student growth component for teacher evaluation scores?

The scoring results from portfolio models, presented in the next section, and the discussions of validity, reliability, and repeatability that follow, address these research questions.

Scoring results of individual growth components

Although only 9 percent of evaluated teachers statewide received growth scores based on portfolios in 2018-19, they were almost four times more likely than teachers receiving individual TVAAS growth scores (about 29 percent of evaluated teachers) to receive a top growth score level of 5. Teachers with individual TVAAS scores outnumbered those with portfolio growth scores (19,000 to 6,000), but top-scoring portfolio teachers outnumbered top-scoring TVAAS teachers by about 750. Portfolio growth scores are weighted in teacher evaluations as individual growth scores and, like individual TVAAS scores, count for 35 percent of a teacher's overall evaluation score. The higher likelihood that teachers using portfolios will earn a top growth level score than teachers using individual TVAAS scores does not, in itself, indicate a problem with the validity or reliability of the scores. But the comparison to TVAAS scores does raise a question as to whether pre-k and kindergarten teachers, the vast majority of portfolio-scored teachers, are significantly more effective in achieving student growth than teachers in tested grades. (See Exhibits 13, 14, and 15.)

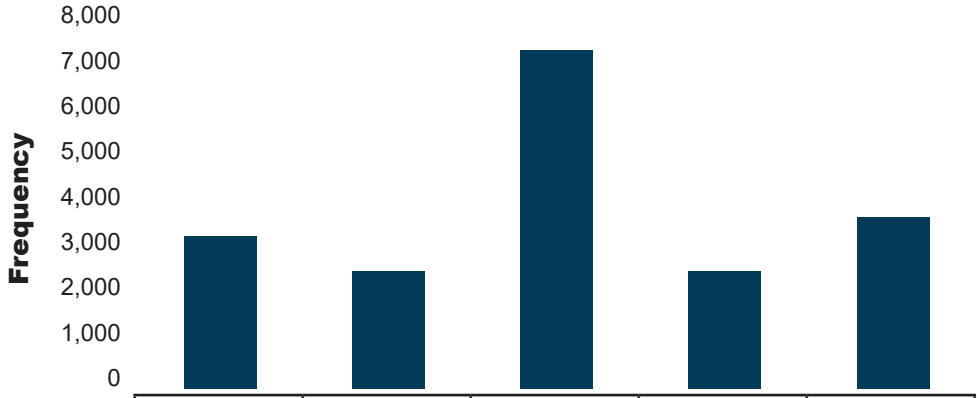
Exhibit 13: Portfolio score distribution, 2018-19



	1	2	3	4	5
Teachers	54	115	381	1,010	4,499
Percent	0.9%	1.9%	6.3%	16.7%	74.3%

Portfolio score

Exhibit 14: TVAAS one-year individual growth score distribution, 2018-19



	1	2	3	4	5
Teachers	3,259	2,638	7,206	2,420	3,753
Percent	16.9%	13.7%	37.4%	12.6%	19.5%

TVAAS individual score

Exhibit 15: Teacher individual growth score levels based on TVAAS and portfolio models, 2018-2019

Growth score level	TVAAS teachers	Percent of TVAAS teachers	Portfolio teachers	Percent of portfolio teachers
Level 1	3,259	16.9%	54	0.9%
Level 2	2,638	13.7%	115	1.9%
Level 3	7,206	37.4%	381	6.3%
Level 4	2,420	12.6%	1,010	16.7%
Level 5	3,753	19.5%	4,499	74.3%
Total	19,276	100%	6,059	100%
Average score	3.04		4.61	

Source for Exhibits 13-15: Tennessee Department of Education portfolio and TVAAS score data, 2018-19.

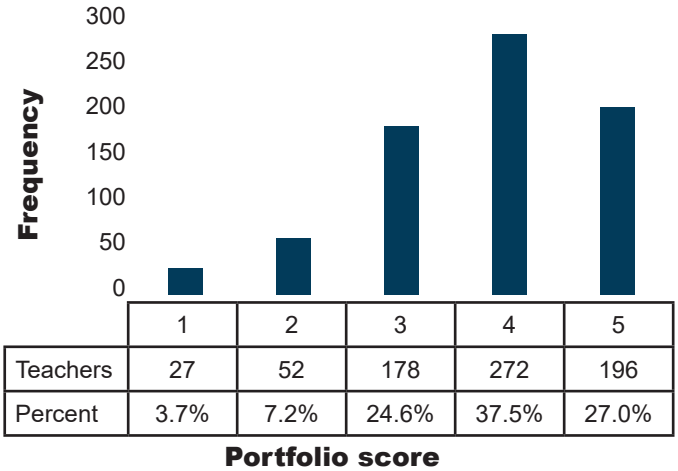
For 2018-19, the average portfolio score for all teachers was 4.61. Teachers in grade-level models (pre-k/kindergarten, 1st, and 2nd grades) scored higher on average, with a 4.73, than those in subject-based models (fine arts, physical education, and world languages), with an average of 3.77. (See Exhibit 16.) Kindergarten and 1st grade teachers produced the highest average portfolio scores, while fine arts and world languages teachers received the lowest. OREA did not find evidence suggesting that teachers in grade-level models are more effective in achieving student growth than those in subject-based models, but did find more scoring agreement among reviewers of grade-level models than subject-based models (see more about scoring reliability, pages 30-33).

Other factors that may impact the variations by model could include:

- ability to identify growth more easily for math and ELA standards assessed in early grades than for performance standards in fine arts, physical education, and world languages,
- more clarity in early grades scoring rubrics, or
- more administrative support or collaboration with colleagues in early grades in compiling a portfolio.

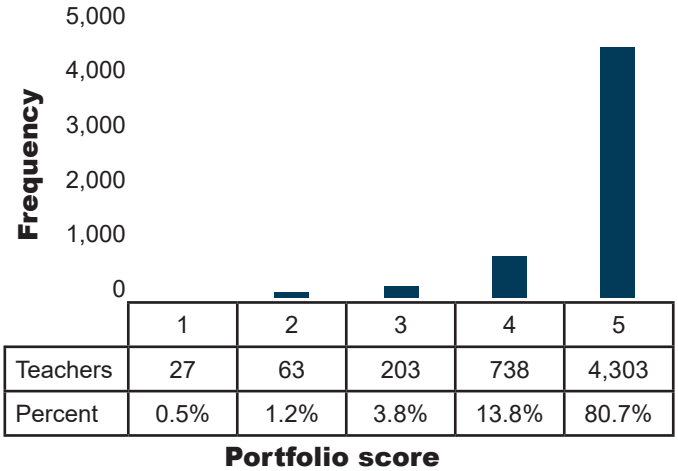
Exhibit 16: Portfolio score distribution, by type of model, 2018-2019

Subjects include fine arts, physical education, and world languages.



	1	2	3	4	5
Teachers	27	52	178	272	196
Percent	3.7%	7.2%	24.6%	37.5%	27.0%

Grade levels include pre-k and kindergarten, 1st grade, and 2nd grade.



	1	2	3	4	5
Teachers	27	63	203	738	4,303
Percent	0.5%	1.2%	3.8%	13.8%	80.7%

Source: Tennessee Department of Education, portfolio score data, 2018-19.

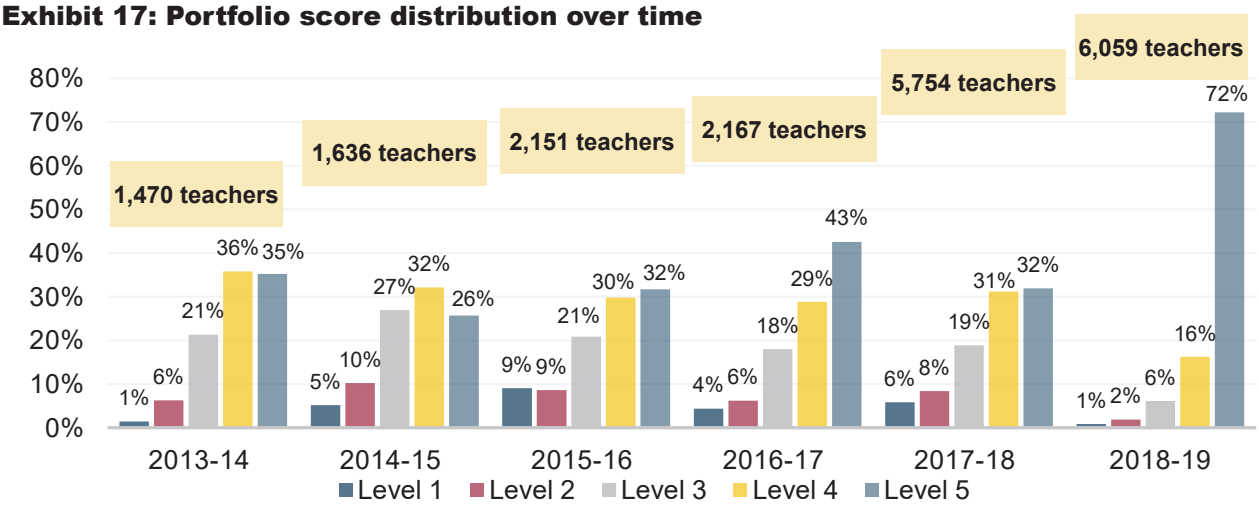
While large numbers of teachers could conceivably achieve top scores on a rating of demonstrated instructional skills, it is less likely that large numbers of teachers will achieve top scores on a measure of student growth levels if the assessment instrument is well-calibrated. Recall that overall portfolio scores are generated from the student growth level ratings assigned to each collection, and that a top growth level rating of 5 means the students represented in the portfolio collection achieved an average academic growth equivalent to **three or more** times the growth expected for a student in a particular grade or subject. A collection growth level rating of 4 means the students, on average, increased their academic performance the equivalent of two or more times (up to three) the expected growth. A growth level rating of 3 is the expected standard for teachers to meet, representing one academic year of growth.[J] (See Exhibit 10 for student growth level rating chart.)

The distribution of portfolio scores has not always skewed to the top score as significantly as it did in 2018-19. As shown in Exhibit 17, the distribution of scores in past years was more differentiated, such as in years 2014-15 through 2016-17. (Note that prior to 2017-18, teachers with incomplete portfolios were automatically given a score of 1, which may have inflated the percentage of teachers at level 1 in these past years. Similarly, prior to 2018-19, portfolio collections with any unscoreable work were automatically given a score of 1, rather than receiving scores based on the scoreable portion of work. In addition, the department indicates that 2017-18 score data is less reliable due to technical glitches with the online platform that year.) As shown in Exhibit 16, the grade level models appear to drive the skewing of the score results more than the other subject models. Decreased differentiation between growth scores suggests that portfolio models may not provide useful information about a teacher's ability to increase student learning relative to other teachers.

OREA could not determine one single cause for the jump in top scores in 2018-19. Several factors impacting the validity and reliability of portfolio models may have partially contributed and are discussed in separate sections below. Teachers' increased familiarity with both submitting and reviewing the portfolio model over the past several years may partially explain the improved scores.

[J] In order to generate an overall portfolio score of 5, as 74 percent of teachers did in 2018-19, teachers must earn a growth level rating of 5 on at least one portfolio collection and all 4s on the other three collections, or a growth level rating of 5 for two portfolio collections, a rating of 4 for another collection, and a rating of no less than 3 for the remaining collection.

Exhibit 17: Portfolio score distribution over time

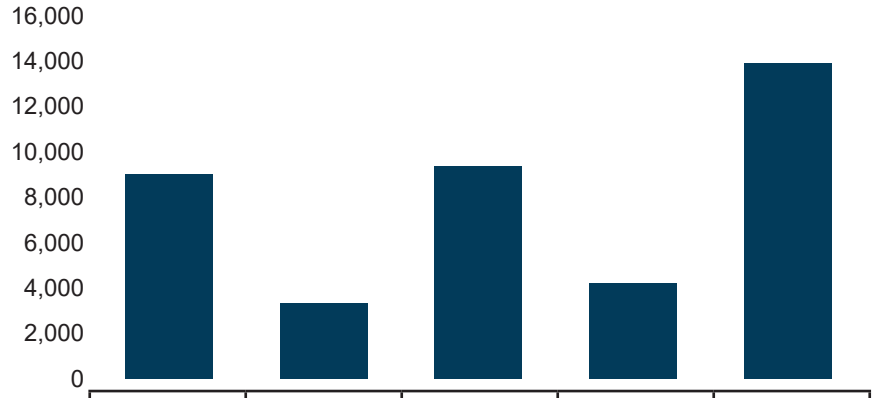


Notes: 2017-2018 and 2018-2019 had 4% and 3% incomplete, respectively. The score distribution for 2018-19 shown here varies slightly from the score distribution shown in Exhibit 15 because they were based on different TDOE data sets that treated incomplete and non-scoreable portfolio submissions differently.
Source: Tennessee Department of Education, portfolio score data, 2013-14 through 2018-19.

For teachers with neither an individual TVAAS score nor an individual portfolio score, a school level TVAAS score is used for the student growth component in their evaluations. Instead of the 35 percent weighting used for individual growth scores, school level scores are weighted at 15 percent of a teacher's total evaluation score, with the difference of 20 percent added to the teacher's classroom observation score weight.

School level TVAAS scores include various combinations of students' growth scores from multiple classrooms across grades and subjects. The 2018-19 distribution of school level TVAAS growth scores shows a similar pattern to individual TVAAS scores. (See Exhibit 18.) While a higher percentage of teachers receive the top growth level score with a schoolwide TVAAS than with an individual TVAAS, the percentage of top growth level teachers with school TVAAS scores is less than half the percentage of top growth level teachers with portfolio scores.

Exhibit 18: Distribution of growth component scores for untested teachers using school level TVAAS scores, 2018-19



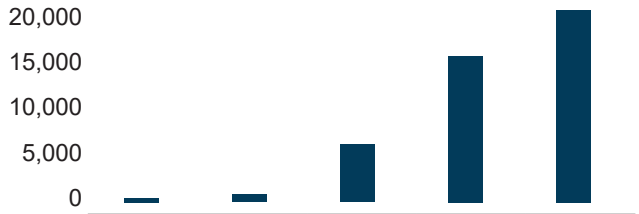
	Level 1	Level 2	Level 3	Level 4	Level 5
Teachers	9,187	3,620	9,343	4,633	13,771
Percent	22.7%	8.9%	23.0%	11.4%	34.0%

Source: Tennessee Department of Education score data, 2018-19.

Untested teachers without portfolio scores have a higher weight placed on classroom observation scores. The graphs below show little difference in the observation scores for any teachers – tested or untested, whether using portfolios or school level scores.

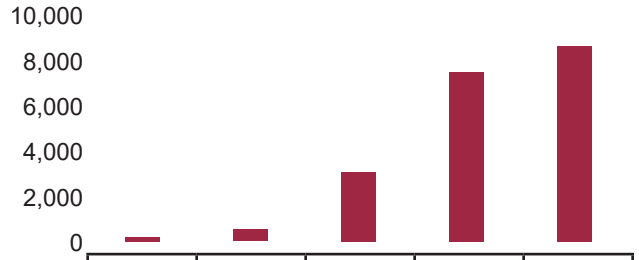
Exhibit 19: 2018-19 Classroom observation component scores for teachers

A. Without individual growth scores



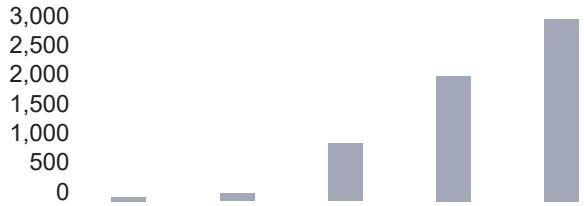
	Level 1	Level 2	Level 3	Level 4	Level 5
Teachers	41	492	5,582	15,633	18,818
Percent	0.1%	1.2%	13.8%	38.5%	46.4%

B. With individual growth scores from TVAAS



	Level 1	Level 2	Level 3	Level 4	Level 5
Teachers	15	345	2,830	7,741	8,345
Percent	0.1%	1.8%	14.7%	40.2%	43.3%

C. With individual growth scores from portfolios



	Level 1	Level 2	Level 3	Level 4	Level 5
Teachers	6	87	867	2,165	2,703
Percent	0.1%	1.5%	14.9%	37.2%	46.4%

Source: Tennessee Department of Education score data, 2018-19.

Teachers with individual TVAAS scores had slightly lower classroom observation scores than non-tested teachers, with about 3 percent fewer teachers receiving the top growth level rating of 5 than teachers with portfolio or schoolwide growth scores. Comparing teachers earning the top two growth levels (4 and 5), shows a difference in observation scores of only 1 or 2 percent, regardless of the growth measure used.

Validity

The validity of an assessment refers to the accuracy of the assessment in measuring what it is intended to measure. For example, a written driving test is a better measure of knowledge about rules of the road than it is a measure of driving ability. The portfolio model therefore should accurately assess students' growth on standards.

Task assignments for collecting student work

Portfolio models, like the TN Ready tests, are intended to measure student performance, from which a student growth score is later calculated. One main difference between portfolios and TN Ready tests is the lack of consistent student assignments for portfolios. These assignments, which are used to measure student performance, are equivalent to the test questions on TN Ready tests. Whereas all students taking a 5th grade math TN Ready test would be scored on the same questions across school classrooms and districts, students given a 5th grade fine arts assignment for student growth portfolios would be scored on the resulting work samples that vary from teacher to teacher and district to district. Some tasks may be better at showing student knowledge of standards or may be better aligned to the standards than others, and some teachers may be better at designing such tasks. A department survey of 2017-18 portfolio users found that 71 percent of responding teachers wanted more support on creating student tasks aligned to the standards and scoring rubrics.

The methods for documenting and collecting student work samples also vary by teacher and district; scans or photos of student papers and projects, audio or video recordings of student presentations, conversations, skills demonstrations, or other activities, are all acceptable to upload on the portfolio platform.

The portfolio guides provided by the department do not set standard assignments or tasks to generate comparable student work samples for all teachers using a portfolio model and do not typically provide examples of appropriate, standards-aligned tasks. Some sample tasks were presented for the pre-k model standards assessed in ELA in 2017-18, but current model guides do not include such samples. Portfolio training provided by the department has included limited examples of student work samples.

Some portfolio guides include tips for the format of work samples for selected standards. For example, pre-k and kindergarten teachers collecting student work for the ELA standards are told:

> When scoring the reading comprehension standard, an oral response about the written piece can be used in conjunction with the written piece to determine the performance level on the reading scoring rubric. Because a primary student's ability to orally retell often surpasses the student's proficiency in expressing those thoughts in written form, evidence such as audio, video, or dictation can also be submitted.[19]

The world languages scoring guide indicates that "best scores are yielded from teacher interviews of students" for one standard, and for another standard, "writing samples [rather than oral samples] have been the best presentational evidence due to the limited speaking skills at the novice proficiency levels."[20]

The department's portfolio resources indicate that the models are designed to give teachers flexibility to use authentic student work from their own classrooms. Portfolio models are labeled as a method for teachers to "demonstrate student growth," and are "not state assessments," yet the state's teacher evaluation policies use portfolio scores as the quantitative student growth component of teachers' evaluations, to stand in place of individual or school level TVAAS student growth scores.[21]

Although there is little research on the use of portfolios in teacher evaluation systems, a few research discussion papers on student growth portfolios suggest that, when used primarily for assessment rather than instructional purposes, portfolios should have more uniformity in the items included in the portfolio, and that high-stakes use of portfolios (such as in teacher evaluation) requires a high degree of specification about the contents of the portfolio to ensure a fair and valid evaluation.[22] (See "Student Growth Portfolios in Other States," on page 37 for more information about the use of student growth portfolios for teacher evaluation outside of Tennessee.)

While Tennessee's portfolio models provide teachers flexibility in determining how best to show their students' learning, their non-standardized design used in a high-stakes setting (teacher evaluations) decrease the content-related validity of portfolios as a measure of student growth. Under the current structure used in Tennessee, portfolios may be more a measure of how well a teacher has designed the tasks that document student learning than a measure of students' actual growth. A teacher's ability to design tasks that are closely aligned with course standards and that allow students to show their full range of learning could be a useful skill to help teachers improve instructional practice; however, such ability at measuring learning is not necessarily correlated with facilitating learning.

Alignment with other evaluation measures

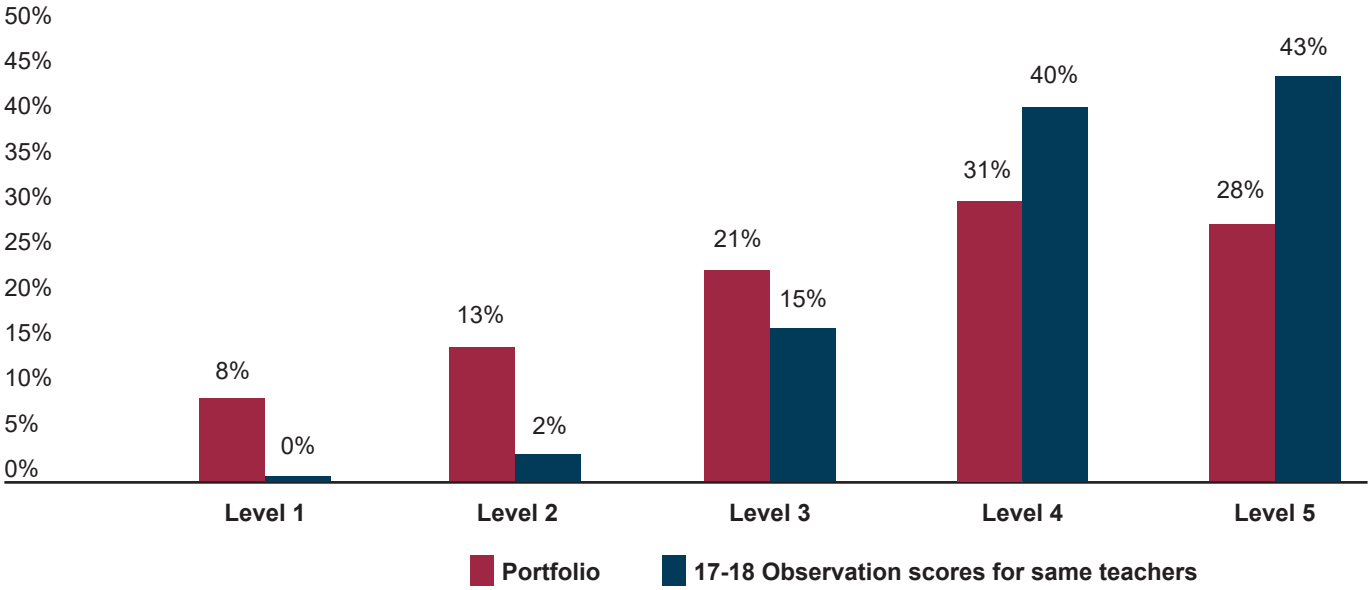
The department compared different evaluation components in the early years of portfolio use and found that teachers' portfolio scores were "well-aligned" with their classroom observation scores.[23] Several steps toward producing a quality student growth portfolio are aligned with several components of the TEAM classroom observation checklist, such as planning and instruction around state standards, student writing assignments that require application of skills, and differentiated instruction. In an early analysis, the department found that teachers using portfolio models scored slightly higher in their classroom observations on three specific instructional areas, and thus on their overall observation score, than other teachers of untested subjects who did not use portfolios.[K] The department concluded that "participating in the portfolio process leads to improvements in key teaching practices."[24] While portfolio use may improve teachers' instructional practice as measured by classroom observation scores, that does not establish a correlation between portfolio scores and student growth, which is the stated purpose of portfolio scores in teacher evaluations.

The department also compared portfolio scores to schoolwide TVAAS growth scores and found approximately two-thirds of teachers who used portfolios achieved the same or higher growth level score than they would have using the schoolwide growth score. In the same January 2017 analysis, the department noted that teachers tended to score lower on the newer portfolio models.

[K] The three instructional areas were activities and materials, teaching types of thinking (analytical, practical, creative, etc.), and teaching types of problem-solving (abstraction, categorization, observing and experimenting, etc.).

In 2017-18, when many teachers were required to use the pre-k/kindergarten portfolio model for the first time, the department compared the distribution of teachers' 2017-18 portfolio scores to their classroom observation scores. That comparison found more differentiation among teachers based on portfolio scores than observation scores, with more teachers receiving low portfolio scores. (See Exhibit 20.)

Exhibit 20: Pre-K and kindergarten teacher portfolio scores compared to observation scores, 2017-18



Source: TDOE presentation to the Portfolio Review Committee, July 23, 2019.

By the next year (2018-19), however, the portfolio score distributions for pre-k and kindergarten teachers had shifted significantly. The percentage of portfolio scores at levels 1 and 2 combined totaled about 1.7 percent in 2018-19. Observation scores of all teachers with portfolios (not just of pre-k and kindergarten teachers) at levels 1 and 2 combined were about 1.6 percent, meaning the portfolio scores no longer provided more differentiation among teachers than observation scores. This may be partially explained by increased familiarity with the pre-k/kindergarten model on the part of teachers.[25]

Rising scores in the years following the introduction of a new assessment model is a pattern that researchers have identified with standardized tests.[26] The scoring results from the current school year, 2019-20, will be useful in determining whether 2018-19's large percentage of top-scoring teachers is a pattern that will continue or whether it was an anomaly.

Rubrics for scoring student work samples

Since the first step of portfolio models is to obtain a measure of student growth, the accurate rating of students' work samples at point A and later in the year at point B to determine the difference between the work samples is crucial. Scoring rubrics, provided by the department, are aligned to Tennessee standards for the subject and grade level. Initially, the rubrics set performance scores for each student work sample on a 1 to 5 scale, with 3 representing work at grade level, 1 and 2 representing work below grade level, and 4 and 5 work above grade level.[L] The difference between scores of a 3 and a 2 or a 5 and a 4 are both 1.0 and represent the same amount of student growth.

Research on scoring rubrics suggests that having an arbitrarily set number of levels may force raters to make artificial distinctions when there are none, or cause them to fail to differentiate between key differences.[27] Since the rubrics involve different grade levels and subjects, it is not a given that all student work will naturally fall into five categories of performance. There is also evidence that overly complex criteria may induce raters to trust their overall impressions of a work sample.[28] That noted, the original five scoring categories do align to all

[L] The world languages model is an exception, with a scoring range of 1-9.

other teacher evaluation categories; teachers' overall level of effectiveness (LOE) scores, classroom observation scores, and individual or schoolwide TVAAS scores are all calculated on five-point scales.

The scoring rubrics have expanded beyond a five-point scale in recent years, mainly in response to teacher requests. In 2017-18, scoring levels of 6 and 7 were added to all portfolio scoring rubrics, primarily intended for the rating of students' second work sample (point B). The scoring levels were added to measure the growth of students who started the school year performing above grade level expectations, at a performance level of 4 or higher.

When comparing the portfolio score distributions in 2017-18 to those of prior years, there was not an apparent increase in student growth scores from the addition of the two top rubric score levels. OREA's analysis of 2018-19 scoring data for submitted portfolios found that 74 percent of teachers self-scored their student work samples using a performance score of 6 or 7 at least once. Peer reviewers or expert reviewers assigned final performance scores including at least one 6 or 7 to 84 percent of the portfolios.[29] Of those portfolios that incorporated at least one 6 or 7 score from reviewers into portfolio score calculations, 12 percent resulted in a higher overall portfolio score than would have otherwise been calculated with the previous maximum performance score of 5. Of these portfolio scores that increased from the use of a 6 or 7, the vast majority increased by one score level (99.5 percent).[30]

After scoring categories of 6 and 7 were added, early grades teachers then suggested that additional scoring categories were needed below the existing level 1 to reflect student work at the beginning of the year (point A) that was substantially below grade level. Such students might have the same span of growth as others, but still complete the year below grade level. For school year 2019-20, the scoring rubrics for only early grades portfolio models (pre-k/kindergarten, 1st grade, and 2nd grade) were revised to include a performance level 0.

The expansion in the range of performance scores for point A and B work samples has the potential to increase student growth scores without necessarily increasing actual student growth. Although teachers and reviewers use scoring rubrics, which describe the attributes of student work at each performance score level, the determination of which score level best matches the work still requires some judgment, and may be more subjective for certain standards and models. (See more about interrater scoring agreement page 30.) The creation of additional performance score levels provides teachers and reviewers more scoring choices, which seems likely to increase the number of growth scores that exceed the 1.0 growth expected for one year. The effects of adding the score level 0 will not be seen until the growth scores for school year 2019-20 are calculated in the summer of 2020.

Purposeful sampling to determine which students' work is included in the portfolios

Assuming expanded scoring levels produce accurate performance ratings to calculate student growth and that student assignments have produced comparable work samples on which to score student performance, another question about the validity of portfolio models is whether the growth scores for selected students reflect average class growth.[M]

The department's portfolio guides explain that teachers should select a *purposeful sample* of work samples from three students, each representative of the emerging, proficient, and advanced student performance tiers, for each standard assessed through the portfolio model.

Generally, portfolio models require student work for four standards be submitted; three students (one at each performance level) for each of the four standards means work samples may be drawn from up to 12 students. Department guidelines do not require 12 students to be selected. The guidelines allow the same student's work to be used in more than one collection. Some teachers may use work from the same three students (emerging, proficient, and advanced) for all four collections of their portfolio, in which case the teacher's individual

[M] Maximum class size limits set by law (TCA 49-1-104) are 25 for kindergarten through 3rd grade and 30-35 in other grades. Voluntary pre-kindergarten class size limits are set at 20 students in the state agreements required to receive the state funds.

growth score is based on three students. The extent to which this practice occurs is unknown, however. Since portfolios are scored by collection and teacher submissions are confidential, neither reviewers nor department staff can document how many students are represented within one teacher's portfolio.

Questions have been raised as to whether work from three students can adequately measure a teacher's impact on an entire class. For comparison, teachers' individual TVAAS growth scores require a minimum of six full-time equivalent (FTE) students assigned to a teacher, with at least six individual students for teachers in grades 3 through 8 and at least 10 individual students for high school grades.

Purposeful (or purposive) sampling, sometimes called subjective sampling, in social science is used when probability or random sampling is not practical. As the term implies, subjective sampling is considered less accurate and rigorous than probability sampling. Tennessee's portfolio models rely on *nonproportional quota sampling*, a type of purposive sampling in which a minimum number of sampled units is required for specified characteristics or categories, in this case one student each in the performance tiers of emerging, proficient, and advanced.

Work in Maryland to develop student growth portfolios for teacher evaluation, which occurred at about the same time as Tennessee's development, recommended that work by the same student not be used more than once in a portfolio. A Maryland school district that piloted the model required four different students for each of the four measures in the portfolio, for a total of 16 students. Later guidelines for districts revised this to "strongly recommend" that different students be used for each portfolio measure.[31]

Tennessee's process of asking teachers to divide a class into three performance tiers and select students from each tier helps ensure that a range of student ability levels is selected. Classifying students into tiers, however, is a subjective determination made by each teacher, with few guidelines and no oversight. The department's guidance recognizes the subjective nature on which these determinations may be made:

> Point A student work . . . should be scored by the teacher and categorized as emerging, proficient, and advanced. While teachers have flexibility in defining these groups, proficient typically refers to artifacts scoring at Level 3, which is the grade-level standard. Emerging typically refers to performance levels below 3, and advanced typically refers to performance levels above 3. However, point A student work sometimes demonstrates limited variance in performance levels across a cohort of students. In these cases, teachers should use their knowledge of students, task-specific expectations, and other assessment data to categorize student work.[32]

Dividing students among three performance tiers is particularly complicated for ELA collections in the early grades because each standard set in ELA is a combination of three standards on which a single student may have varying proficiency.

> Because collections in the ELA portfolio include multiple standards, teachers should group students into an overall category of emerging, proficient, or advanced based on the collective performance on the three standards (foundational, reading, and writing). This categorization should be determined by totaling the three individual scores and using that total . . . to rank order the artifacts. At this point, the teacher observes patterns that indicate emerging, proficient, and advanced differentiated groups.
>
> Of course, teachers have the flexibility to use their own professional judgment to make adjustments during categorization by considering their knowledge of students. For example, a student in the emerging category based on the rank order might be sorted into the proficient category based on factors such as universal screening data, etc.[33]

Teachers may use a limited number of students for their portfolio due to various reasons. Some students transfer or move in or out during the school year and thus do not have both point A and B work samples

available; there may be fewer students in one of the three performance level groups from which to select (for example, there may be only two advanced students in the class, so work samples for the advanced performance level on the four portfolio standards assessed would need to be selected from those two students). Further, some teachers have cited the time burdens of collecting portfolio work; collecting from fewer students may be perceived as less burdensome. Given the high-stakes use of portfolio scores in teachers' evaluations, setting a higher minimum number than three students to include in this growth measure and developing a method to monitor the number of students actually represented by each portfolio would increase validity.

Reliability and repeatability

Reliability refers to consistency in measurement. There are several types of reliability, often associated with large standardized tests. Two types of reliability are of interest for assessments where reviewers use rubrics to assign scores. Interrater reliability is the consistency of scores among multiple reviewers of the same portfolio. In Tennessee's portfolio process, at least two people review each portfolio collection, the submitting teacher and the first peer reviewer. If their scores are not in consensus, a second peer reviewer and potentially an expert reviewer may also score the same work, meaning some collections may have four reviews.

Intrarater reliability is the consistency of one reviewer in scoring multiple portfolio collections according to the rubrics. In the portfolio process, each reviewer is generally expected to score 40 collections, equivalent to about 10 completed portfolios.

Although the department's current vendor contract with Portfolium for the online platform includes a provision for interrater reliability statistics, this provision has not been activated by the department and no reliability statistics (either inter- or intra-) have been collected by the vendor or the department.

Research suggests that clear, well-defined scoring rubrics and reviewer training are methods to increase both kinds of scoring reliability.[34] The department has revised scoring rubrics for portfolio models after the state's academic standards are updated, but small revisions to the rubrics to improve clarity are also made between standards updates. Scoring accuracy is generally higher when reviewers understand the content of the work they are assessing. Because peer reviewers are fellow Tennessee teachers in the same grade or subject area as the teachers submitting portfolios, reviewers are well placed to understand the work they are scoring. A report on a pilot student growth portfolio program in Maryland concluded that interrater reliability should be confirmed before portfolio scores were used for high-stakes purposes.[35] TDOE has not collected rater reliability statistics.

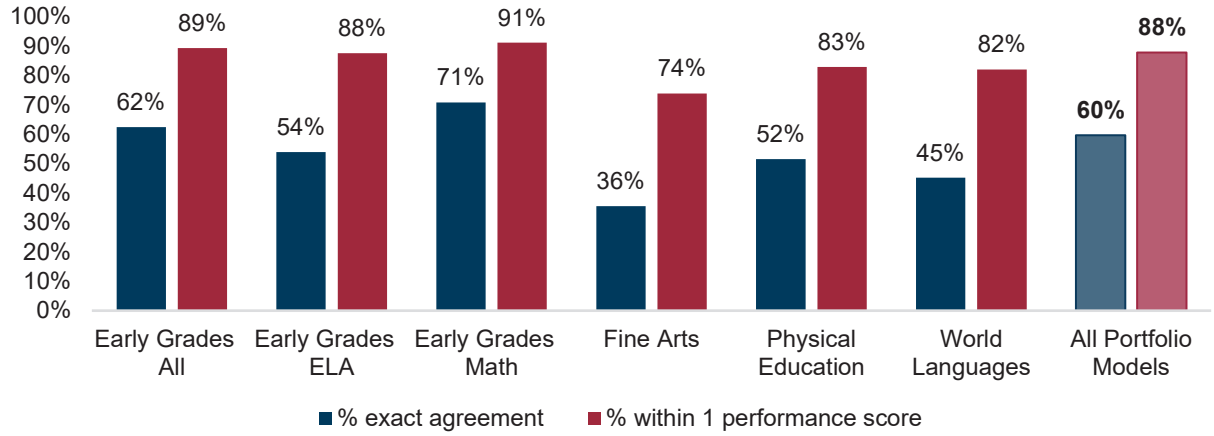
OREA's analysis of interrater reliability agreement found low rates of exact agreement, although some models had stronger agreement than others. The reliability of Tennessee's portfolio model may be impacted by certain scoring procedures, such as the use of peer reviewers who are also submitting teachers. Short-term repeatability and longer year-to-year comparability analyses of portfolio model assessments were not feasible.

Interrater agreement is low and varies between models and between standards assessed within a single model

Interrater reliability statistics have not been compiled through the online portfolio platform to date, but OREA analyzed interrater agreement, one type of consistency measure that is often grouped with reliability statistics.[36] Interrater agreement measures the degree to which evaluators give the same rating to an observation.

In Tennessee's portfolio process, teachers must first score their students' work samples. Each collection of work samples – student work samples assessed at two different points in time based on a specific standard or set of standards – is randomly assigned to a peer reviewer. OREA reviewed each scored work sample to determine agreement rates between teachers and peer reviewers. (Because OREA could not track peer reviewers' collection assignments, peer reviewers' rating patterns were not analyzed.) Each teacher or peer reviewer typically provided six to 18 individual ratings per collection depending on the portfolio model and standard(s) assessed.[37]

Exhibit 21: Averaged collection interrater agreement between teachers' self-scores and first peer reviewer scores, by portfolio model, 2018-19



Source: OREA analysis of Tennessee Department of Education portfolio score data, 2018-19.

When comparing teachers' self-scores with those of their first peer reviewers, OREA found that performance scores assigned to early grades portfolio collections were in exact agreement an average of 62 percent of the time and within one score level of each other an average of 89 percent of the time. This was the highest observed agreement rate among the four types of portfolio models (early grades, fine arts, physical education, and world languages). Although specific research guidelines for interrater agreement involving self-raters were not identified, general research guidelines on acceptable interrater agreement suggest exact match agreement rates of between 75 percent and 90 percent for high-stakes decisions.[38] If there are more than five to seven rating levels – as is the case with the world languages model, which has nine rating levels – exact agreement rates near the lower end of that range are regarded as acceptable.

Fine arts portfolio scores had the lowest agreement rates, with an averaged collection rate of 36 percent of scores between teachers and peer reviewers in exact agreement. In addition, although they accounted for only 9 percent of all portfolio submissions in 2018-19, fine arts portfolios made up 38 percent of portfolios deemed incomplete submissions.[39]

Further analysis of early grades scoring found math work samples had agreement rates of 71 percent compared to 54 percent for ELA samples.[40] Analysis by individual standards within a portfolio found similarly wide ranges. For example, one pre-k math standard on counting and cardinality had an exact agreement rate of 88 percent, while the exact agreement rate for a pre-k ELA option blended standard was 51 percent.

OREA also analyzed agreement rates between the first and second peer reviewers for portfolio collections that received a second review. Agreement rates between peer reviewers were generally lower than those between teachers and their first peer reviewers. The exceptions were physical education, where the average agreement rates remained nearly the same, and fine arts, where the average agreement rates were about 6 percentage points better. A higher agreement rate between peer reviewers than between teachers and reviewers might be expected since all reviewers receive training and must demonstrate their competency to serve as peer reviewers by passing certification criteria.[N]

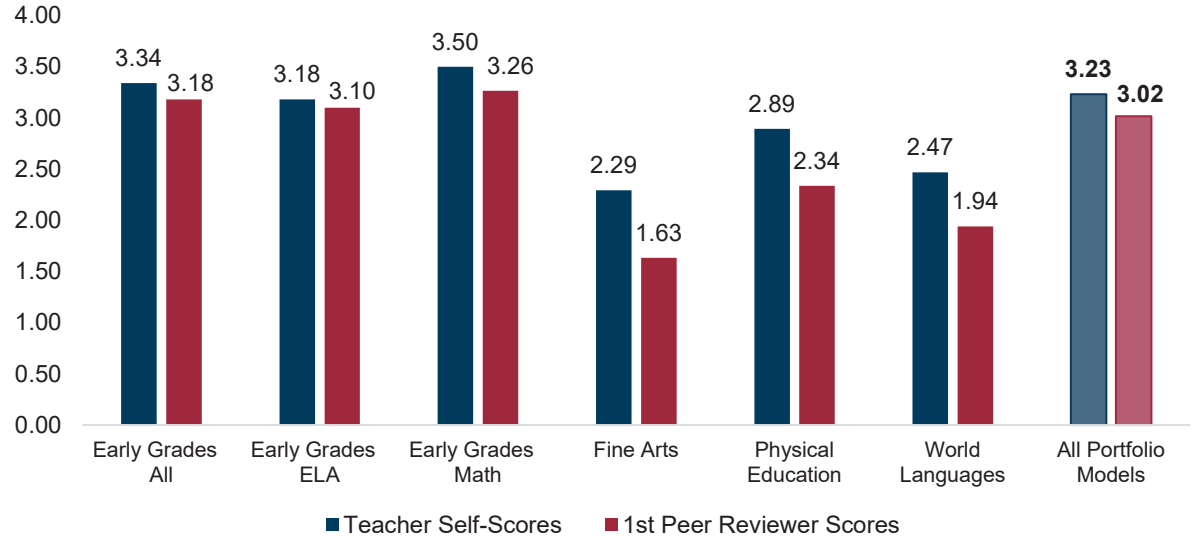
Although the portfolio model scoring process is based on consensus scoring rather than exact matching, low interrater agreement rates and variance in interrater agreement by model and scoring rubric may suggest refinements are needed in reviewer training and/or rubric construction.

[N] Because a second peer reviewer is assigned only when consensus is not reached between the teacher and first peer reviewer (9 percent of collections in 2018-19), they may rate a disproportionate amount of "difficult-to-score" work samples.

Current scoring procedures that can impact reliability

In addition to the interrater agreement analysis of performance scores discussed in the previous section, OREA analyzed the collection growth scores (the calculated student growth before conversion to a growth level rating) for teachers and for first peer reviewers to compare scoring differences. The average collection growth scores calculated from all teachers' self-scores in the 2018-19 school year was 3.23, compared to 3.02 calculated from peer reviewers' scores. An analysis of these means found that the difference in collection growth scores between the groups was statistically significant. The difference in collection growth scores also varied by portfolio model category, with the most pronounced difference for fine arts portfolio collections: teachers' ratings generated a mean collection-level growth of 2.29 compared to 1.63 for peer reviewers.

Exhibit 22: Comparison of average collection growth scores by portfolio model, 2018-19



Source: OREA analysis of Tennessee Department of Education portfolio score data, 2018-19.

The data suggests that teachers, not unexpectedly, assign scores to student work that result in somewhat higher growth scores than do peer reviewers, although it should be noted that averages for both teachers and reviewers are high in absolute terms. Average collection growth scores of 3.0 or above translate to a growth level rating of 5, the top rating possible. The data further suggests that scoring reliability issues may differ among the portfolio models.

The portfolio scoring process begins with teachers self-scoring their own portfolios. While this is a useful technique to encourage teachers to reflect on and improve their instructional practice, self-scoring does not contribute to more objective or reliable measures of student growth.

The use of peer reviewers, most of whom are also participants submitting their own portfolios, has pros and cons that can impact reliability. Department staff recognize that peer reviewers – as fellow teachers working in the same grades and subjects as teachers whose portfolios they review – are familiar with the relevant state standards and the types of student work included in portfolios but may be hesitant to assign a fellow teacher scores that result in low ratings, knowing they may impact a colleague's overall evaluation score and that fellow teachers will be reviewing their own portfolio collections. The final report of the ad hoc Portfolio Review Committee stated that there has been a lack of clarity among peer reviewers about how to score portfolios, suggesting that improvements could be made to the training provided by the department to portfolio reviewers.[O]

If a reviewer finds that they personally know the teacher whose portfolio collection they are scoring or the student whose work is a sample in the portfolio collection, department policy encourages the reviewer to recuse himself or herself but does not require it. In the current platform, reviewers must contact the state to have the collection reassigned.

[O] The Portfolio Review Committee was created by Public Chapter 376 (2019) to review Tennessee's pre-k/kindergarten portfolio model and make recommendations for improvement.

Repeatability of assessment results

Repeatability is another type of reliability, also referred to as test-retest reliability. It refers to the consistency of scores when assessments are repeated under identical conditions within a certain period of time. An IQ test is an example of a test one might expect to yield similar results if adults were to be tested twice within a short timeframe. To accurately assess repeatability of portfolio models as an assessment, researchers would have to precisely repeat the portfolio scoring process in a short amount of time using the same student work samples → assigned to the same peer reviewers → using the same scoring rubrics → performed in the same location and conditions.

Repeatability tests have not been conducted by the department due to practical, logistical, and cost factors. There are time constraints within which portfolios must be scored, converted to an evaluation growth score, and included in teachers' LOE scores. Most peer reviewers are teachers who must complete and submit their own portfolios for review in addition to carrying out their duties as peer reviewers. Requesting or requiring peer reviewers to also participate in repeatability tests is not feasible for reasons of cost and practicality.

Year-to-year comparisons of assessment results

While repeatability focuses on short-term duplication of an assessment, year-to-year comparisons focus on longer-term trends over multiple years. On the TN Ready tests, for example, certain test questions are repeated annually to help ensure the test remains comparable from one year to the next. Repeating specified tasks each year to ensure year-to-year comparability is not done with portfolio assessments, however. Each teacher selects the assigned tasks for students to demonstrate growth, and these tasks can change each year at the teacher's discretion.

Scoring changes over several years also limit the usefulness of year-to-year comparisons. As noted earlier, score categories have expanded, from 1 through 5, to 1 through 7, and now 0 through 7 for all early grade portfolio models. In addition, policies about scoring and rating mismatched student work samples have changed. If student work samples are not matched correctly – that is, a point B sample is not correctly matched to a point A sample from the same student within a collection or work samples are not appropriately matched to the same standard – the work samples are deemed "unscoreable." Similarly, if a student work sample cannot be read or played (in the case of audio or video), it will also be deemed "unscoreable." Previously, if a collection included unscoreable work, the whole collection was automatically assigned a student growth level rating of 1. Beginning in 2018-19, reviewers could assign scores to the remaining readable or usable work samples in a collection and teachers could still earn a collection score based on the portion of scoreable work. If the whole collection is unscoreable, however, it will be assigned a growth level rating of 1.

Also changed is the policy for handling incomplete or unsubmitted portfolios. Prior to 2017-18, if a portfolio was incomplete, teachers would receive a portfolio score level of 1. The department changed the policy beginning with the 2017-18 school year so that incomplete or unsubmitted portfolios no longer receive an automatic score level of 1 but instead receive no portfolio growth score. Without a portfolio score for the student growth component of their evaluation, teachers would not receive an overall evaluation score, and could face potential repercussions on tenure, retention, and compensation decisions.

Portfolio scores from multiple years (see Exhibit 17) are included in this report for descriptive purposes only and should not be used to draw conclusions about trends in teachers' instructional skills or student growth.

Implementation Challenges

Summary: The request for this study specifically asked for review of two issues: (1) the time it takes teachers to prepare and submit their portfolios and (2) specific issues unique to rural districts in implementing portfolio models.

OREA found that some teachers have reported significant time burdens associated with preparing portfolios, though teachers reported that submitting portfolios in 2018-19 was easier and less time-consuming following the state's shift to Portfolium, a new online platform vendor.

A review of various department surveys related to the portfolio process and OREA interviews with staff in 14 districts found no significant distinctions between rural and urban districts in implementing portfolio models.[41]

Time to prepare and submit portfolios

Teachers report that the portfolio model takes time away from classroom practice and requires time spent after hours. State-required teacher assessments, whether based on standardized tests or on student growth portfolios, require time spent on preparation and administration. Education officials must balance the time away from instruction with the feedback from test results that teachers can use to improve the instruction they provide to students.

A department survey of teachers using portfolios in 2017-18 found that 81 percent of all responding teachers (3,404) spent more than eight hours on portfolio preparation (e.g., uploading student work, adding explanatory comments, and completing self-scoring). Teachers using the world languages portfolio (24) reported spending the least amount of time on portfolio preparation, with 46 percent reporting they spent more than eight hours.

The department also surveyed districts on how much release time administrators granted teachers to work on portfolios during the 2017-18 and 2018-19 school years. In both school years, approximately one-third of responding districts reported that teachers were given two days release time. The districts containing the largest concentrations of portfolio teachers (Shelby, Davidson, Knox, and Montgomery) did not respond to time burden-related survey questions for either school year.

Exhibit 23: Release time for portfolio teachers, 2017-18 and 2018-19

	Half a day or less	A full day	Two days	More than two days
Number of districts, 2017-18	7	26	34	27
Percent of districts responding, 2017-18	7.5%	27.7%	36.2%	28.7%
Number of districts, 2018-19	13	30	31	19
Percent of districts responding, 2018-19	13.9%	32.3%	33.3%	20.4%

Note: A total of 94 districts responded to this survey question for 2017-18, and 93 responded for 2018-19.
Source: Tennessee Department of Education, district survey portfolio responses, 2017-18 and 2018-19.

The portfolio process places demands on district administrators' time as well, and these demands appear to be increasing based on changes in districts' portfolio responsibilities outlined in state policy. As the statewide use of portfolio models expanded and the department understood the level of administrative support needed for successful portfolio implementation, the state's requirements of districts grew. Prior to the 2016-17 school year, the state's policy for districts stated:

> *In order to implement one of the student growth portfolio models . . . LEAs must:*
>
> *a. Provide training to evaluators [reviewers] to assess whether the students instructed by the educator being evaluated have demonstrated sufficient growth for the chosen measure, and*
>
> *b. Implement the state's multiple rating categories to measure levels of performance for the chosen measure.*[42]

In April 2017, prior to the first year of required pre-k/kindergarten portfolio implementation, the department requested a change in State Board of Education policy based on feedback from its Student Growth Portfolio Advisory Group. The State Board adopted the department's revisions:

> In order to implement one of the student growth portfolio models . . . LEAs must:
>
> *a. Assign a district portfolio lead to verify portfolio submissions and to facilitate committee reviews as needed.*
>
> *b. Select and provide portfolio evaluators [reviewers] at a ratio of one (1) evaluator for every ten (10) portfolios in each content area.*
>
> *c. All portfolio evaluators must be trained and credentialed by the department to assess student growth according to the portfolio model.*
>
> *d. Implement the state's multiple rating categories to measure levels of performance on the growth model.*[43]

After the first year in which pre-k/kindergarten portfolios were implemented in nearly all districts, the department proposed additional changes to districts' portfolio responsibilities, and the State Board approved these new requirements in October 2018. The state required that each district assign a portfolio technology lead to provide teachers support with the online platform and expanded the duties of the district portfolio leads as follows:

> *Assign a district TEAM portfolio lead to maintain accurate teacher rosters, distribute portfolio-related information and resources, monitor and support timely portfolio submissions for all teachers, and ensure portfolio peer reviewing.*[44]

Research has shown that comprehensive support is key to successful implementation of portfolio models.[45]

Another concern of early grades teachers, in particular, relates more to the logistics of managing a classroom while also documenting the task performance of a selected student for a portfolio collection, such as recording audio of video of a student. For example, one pre-k supervisor indicated that portfolio collections were easier for pre-k teachers to put together than kindergarten teachers because pre-k teachers have a full-time teacher's aide in their classrooms.

Rural districts' unique challenges

Several potential challenges for rural and smaller districts arise from statewide implementation of portfolios.[P] Approximately 49 percent of school districts in Tennessee that use portfolios are rural, and in 2018-19, about 24 percent (1,479) of all portfolio teachers were from rural school districts.[46] Rural districts may have fewer resources, both technological and personnel-related, to administer the portfolio system than larger districts.

Issues with broadband accessibility and network capabilities can pose a challenge to small, rural districts. Uploading portfolio submissions to the online platform causes obstacles when network speeds are slow, a problem more common in rural areas. The portfolio system does not mandate additional technology upgrades; TDOE encourages the use of district-owned devices to complete portfolios and notes that no significant technological investment should be made by districts.

Despite these possible barriers, no rural districts that responded to an OREA survey reported that technological problems affected their ability to complete the portfolio requirement for the 2018-19 school year. Rural teachers received, on average, slightly higher scores on their portfolios than urban teachers for the 2018-19 school year. Several districts indicated there were problems in the past, however, caused by the online platform's insufficient load capacity and other implementation challenges.

Exhibit 24: 2018-19 Portfolio teacher and portfolio reviewers by district classifications

Designation	Average teacher effectiveness indicator or "portfolio score"	Number of portfolio teachers	Percent of portfolio teachers	Number of portfolio reviewers	Percent of portfolio reviewers	Number of participating school districts	Percent of participating school districts
Rural	4.8	1,479	24.4%	224	30.3%	67	48.9%
Urban	4.6	4,580	75.6%	504	68.2%	70	51.1%
Total	**4.6**	**6,059**	**100%**	**728**	**98.5%**	**137**	**100%**

Note: There were 11 TDOE staff certified as expert reviewers, representing approximately 2 percent of certified reviewers. No TDOE staff served as official reviewers and are not included in this exhibit.
Source: Tennessee Department of Education, portfolio scoring data, 2018-19.

Department staff indicated that recruiting reviewers from rural districts is challenging. The department suggested that these districts may have a limited number of teachers to provide one peer reviewer for every 10 portfolios in each content area, a requirement under SBE's educator evaluation policy.

By nature of their size, urban districts are more likely to draw from a larger pool of possible peer reviewers than smaller, rural districts. Smaller districts may also not have staff available to help with the administration of the portfolio program. Larger districts, by contrast, may even employ a portfolio coordinator position in some schools and use economies of scale to share information and resources among districts.

Approximately 30 percent of peer reviewers for 2018-2019 were from rural school districts, while rural teachers made up 24 percent of the total portfolio teacher population. (See Exhibit 23.) While there were fewer rural peer reviewers than urban, the ratio of rural to urban peer reviewers (30 percent versus 68 percent) was similar to the ratio of rural to urban portfolio teachers (24 percent versus 76 percent).

[P] While most rural districts tend to be smaller in both student enrollment and teachers employed than their urban counterparts, some small districts – typically municipal or special school districts – are in urban areas and are small because of their geographic boundaries. Although this section describes rural districts' challenges with portfolios, some of the same challenges also apply to small municipal and special districts.

OREA did not find evidence to suggest that rural districts struggled more than urban districts in meeting the peer reviewer requirement. All school districts classified as "rural-remote," the most rural designation, exceeded or met the required number of peer reviewers. In contrast, several urban school districts including Metro Nashville, Hamilton County, and Shelby County contributed fewer peer reviewers than the number mandated by the state's 1:10 requirement. This requirement may soon be removed, however.

The department presented potential revisions of several aspects of portfolio models to the State Board of Education in November 2019. The proposed revisions remove the requirement that districts provide one peer reviewer for every 10 portfolios they submit in each content area. The State Board is expected to consider final approval of the portfolio revisions in February 2020, but if the policy does change, it is not yet known how portfolio peer reviewers will be selected.

Student Growth Portfolios in Other States

Summary: Student growth portfolios have been used in other states for several kinds of student and teacher assessment and by national organizations to assess teachers' instructional practice for licensure and advanced certification. Tennessee's portfolio model is unique in both purpose and scale. Tennessee appears to be the only state using portfolios as a quantitative measure of student growth in annual teacher evaluations. Tennessee is also the only state requiring portfolio use by all local districts.

While no other state mandates that local districts adopt portfolios, some states allow portfolios as a district option for teacher evaluations. In these cases, local school districts typically select a measure from approved options outside of the mandated state assessments. Districts appear to rarely select portfolios as an option or do not report their selected options to their states and are unknown. Portfolios may also be used by districts as an alternative to standardized assessments for students with disabilities or English language learners, for example.

Districts may more commonly choose to use student learning objectives, or SLOs, a portfolio-like tool that can be used as a student growth measure in annual teacher evaluations for teachers that do not teach classes or grades covered by state standardized tests.

Portfolios in teacher preparation and licensure

Portfolios originally gained popularity in the 1990s as a way for student teachers in teacher preparation programs to document and reflect on applying good instructional practices. The National Board of Professional Teaching Standards (NBPTS) uses portfolios as a method of assessing teacher practice for its advanced certification program, and ETS uses them for its licensing assessment programs. In these cases, student work is collected to demonstrate the teacher's instructional practice and impact on student learning, but not to quantitatively measure student growth.

Some states, such as Minnesota and Vermont, use portfolios for teacher licensure. Teacher portfolios have been used in Iowa for a mentoring and induction program in which beginning teachers submit evidence of their teaching practice, and mentors review their work online and provide feedback as teachers work toward an evidence collection for licensure. Iowa state law permits school districts to create local licensure programs. Typically, the quality of a beginning teacher's portfolio is factored into a pass/fail decision about licensure as opposed to the use of the portfolio as a quantitative measure of student growth.

Maryland began exploring the use of portfolios in teacher evaluations as a quantitative student growth measure in 2011, around the same time teachers in the former Memphis City Schools were developing the fine arts portfolio model in Tennessee. A pilot project with one Maryland district in 2011-12, using a model similar to Tennessee's current model, produced some similar outcomes, including scores that skewed relatively

high and some significant time burdens for teachers, particularly around using the required technology.[Q] Most Maryland school districts, like many others in states that include a student growth component in teacher evaluations, use student learning objectives, described in more detail below.

Student learning objectives: A portfolio-like tool

Several school districts across the nation, such as Austin Independent School District in Texas, Box Elder School District in Utah, and Denver Public Schools in Colorado, have used student learning objectives (SLOs) to measure an educator's contribution to student growth. Like Tennessee's portfolio models, SLOs can be used as a student growth measure in annual teacher evaluations for teachers who do not teach classes or grades covered by state standardized tests.

Student learning objectives are laid out in a teacher-created plan based on state standards, or district or school-level goals. Within the SLO, a teacher outlines specific growth target areas for students, assessment tools to measure students' achievement of target areas, and expected outcomes for the year. Some districts require approval of the SLO before the teacher begins implementing the plan. These districts typically require that teachers meet with evaluators throughout the year to check the progress of the SLO. At the end of the school year, the SLO with relevant student data is submitted to a principal, administrative leader, or other evaluator to determine how well the teacher helped students meet the growth goals set by the teacher. Grading policies vary by state and district. Some districts grade teachers with levels based on growth data – e.g., failing, needs improvement, proficient, or distinguished – or use a checklist aligned with state standards. Like a portfolio, an SLO tracks student progress during the year and measures a student's skills and knowledge as a component of teacher evaluation.

One difference between SLOs and portfolios is which students are chosen to assess academic growth. For portfolios, teachers select a sample of students to represent each of the different proficiency levels in a classroom. (The three different proficiency levels used for portfolios in Tennessee are emerging, proficient, and advanced.) With SLOs, teachers may choose a certain student group (for example, English learners) to target or set learning objectives for each student in their class. There is typically less standardization across classrooms with the SLO process than with Tennessee's statewide portfolio process.

SLOs may be subject to some of the same validity and reliability weaknesses as portfolios if used as a component of teacher evaluation. Like portfolios, SLOs are more subjective than standardized state assessments, and SLOs allow for significant direction by the same teachers who are being evaluated. The individualistic nature of SLOs makes it difficult to compare teacher evaluation results across districts or schools.

Conclusion: Portfolios as a Component of Teacher Evaluation

Current use

The number of teachers who receive portfolio scores in Tennessee is low: for the 2018-19 school year, 9 percent of teachers received portfolio scores. Most of these teachers taught pre-k or kindergarten and are required to use portfolios because their districts accepted state Voluntary Pre-k (VPK) program funding. Of the other 91 percent of teachers, approximately 29 percent received an individual TVAAS score while the other 62 percent received school level TVAAS scores. The number of districts adopting portfolio models as their student growth component for teacher evaluations has generally been declining, other than when needed to meet state requirements.

Several factors in the design of portfolio models weaken their validity and reliability as a quantitative measure of student growth. Such factors include the lack of standardized assessment tasks, the expansion of score categories in the rubrics used to evaluate student work, the subjective nature of purposeful sampling, and certain scoring

[Q] The Maryland district that pilot-tested student growth portfolios no longer uses them.

procedures. OREA analysis of 2018-19 scoring results found low interrater score agreement between teachers' self-scores and first peer reviewers' scores, as well as between first and second peer reviewers' scores.

Other design factors, more qualitative in nature, require teachers to undertake steps that are likely to help improve their instructional practice, such as deepening their knowledge of key state standards, strengthening alignment of classroom assignments with the standards, identifying students by skill levels (emerging, proficient, advanced), planning lessons and differentiating instruction to help all students achieve at least one year of growth, collaborating with colleagues, and preparing narratives to explain the context of student work samples. Increased collaboration among teachers, cited by the department as one of the benefits of the portfolio process, may result in more teachers gaining skills in compiling top-scoring portfolios. The development of such skills is desirable when linked to improved instructional practice, such as deeper understanding of state standards or better strategies to help students at different performance levels gain mastery of those standards. When skill at compiling a top-scoring portfolio is not linked to instructional practice, teachers' high portfolio scores are less valid and reliable measures of student academic growth.

Although the ultimate goal of improving instruction is, of course, to increase student learning, a tool that helps improve instruction is not necessarily an appropriate tool to objectively measure student growth. The process teachers go through to compile a strong student growth portfolio may be a more effective way to provide teacher professional development and improve teacher instructional practice than its current use in teacher evaluations as a quantitative measure of student growth.

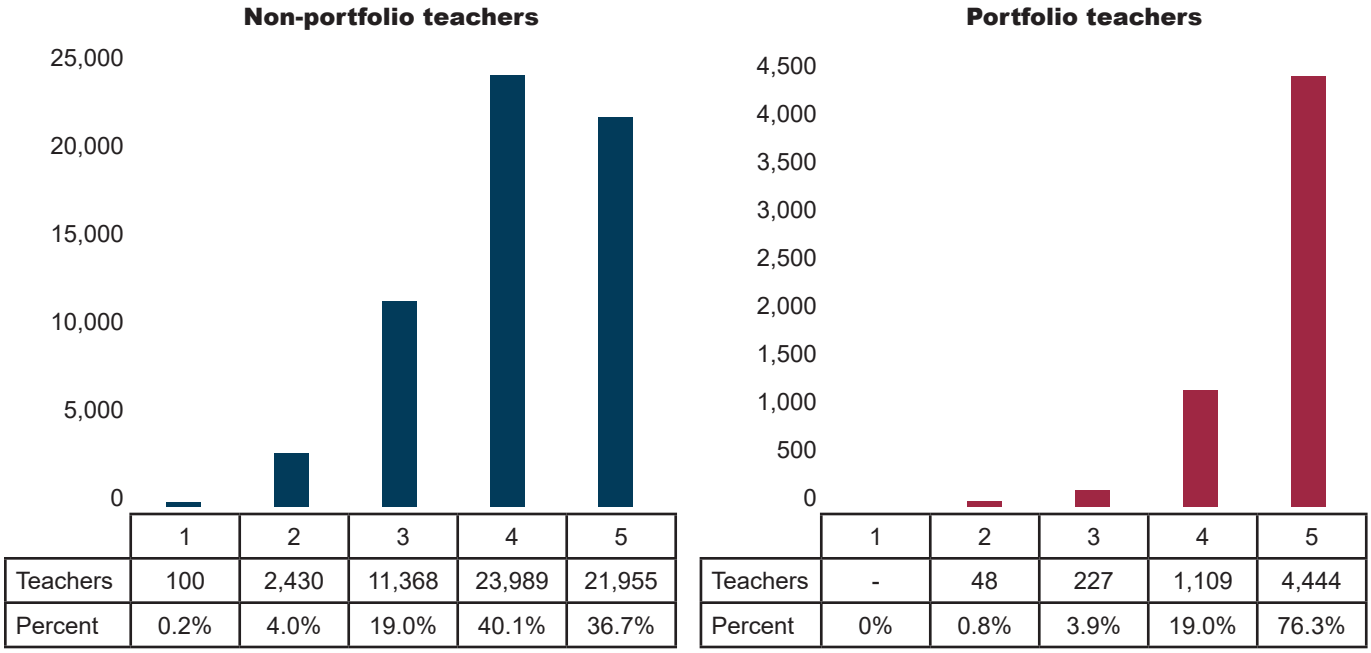
Changing goals

While portfolio models were initially developed to give more teachers an individual growth score for evaluation purposes as opposed to using school composite growth scores, the legislature's mandate that all pre-k and kindergarten teachers in VPK districts (96 percent of all districts) be required to use the portfolio model for their evaluation growth score was driven by a goal of improving pre-k programs across the state. The state's setting of 3rd grade reading goals in its Read to be Ready initiative seems to have also linked the use of portfolio models with a broader goal – that of improving early grades instruction in order to increase the number of students reading on grade level by 3rd grade. Yet, portfolio use is not required for 1st and 2nd grade teachers, and no portfolio model exists for 3rd grade teachers. (Although 3rd grade students take TN Ready tests, no TVAAS growth score can be calculated for those students unless they are in districts that use the optional 2nd grade standardized assessment. About 106 districts (75 percent) have opted to use the 2nd grade assessment.)

It is too soon to determine whether requiring portfolios for pre-k and kindergarten teachers has impacted the quality of pre-k and other early grades so that (1) the benefits of pre-k persist into later grades and (2) more students are reading on grade level by 3rd grade. The pre-k/kindergarten portfolio requirement was implemented statewide in 2017-18. Students in those classes will not take optional 2nd grade assessments or their first year of TN Ready tests as 3rd graders until spring of 2021. Yet the lack of differentiation in portfolio scores from 2018-19 suggests they are of limited use to districts in identifying and rewarding the most effective pre-k and kindergarten teachers, providing additional support and training to less effective teachers, or removing ineffective teachers.

Teachers with individual portfolio growth scores are more likely to receive a top overall evaluation score (LOE score) than "non-portfolio" teachers who have either an individual TVAAS or schoolwide composite TVAAS growth score. (See Exhibit 25.) Districts have the option to allow teachers who score a level 4 or 5 on individual growth to use their individual growth score for their entire evaluation score. Overall evaluation scores are used by districts for decisions about tenure and may be used for decisions about retention and compensation. If districts use individual growth scores and/or overall evaluation scores to compare teachers within their district, teachers with portfolio scores as their growth component may be judged more positively due to the tendency for portfolio scores to skew high.

Exhibit 25: Teacher overall evaluation (level of effectiveness or LOE) scores, 2018-19



	1	2	3	4	5
Teachers	100	2,430	11,368	23,989	21,955
Percent	0.2%	4.0%	19.0%	40.1%	36.7%

	1	2	3	4	5
Teachers	-	48	227	1,109	4,444
Percent	0%	0.8%	3.9%	19.0%	76.3%

Source: Tennessee Department of Education, 2018-19 score data.

Other states' use

OREA's review could find no other states that require student growth portfolios as a quantitative measure of student growth for teacher evaluation purposes. Although portfolios might be a district option, OREA could not identify districts in other states that have adopted portfolios for teacher evaluation. Districts and states that used student growth portfolios in earlier years have discontinued such use.

States that include student growth components in teacher evaluations have typically let districts decide what student growth measure should be used for teachers in subjects and grades without standardized tests. One common choice has been student learning objectives (SLOs) set by teachers for each student or for a subset of students. SLOs, however, appear to have some of the same weaknesses in validity and reliability as portfolio models, due to their lack of standardization and their subjectivity. Other states are opting to discontinue the use of student growth components in teacher evaluations altogether, as noted in a recent report from the National Council on Teacher Quality (NCTQ).[47]

Portfolios' future

In 2019, the General Assembly revised state law requirements for districts using the pre-k/kindergarten portfolio model for teacher evaluations if they accepted VPK state funds.[48] The revised law allows districts to use a "comparable alternative measure of student growth approved by the State Board of Education" as an alternative to the portfolio model. In November 2019, the department presented to the State Board of Education the proposed guidelines for approving alternative growth measures, including that the measures be nationally normed, evidence based, and able to evaluate all students in pre-k and kindergarten on math and English language arts.

Some of the early alternative growth measures proposed by districts are benchmark tests that could be given to pre-k and kindergarten students at the beginning of the school year and again at the end of the year in order to calculate growth. Some of these tests have already been approved by the State Board for use as measures for the 15 percent of a teacher's evaluation based on student achievement scores. A number of districts already use benchmark tests to measure students' progress throughout the year. Such benchmark tests, if they meet the department's proposed guidelines, would likely be more standardized, less subjective, and could be used to measure growth of all students in the classroom, rather than just a sample of selected students, and thus

have the potential to be a more valid and reliable measure of student growth than portfolio models. Although the proposal for such tests allows districts to choose among options and would thus not be consistent from one district to another, teacher evaluation results are used at the district level, not statewide, for human capital decisions. District options for a student growth measure would be similar to district options to adopt portfolios in place of school level TVAAS scores (for all but the pre-k/kindergarten model) and to the district, administrator, and teacher options for the student achievement measure.

Policy Considerations

The Tennessee Department of Education should consider the following methods to increase the validity and reliability of portfolio models, particularly if they continue to be used as a quantitative measure for student growth in teacher evaluations. The potential revisions suggested below would likely require investments of staff time, financial resources, or both. Such resource costs should be weighed against current investments of teacher and district staff time and local and state dollars used to produce student growth scores that do not meaningfully differentiate among teachers.

1. Develop standardized student assignments, aligned to the academic standards used in each portfolio model, and provide them to teachers through portfolio training and guidelines. Such standardization would better ensure that teachers' growth scores reflect gains in student learning, rather than ability to design an effective portfolio assignment, and would help make their portfolio growth scores more comparable.

2. Reassess rubrics on a case-by-case basis and determine whether the number of scoring categories (such as 0 through 7 for early grades portfolios) should be revised to make rating more reliable and more effective at making distinctions between students' abilities to master the relevant state standards.

3. Consider removing teacher self-scoring and the consensus review approach from portfolio scoring for student growth measure purposes. Self-scoring may be valuable as a tool for self-reflection and improved practice but serves little purpose in producing an objective measure of student growth.

4. Develop a small, committed, and well-trained pool of portfolio reviewers as recommended in the 2019 Portfolio Review Committee final report. Using educators with expertise in state standards who are not also submitting their own portfolios for scoring would help minimize potential for conscious or unconscious scoring bias.

5. Require peer reviewers to recuse themselves from reviewing portfolio collections for which they believe they know the teacher submitting the collection or any of the students whose work is included. Consider whether safeguards can be programmed in the portfolio online platform to prevent collections being assigned to reviewers in the same school district.

6. Revise platform specifications so that they mask the performance group that a student work sample represents and whether the work sample is from the point A or point B time frame.

7. Set more rigorous criteria for certifying portfolio reviewers. For example, require a specified level of exact matches of interrater score agreement, factoring in the number of assigned scores per collection and the number of scoring categories available to reviewers, rather than adjacent growth level ratings calculated from growth scores.

8. Activate the provisions in the vendor contract for the portfolio online platform to collect and provide interrater reliability statistics.

9. Consider revising the scoring process to reduce the number of score conversions, and thereby reduce the artificial cut-offs for different score categories. For example, each collection's raw score, which is the average of individual student growth scores, could be retained until it is averaged with the other collections scores. Then, only the final overall portfolio score could be converted to a growth rating.

10. Set requirements for the minimum number of students that must be included in a teacher's complete portfolio to ensure adequate representation across a class. Consider having portfolio leads in each district monitor the number of students selected for each complete portfolio or track students chosen as part of purposeful sampling to see if proficiency levels exhibited in portfolios correlate with related external benchmarks over time.

11. Develop district and department capacity to monitor the validity and quality of data throughout the portfolio process, as recommended by the Portfolio Review Committee. This may include conducting periodic checks of rater agreement against the scoring rubrics or randomly reviewing a subset of raters' scores on work samples.

Some of the issues related to portfolios' validity and reliability attach to their use a quantitative measure of student growth, which is then incorporated into a high-stakes teacher evaluation score. If portfolio models were used primarily as a tool to improve teachers' instructional practice, rather than as a measure of student growth, some of their inherent weaknesses – such as lack of standardization – become less of a concern. The General Assembly, Department of Education, and local districts may wish to consider the best way to use portfolio models in the future to capitalize on their strengths to meet objectives such as developing teachers' instructional skills, ensuring pre-k quality, and increasing 3rd grade reading attainment. Some options include:

1. Reviewing the statutory requirements that all districts use at least one portfolio model for teacher evaluation and that all districts that accept Voluntary Pre-K funding implement the pre-k/kindergarten portfolio model for teacher evaluation. Successful implementation of portfolio models requires sufficient district administrative support. Providing districts the option to adopt portfolio models for teacher evaluation could help ensure that districts are ready and willing to provide the needed support.

   If districts were not required to adopt a portfolio model, those that chose not to would revert to using school level TVAAS scores for teacher evaluations. In 2018-19, 62 percent of Tennessee teachers received school level TVAAS scores as the student growth component of their overall evaluation score.

   It is unknown as of December 2019 how the newest statutory provision allowing districts to choose an alternative student growth measure to the pre-k/ kindergarten portfolio model will impact portfolio use until policies regarding such alternatives are finalized by the State Board of Education and districts pilot conditionally-approved alternatives.

2. Retaining some requirement for portfolio use in teacher evaluations, but shifting it to part of the qualitative component of evaluations. For example, teachers in early grades (pre-k through grade 3) could be required to prepare a portfolio in lieu of one classroom observation. The portfolio models' focus on state standards, on helping students at varying performance levels achieve growth, and on collaboration with other teachers would seem to be particularly helpful in developing strong instructional practice skills among new and/or struggling teachers or improving skills among experienced teachers. Organizations, such as the National Board for Professional Teaching Standards (NBPTS), for example, use portfolios as part of an advanced certification process.

3. Reducing the weight of student growth portfolio scores in overall teacher evaluations to 15 percent, matching the weight of schoolwide TVAAS scores, rather than matching the weight of individual TVAAS scores. Comparison of 2018-19 score distributions for teachers using each type of student growth score indicate that portfolios are not providing enough differentiation among teachers to warrant treating them as equivalent to individual TVAAS scores.

4. Blending some type of required portfolio use with district options. For example, new teachers in early grades could be required to use the portfolio model under the guidance of a more experienced or mentor teacher for the student growth component of their evaluation. Another example might be for teachers scoring poorly on evaluations to be required to prepare a portfolio in the following year to improve their instructional practice.

Appendix A - District portfolio adoptions, 2018-19, 2019-20

District	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
	2018-2019				2019-2020			
Achievement School District (ASD)	Pre-K/K	1st grade			Pre-K/K			
Anderson County	Pre-K/K				Pre-K/K			
Clinton City	Pre-K/K	1st grade			Pre-K/K			
Oak Ridge City	Pre-K/K				Pre-K/K			
Bedford County	Pre-K/K				Pre-K/K			
Benton County	Pre-K/K				Pre-K/K	1st grade		
Bledsoe County	Pre-K/K				Pre-K/K			
Blount County	Pre-K/K				Pre-K/K			
Alcoa City	Pre-K/K				Pre-K/K			
Maryville City	Pre-K/K				Pre-K/K			
Bradley County	Pre-K/K				Pre-K/K			
Cleveland City	Pre-K/K				Pre-K/K			
Campbell County	Pre-K/K				Pre-K/K			
Cannon County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Carroll County	n/a				n/a			
Hollow Rock Bruceton SSD	Pre-K/K				Pre-K/K			
Huntingdon SSD	Pre-K/K				Pre-K/K			
McKenzie SSD	Pre-K/K				Pre-K/K			
South Carroll SSD	Pre-K/K				Pre-K/K			
West Carroll SSD	Pre-K/K				Pre-K/K			
Carter County	Pre-K/K				Pre-K/K			
Elizabethton City	Pre-K/K	1st grade			Pre-K/K	1st grade		
Cheatham County	Pre-K/K				Pre-K/K			
Chester County	Pre-K/K				Pre-K/K			
Claiborne County	Pre-K/K				Pre-K/K			
Clay County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Cocke County	Pre-K/K				Pre-K/K			
Newport City	Pre-K/K	1st grade			Pre-K/K	1st grade	2nd grade	
Coffee County	Pre-K/K				Pre-K/K			
Manchester City	Pre-K/K	1st grade			Pre-K/K			
Tullahoma City	Pre-K/K				Pre-K/K			
Crockett County	Pre-K/K				Pre-K/K			
Alamo City	Pre-K/K				Pre-K/K			
Bells City	Pre-K/K				Pre-K/K			
Cumberland County	Pre-K/K	1st grade	2nd grade	Fine Arts	Pre-K/K			
Davidson County (MNPS)	Pre-K/K				Pre-K/K			
Decatur County	Pre-K/K				Pre-K/K			
DeKalb County	Pre-K/K	1st grade	2nd grade		Pre-K/K	1st grade	2nd grade	
Dickson County	Pre-K/K				Pre-K/K			
Dyer County	Pre-K/K				Pre-K/K			

District	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
	2018-2019				2019-2020			
Dyersburg City	Pre-K/K				Pre-K/K			
Fayette County	Pre-K/K				Pre-K/K			
Fentress County	Pre-K/K				Pre-K/K			
Franklin County	Pre-K/K				Pre-K/K			
Gibson County - no district								
Bradford SSD	Pre-K/K	1st grade			Pre-K/K	1st grade		
Gibson County SSD	Pre-K/K				Pre-K/K			
Humboldt City	Pre-K/K	1st grade			Pre-K/K	1st grade		
Milan SSD	Pre-K/K				Pre-K/K			
Trenton SSD	Pre-K/K				Pre-K/K			
Giles County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Grainger County	Pre-K/K				Pre-K/K			
Greene County	Pre-K/K				Pre-K/K			
Greeneville City	Pre-K/K				Pre-K/K			
Grundy County	Pre-K/K				Pre-K/K	1st grade	2nd grade	
Hamblen County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Hamilton County	Pre-K/K				Pre-K/K			
Hancock County	Pre-K/K				Pre-K/K			
Hardeman County	Pre-K/K				Pre-K/K			
Hardin County	Pre-K/K				Pre-K/K			
Hawkins County	Pre-K/K				Pre-K/K			
Rogersville City	Pre-K/K				Pre-K/K			
Haywood County	Pre-K/K				Pre-K/K			
Henderson County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Lexington City	Pre-K/K				Pre-K/K			
Henry County	Pre-K/K	1st grade			Pre-K/K			
Paris SSD	Pre-K/K	1st grade			Pre-K/K	1st grade		
Hickman County	Pre-K/K				Pre-K/K			
Houston County	Pre-K/K				Pre-K/K			
Humphreys County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Jackson County	Pre-K/K	1st grade	Physical Education		Pre-K/K	1st grade	Physical Education	
Jefferson County	Pre-K/K				Pre-K/K			
Johnson County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Knox County	Pre-K/K				Pre-K/K			
Lake County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Lauderdale County	Pre-K/K				Pre-K/K			
Lawrence County	Pre-K/K				Pre-K/K			
Lewis County	Pre-K/K				Pre-K/K			
Lincoln County	Pre-K/K				Pre-K/K			
Fayetteville City	Pre-K/K	1st grade			Pre-K/K	1st grade		
Loudon County	Pre-K/K				Pre-K/K			
Lenoir City	Pre-K/K				Pre-K/K			
Macon County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Madison County	Pre-K/K	1st grade	Fine Arts		Pre-K/K	1st grade	Fine Arts	
Marion County	Pre-K/K	Fine Arts			Pre-K/K	Fine Arts		

District	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
	2018-2019				2019-2020			
Richard City	Pre-K/K	1st grade			Pre-K/K	1st grade	2nd grade	
Marshall County	Pre-K/K				Pre-K/K			
Maury County	Pre-K/K				Pre-K/K			
McMinn County	Pre-K/K				Pre-K/K			
Etowah City	Pre-K/K				Pre-K/K			
Athens City	Pre-K/K				Pre-K/K			
McNairy County	Pre-K/K				Pre-K/K	1st grade		
Meigs County	Pre-K/K				Pre-K/K			
Monroe County	Pre-K/K				Pre-K/K	1st grade		
Sweetwater City	Pre-K/K				Pre-K/K			
Montgomery County	Pre-K/K	Fine Arts			Pre-K/K	Fine Arts		
Moore County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Morgan County	Pre-K/K				Pre-K/K			
Obion County	Pre-K/K				Pre-K/K			
Union City	Pre-K/K				Pre-K/K			
Overton County	Pre-K/K				Pre-K/K			
Perry County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Pickett County	Pre-K/K				Pre-K/K			
Polk County	Pre-K/K				Pre-K/K			
Putnam County	Pre-K/K				Pre-K/K			
Rhea County	Pre-K/K				Pre-K/K			
Dayton City	Pre-K/K				Pre-K/K			
Roane County	Pre-K/K				Pre-K/K			
Robertson County	Pre-K/K				Pre-K/K			
Rutherford County	Pre-K/K				Pre-K/K			
Murfreesboro City	Pre-K/K				Pre-K/K			
Scott County	Pre-K/K				Pre-K/K			
Oneida City	Pre-K/K				Pre-K/K			
Sequatchie County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Sevier County					Pre-k only			
Shelby County	Pre-K/K	World Languages	Fine Arts	Physical Education	Pre-K/K	World Languages	Fine Arts	Physical Education
Arlington City					Fine Arts			
Bartlett City	Pre-K/K				Pre-K/K			
Collierville City	Pre-K/K				Pre-K/K			
Germantown City					Physical Education			
Lakeland City					World Languages			
Millington City	Pre-K/K				Pre-K/K			
Smith County	Pre-K/K				Pre-K/K			
Stewart County	Pre-K/K				Pre-K/K			
Sullivan County	Pre-K/K				Pre-K/K			
Bristol	Pre-K/K				Pre-K/K			
Kingsport	Pre-K/K				Pre-K/K			
Sumner County	Pre-K/K				Pre-K/K			
Tipton County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Trousdale County	Pre-K/K	1st grade	2nd grade		Pre-K/K	1st grade	2nd grade	

District	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
	2018-2019				2019-2020			
Unicoi County	Pre-K/K				Pre-K/K			
Union County	Pre-K/K				Pre-K/K			
Van Buren County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Warren County	Pre-K/K	Fine Arts			Pre-K/K	Fine Arts	1st grade	
Washington County	Pre-K/K				Pre-K/K			
Johnson City	Pre-K/K				Pre-K/K			
Wayne County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Weakley County	Pre-K/K	1st grade			Pre-K/K	1st grade		
White County	Pre-K/K	1st grade			Pre-K/K	1st grade		
Williamson County	Pre-K/K				Pre-K/K			
Franklin SSD	Pre-K/K				Pre-K/K			
Wilson County	Pre-K/K	1st grade	Physical Education		Pre-K/K	1st grade	Physical Education	
Lebanon SSD	Pre-K/K	1st grade			Pre-K/K	1st grade		

Sources: Districts' models for 2018-19 come from actual score data provided by the Tennessee Department of Education. Districts' models for 2019-20 come from a department flexibility survey in which districts indicated the model they planned to adopt for the 2019-20 school year.

Endnotes

[1] Public Chapter 2, First Extraordinary Session, 2010, https://publications.tnsosfiles.com/acts/106/pub/pc0002EOS.pdf (accessed Sept. 17, 2019).

[2] Ibid., p. 5.

[3] Public Chapter 105, 2013, p.1, https://publications.tnsosfiles.com/acts/108/pub/pc0105.pdf (accessed Jan. 8, 2020).

[4] Public Chapter 703, 2016, https://publications.tnsosfiles.com/acts/109/pub/pc0703.pdf (accessed Jan. 8, 2020).

[5] Tennessee Department of Education, Division of Data and Research, *The Rise of Student Growth Portfolio Models in Tennessee*, Jan. 2017, p.4, https://files.eric.ed.gov/fulltext/ED572949.pdf (accessed Oct. 11, 2019); Tennessee Department of Education, 2019 portfolio score data provided under Oct. 3, 2019 Memorandum of Agreement.

[6] Emerald Academy, a charter school in Knox County, had 10 teachers that used portfolio models not adopted by Knox County Schools. Similarly, a charter school in Shelby County, Memphis Scholars, had two teachers using the 1st grade model. Shelby County did not adopt the 1st grade model as a district.

[7] Calculations based on data from Division of Data and Research, Tennessee Department of Education, *The Rise of Student Growth Portfolio Models in Tennessee*, Jan. 2017, p. 5, https://files.eric.ed.gov/fulltext/ED572949.pdf (accessed Oct. 11, 2019).

[8] Tennessee Department of Education, 2018-19 reviewer data provided under Oct. 3, 2019 Memorandum of Agreement.

[9] State Board of Education, Teacher and Principal Evaluation Policy 5.201, effective May 31, 2019.

[10] Tennessee Department of Education, 2019 portfolio score data provided under Oct. 3, 2019 Memorandum of Agreement.

[11] For payment data: Tennessee Department of Education, email, Oct. 2, 2019; OREA review of Edison vendor payment data, Oct. 24, 2019. For portfolio submission data: Exhibit 17 of this report.

[12] Tennessee Department of Education, email Oct. 10, 2019; OREA analysis of vendor payment data from Edison and Tennessee Department of Education emails, Oct. 2 and Oct 10, 2019.

[13] Tennessee Department of Education, emails, Oct. 2 and Oct. 10, 2019.

[14] OREA review of Edison vendor payment data, Sept. 25, 2019; Tennessee Department of Education, interview, Nov. 1, 2019.

[15] Calculated: $216,496 + $607,282 + 26,100 in 2018-19 payments for portfolio software platform, reviewer expenses, and portfolio consultants = $849,878 total, divided by 6,059 portfolio submissions in 2018-19 = $140.27 per submission.

[16] SAS Invoice approved for payment on Sept. 13, 2019 (from Edison filenet Dec. 13, 2019); Tennessee Department of Education 2018-19 score distribution data provided under Oct. 3, 2019 Memorandum of Agreement.

[17] Tennessee Department of Education, "Voluntary Pre-Kindergarten (VPK) Application for 2019-20 School Year," p. 9, https://www.tn.gov/content/dam/tn/education/early-learning/pre-k/vpk_application_packet_2019.pdf (accessed Sept. 26, 2019)

[18] Research suggests that placing high-stakes consequences on test results in social sciences can skew the results and reduce the validity of the test as an objective measure. See, for example, Donald T. Campbell, *Assessing the Impact of Planned Social Change*, Dec. 1976, pp. 51-52, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.170.6988&rep=rep1&type=pdf (accessed Oct. 24, 2019) and Gail Goldberg, "Portfolio Assessment as a Tool for Teacher Evaluation: Outline and Key Points of Presentation to the Maryland Council for Educator Effectiveness," May 2011, p.2, http://archives.marylandpublicschools.org/NR/rdonlyres/69B8F5FC-6A4A-468F-8886-E089A2605AD8/28669/Portfolio_Assessment_Presentation_Outline_051611.pdf (accessed Oct. 8, 2019).

[19] Tennessee Department of Education, *Team Portfolio Resource Guide – Early Grades 2019-20*, p.5, https://gallery.mailchimp.com/b28b453ee164f9a2e2b5057e1/files/91515d2f-72b3-48db-8660-f7590db6c6b6/2019_20_Early_Grades_TEAM_Portfolio_Resource_Guide.02.pdf (accessed Aug. 27, 2019).

[20] Shelby County Schools, *Scoring Guide to Portfolio of Performance-Based Practice*, World Languages, 2015, p.6, https://team-tn.org/wp-content/uploads/2016/07/World-Language-Portfolio-Scoring-Guide-2016.pdf (accessed Oct. 1, 2019).

[21] Tennessee Department of Education, PowerPoint presented at Portfolio Review Committee, July 23, 2019, slide 8.

[22] Judith A. Arter, Vicki Spandel, and Ruth Culham, *Portfolios for Assessment and Instruction*, ERIC Digest, 1995, p.2, https://eric.ed.gov/?id=ED388890 (accessed Oct. 7,2019); National Comprehensive Center for Teacher Quality, *Measuring Teachers' Contributions to Student Learning Growth for Nontested Grades and Subjects*, March 2011, p. 10, https://gtlcenter.org/sites/default/files/docs/MeasuringTeachersContributions.pdf (accessed July 15, 2019); Gail Goldberg, "Portfolio Assessment as a Tool for Teacher Evaluation: Outline and Key Points of Presentation to the Maryland Council for Educator Effectiveness," May 2011, p.2.

[23] Tennessee Department of Education, Division of Data and Research, *The Rise of Student Growth Portfolio Models in Tennessee*, Jan. 2017, p. 6, https://files.eric.ed.gov/fulltext/

ED572949.pdf (accessed Oct. 11, 2019).

[24] Ibid., p. 7, https://files.eric.ed.gov/fulltext/ED572949.pdf (accessed Oct. 11, 2019). The department's analysis was published in January 2017, at a time when teachers used the GLADiS platform and districts' use of portfolio models was optional.

[25] Department staff indicate that observation scores for all teachers, not just portfolio teachers, have increased over time.

[26] Susannah Faxon-Mills, Laura S. Hamilton, Mollie Rudnick, and Brian M. Stecher, "Conditions that Influence Educators' Responses to Assessment," in *New Assessments, Better Instruction?*, RAND Corporation, 2013, p.7, https://www.jstor.org/stable/10.7249/j.ctt5hhtkh.12?seq=7#metadata_info_tab_contents (accessed Nov. 18, 2019); National Research Council, *Incentives and Test-Based Accountability in Education*, 2011, p. 54, http://download.nap.edu/cart/download.cgi?&record_id=12521 (accessed Dec. 5, 2019).

[27] Stephen Mark Humphry and Sandra Allison Heldsinger, "Common Structural Design Features of Rubrics May Represent a Threat to Validity," *Educational Researcher*, Vol. 43, No.5, 2014, pp. 256-257, 261-262, https://www.jstor.org/stable/24571208 (accessed June 11, 2019).

[28] D. Royce Sadlre, "Indeterminacy in the use of preset criteria for assessment," *Assessment & Evaluation in Higher Education*, Vol. 34, No.2, 2009, p. 166, https://www.researchgate.net/publication/248966020_Indeterminacy_in_the_use_of_preset_criteria_for_assessment_and_grading (accessed Aug. 15, 2019).

[29] Because world languages portfolios are rated on a 1 to 9 scale, they were unaffected by the addition of scoring levels 6 and 7. World languages portfolios were therefore omitted from percentage calculations of affected portfolios.

[30] Analysis was conducted under the assumption that level 6 and 7 performance scores would translate to a level 5 performance score in their absence. It was also assumed that the score resolution process would terminate with the same reviewer that assigned the final scores in 2018-19, e.g., if an expert reviewer assigned the final scores, then that reviewer's scores were used to generate the counterfactual figures.

[31] "Getting Started with Student Growth Portfolios," a handout provided by Gail Goldberg, consultant to Maryland Department of Education, based on guidelines for 2011-12 pilot program at Queen Anne County Public Schools, Oct. 22, 2012, p.1.

[32] Tennessee Department of Education, "2019-20 TEAM Portfolio Guidebook for Administrators and Teachers," p. 5, https://gallery.mailchimp.com/b28b453ee164f9a2e2b5057e1/files/2c7bf67d-7cd3-4172-9388-8f577293f968/2019_20_TEAM_Portfolio_Guidebook_for_Administrators_and_Teachers.01.pdf (accessed Oct. 17, 2019).

[33] Tennessee Department of Education, *TEAM Portfolio Resource Guide, Early Grades, 2019-20*, p. 15, https://gallery.mailchimp.com/b28b453ee164f9a2e2b5057e1/files/91515d2f-72b3-48db-8660-f7590db6c6b6/2019_20_Early_Grades_TEAM_Portfolio_Resource_Guide.02.pdf (accessed Aug. 27, 2019).

[34] Barbara M. Moskal and Jon A. Leydens, "Scoring Rubric Development: Validity and Reliability," *Practical Assessment, Research, & Evaluation*, Vol. 7, No.10, 2000, p.4, https://www.researchgate.net/publication/242637290_Scoring_Rubric_Development_Validity_and_Reliability (accessed Aug. 27, 2019); Matthew Graham, Anthony Milanowski, and Jackson Miller, *Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings*, Center for Education Reform, 2012, p. 15, https://files.eric.ed.gov/fulltext/ED532068.pdf (accessed Nov. 10, 2019).

[35] *Final Report on the Queen Anne's County Portfolio Pilot: The Design, Development, and Preliminary Piloting of a Student Growth Portfolio as a Measure of Teacher Effectiveness*, provided by Gail Goldberg, consultant to Maryland Department of Education, p. 16, circa 2012.

[36] Some academics consider interrater agreement as a method to measure interrater reliability. Other academics define interrater reliability as a measure of consistency when evaluators determine rank order or relative standing. This definition would exclude interrater agreement as a rater reliability measure since portfolio scoring is based on score categories rather than rank ordering of students. OREA conducted an analysis of rank correlation between the mean collection growth scores assigned by teachers and the mean collection growth scores assigned by the first peer reviewers. The collection growth scores were found to be positively correlated with a Kendall's tau correlation coefficient of approximately 0.5.

[37] An exception at the low end for the number of work samples per collection is the fine arts portfolio model, in which a single collection may have only two work samples, a Point A and Point B sample of a group performance. At the high end, the early grades models include ELA blended standards in which scores are assigned to three strands within one blended standard, at both Point A and Point B, for three students.

[38] Matthew Graham, Anthony Milanowski, and Jackson Miller, *Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings*, Center for Educator Compensation Reform, p. 9, https://files.eric.ed.gov/fulltext/ED532068.pdf (accessed Nov. 12, 2019). It should be noted that the acceptability guidelines referenced in this source most readily apply to situations with external reviewers and may not be as applicable to assessments that employ self-scoring.

[39] About 10 percent of incomplete submissions were not associated with a specific portfolio model and were therefore omitted from the calculated percentages.

[40] OREA also conducted an analysis that factored out chance agreement between teachers' self-scores with those of their first peer reviewers. This analysis weighted corresponding scores that

were 1 point apart as being ¾ correct. Any score differences that were more than 1 point apart were assigned a weight of 0. For early grades portfolios, agreement rates factoring out chance agreement with the aforementioned weights fell within general research guidelines of acceptability.

[41] OREA interviews with school officials from Clay County, Cocke County, Decatur County, Dyer County, Hancock County, Marshall County, Metro Nashville Public Schools, Polk County, Rhea County, Robertson County, Scott County, Sevier County, Shelby County, and Weakley County, fall 2019.

[42] State Board of Education, Teacher and Principal Evaluation Policy 5.201, effective Jan. 29, 2016, p.2.

[43] State Board of Education, Teacher and Administrator Evaluation Policy 5.201, second reading July 28, 2017, p.2, https://www.tn.gov/content/dam/tn/stateboardofeducation/documents/meetingfiles/7-28-17_IV_B_Teacher__Principal_Evaluation_Policy_5_201_Attachment_Clean_Copy.pdf (accessed Nov. 8, 2019).

[44] State Board of Education, Teacher and Administrator Evaluation Policy 5.201, revision 1st reading July 27, 2018, p. 3, https://www.tn.gov/content/dam/tn/stateboardofeducation/documents/2018_sbe_meetings/june-27%2c-2018-sbe-meeting/7-27-18%20II%20K%20Teacher%20%26%20Administrator%20Evaluation%20Policy%205.201%20Attachment%20Clean%20Copy.pdf (accessed Nov. 8, 2019).

[45] Daniel Koretz, *The Vermont Portfolio Assessment Program: Interim Report on Implementation and Impact*, Office of Educational Resource and Improvement, Washington, D.C., p. 10, https://eric.ed.gov/?id=ED351345 (accessed Oct. 4, 2019); Karen Shakman, Nicole Breslow, Julie Kochanek, Julio Riordan, and Tom Haferd, *Changing Cultures and Building Capacity: An Exploration of District Strategies for Implementation of Teacher Evaluation Systems*, Learning and Teaching Division Education Development Center, Inc., pp. 8, 14, 15, https://eric.ed.gov/?id=ED566724 (accessed June 24, 2019).

[46] Calculations using data from the National Center for Education Statistics, 2016-17, and from Tennessee Advisory Commission on Intergovernmental Relations, *Just How Rural or Urban are Tennessee's 95 Counties? Finding a Measure for Policy Makers*, Aug. 2016, https://www.tn.gov/content/dam/tn/tacir/documents/2016JustHowRuralOrUrban.pdf (accessed Oct. 10, 2019).

[47] E. Ross and K. Walsh, *State of the States 2019: Teacher and Principal Evaluation Policy*, National Council on Teacher Quality, Washington, D.C., 2019, pp. 3, 5, https://www.nctq.org/pages/State-of-the-States-2019:-Teacher-and-Principal-Evaluation-Policy (accessed Oct. 8, 2019).

[48] Public Chapter 376, 2019, https://publications.tnsosfiles.com/acts/111/pub/pc0376.pdf (accessed Jan. 8, 2020).

Office of Research and Education Accountability

Russell Moore | *Director*
425 Fifth Avenue North
Nashville, Tennessee 37243
615.401.7866
www.comptroller.tn.gov/OREA/