

Delving into the Data: The Mathematics of Teacher Effectiveness

David Scott

Director of Accountability and Research

Tucson Unified School District

Arizona's Framework for Evaluating Educator Effectiveness

Key elements of the Framework:

- Teacher evaluation instruments should integrate student academic progress data with data derived through classroom observations – neither should stand alone
- Academic progress data shall account for at least 33% of evaluation outcomes and shall not exceed 50%
- The classroom observation component of the evaluation shall account for between 50% and 67% of evaluation outcomes

Arizona's Framework for Evaluating Educator Effectiveness

Academic progress portion of the evaluation:

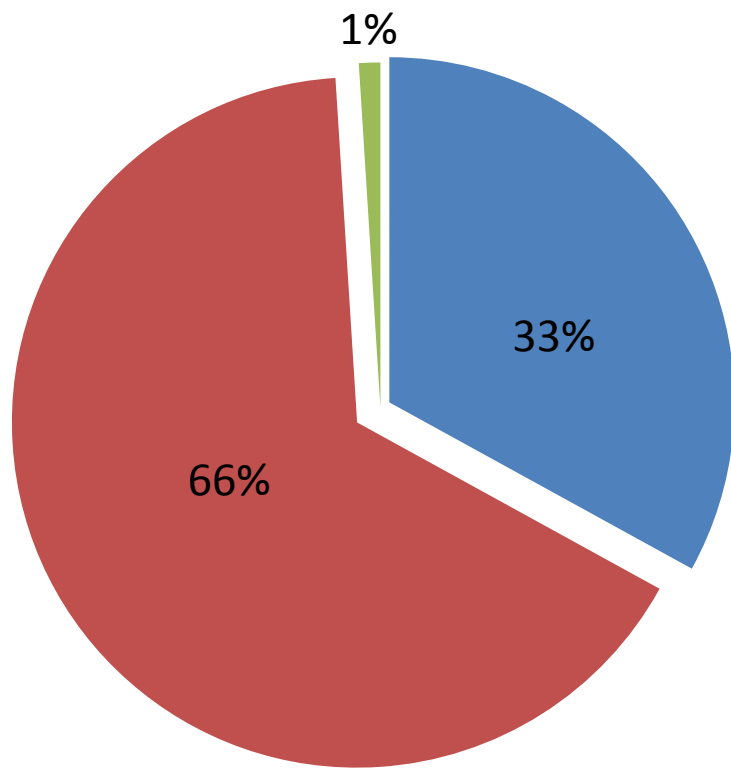
- Teachers shall be divided into two groups (A and B)
- Group A teachers are those with available classroom-level student achievement data that are valid and reliable, aligned to Arizona's academic standards, and appropriate to individual teachers' content areas
- If available and appropriate to a teacher's content area, data from statewide assessments (e.g. AIMS, SAT 10, etc.) shall be used as at least one of the classroom-level data elements

Arizona's Framework for Evaluating Educator Effectiveness

The classroom observation component of the evaluation must:

- be based upon multiple classroom observations
- include rubrics that are aligned to national teaching standards

Tucson Unified School District 2013-14 Teacher Evaluation Model



Teacher Evaluation Components

- Student Growth (33%)
- Danielson Framework (66%)
- Teacher Self Reflection (1%)

Student Growth (33%)

Student Growth (33%)

To be included in the TUSD growth calculation a student must:

- have two years of data (grade 2 Stanford 10, grades 3-8 and high school AIMS)
- be enrolled in the school for 100 days prior to the test
- be enrolled in a teacher's course for 85% of the scheduled class periods (Group A)

Student Growth (33%)

To receive a Group A growth calculation a teacher must:

- teach math, language arts or both
- have at least 15 students remaining in his/her student group after the previously listed student inclusion rules are applied

All remaining teachers are placed in Group B

Student Growth (33%)

- Growth is calculated for reading and math separately
- Reading and math scores are standardized
- A change score is calculated by subtracting the standardized pre-score from the standardized post-score

Student Growth (33%)

- Change scores are averaged at either the school level (Group B) or teacher level (Group A), and a 95% confidence interval is calculated
- The confidence interval is used to construct three categorical groups:
 - Below Average (1)
 - Average (2)
 - Above Average (3)

Student Growth (33%)

Group B values and Group A values for teachers who teach both reading and math are the average of the reading and math categorical values. Averaging expands the possible growth points scale from three to five values:

1.0 - Low

1.5 - Medium-Low

2.0 - Medium

2.5 - Medium-High

3.0 - High

Student Growth (33%)

Based on prior year impact data, it was estimated:

- that no more than 25% of TUSD teachers would receive a Group A growth calculation
- that the overall growth distribution (groups A/B combined) would be:

11% (1.0) Low

12% (1.5) Medium-Low

51% (2.0) Medium

16% (2.5) Medium-High

11% (3.0) High

Teacher Performance (66%)

Teacher Performance (66%)

- During the 2012-13 school year, principals piloted a District developed observation rubric.
- The rubric was found to be time consuming and difficult to bring to scale.
- After consulting with other districts, TUSD decided to adopt the Danielson Framework for Teaching as the measure of teacher performance, for the 2013-14 school year.

Teacher Performance (66%)

- To facilitate the evaluation process, TUSD contracted with a Danielson partner, Teachscape Inc., for evaluator training and software-system support.
- TUSD principals completed approximately 80 hours of professional development and were required to pass an observer certification test before they could begin using the Teachscape/Danielson system to evaluate teachers.

Teacher Performance (66%)

The Danielson Framework has four Domains with 22 discrete measures

Domain 1: Planning and Preparation (6)

- 1a Demonstrating Knowledge of Content and Pedagogy
- 1b Demonstrating Knowledge of Students
- 1c Setting Instructional Outcomes
- 1d Demonstrating Knowledge of Resources
- 1e Designing Coherent Instruction
- 1f Designing Student Assessments

Domain 2: Classroom Environment (5)

- 2a Creating an Environment of Respect and Rapport
- 2b Establishing a Culture for Learning
- 2c Managing Classroom Procedures
- 2d Managing Student Behavior
- 2e Organizing Physical Space

Domain 4: Professional Responsibilities (6)

- 4a Reflecting on Teaching
- 4b Maintaining Accurate Records
- 4c Communicating with Families
- 4d Participating in a Professional Community
- 4e Growing and Developing Professionally
- 4f Showing Professionalism

Domain 3: Instruction (5)

- 3a Communicating With Students
- 3b Using Questioning and Discussion Techniques
- 3c Engaging Students in Learning
- 3d Using Assessment in Instruction
- 3e Demonstrating Flexibility and Responsiveness

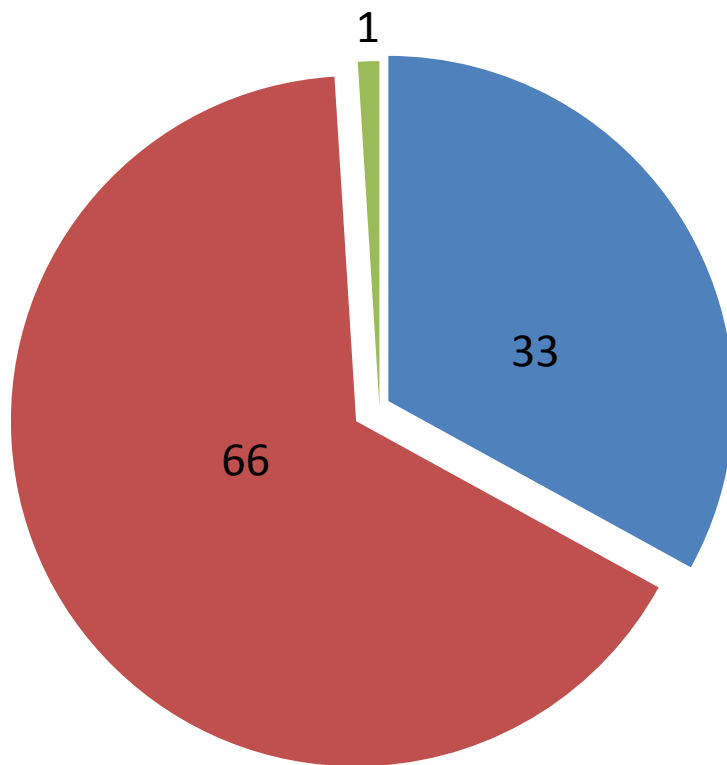
Teacher Performance (66%)

Each of the 22 Danielson Framework elements are scored on a four point rubric:

- Unsatisfactory
- Basic
- Proficient
- Distinguished

Putting it all Together

Teacher Evaluation Components



Total Points = 100

■ Student Growth

- Possible Growth Values (1.0, 1.5, 2.0, 2.5, 3.0)
- Multiply by 11 and round
- **Resulting score range (11, 17, 20, 28, 33)**

■ Danielson Framework

- Twenty two elements
- Assign 0-3 as rubric values (Unsatisfactory = 0, Basic = 1, Proficient = 2, Distinguished = 3)
- **Resulting score range: 0-66**

■ Teacher Self Reflection

- **One point for completion**

Establishing Cuts

Establishing Cuts

TUSD Teacher Performance Categories:

- Ineffective (0-39 points)
- Developing (40-55 points)
- Effective (56-73 points)
- Highly Effective (74-100 points)

Establishing Cuts

The tension between two important questions:

- Face validity

Will teachers and the public understand and agree with the standards?

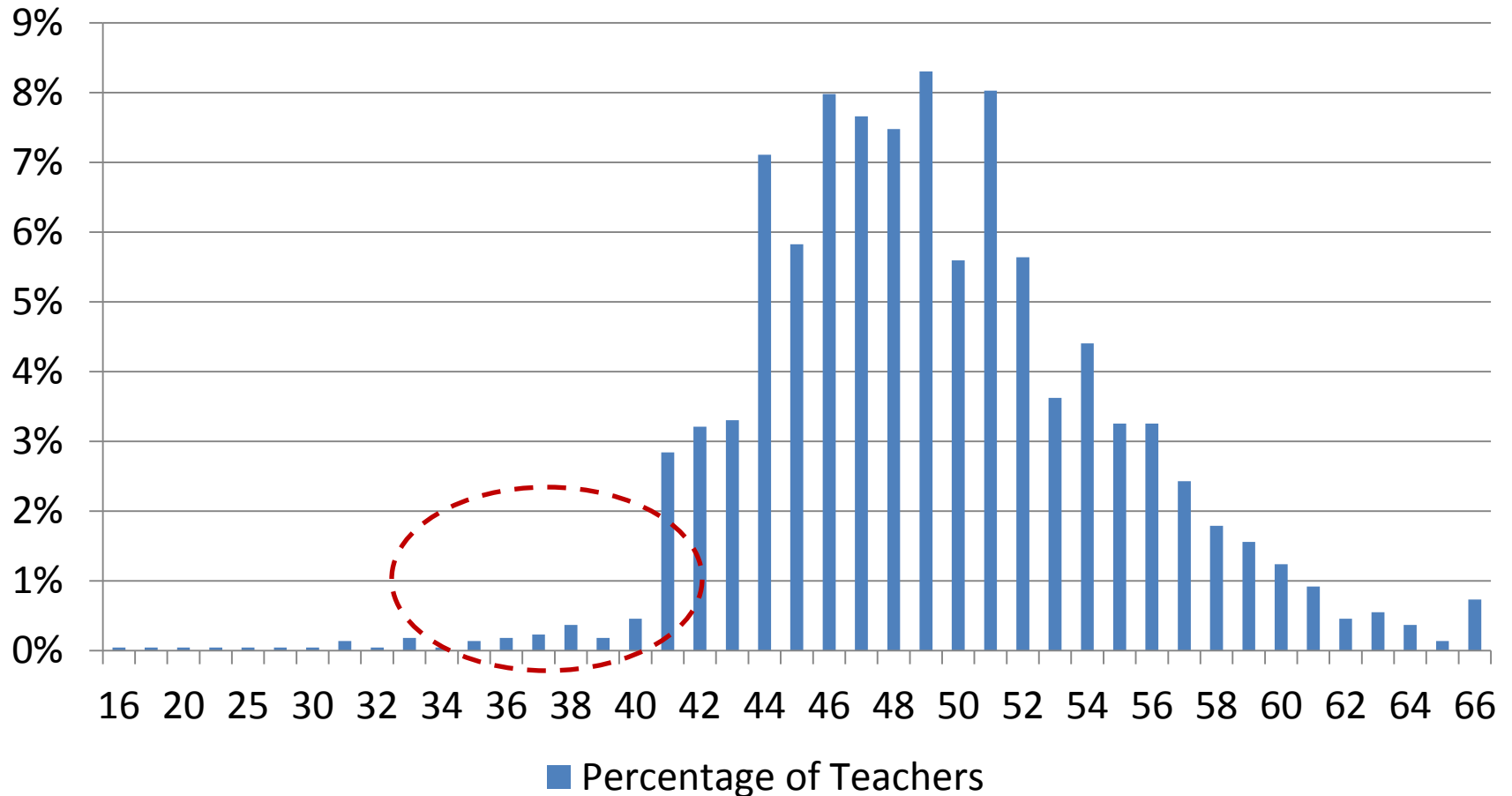
- Consequential validity

Will the outcomes categorize teachers into manageable groups?

Establishing Cuts – Impact Data

Comparison District – Danielson Framework Total Score

(Unsatisfactory=0, Basic=1, Proficient=2, Distinguished=3)



(Mean=49, Standard Deviation=5.6)

Establishing Cuts – Impact Data

TUSD Growth Groups vs. Comparison District – Danielson Framework Total
(with proposed cuts - based on distribution statistics)

| TUSD Growth Group | Percentage of Teachers | Growth Points | Developing | Effective | Highly Effective |
|-------------------|------------------------|---------------|-------------------|---|------------------|
| Low | 10% | 11 | 50 51 52 53 54 55 | 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 | |
| Medium-Low | 12% | 17 | 50 51 52 53 54 55 | 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 | |
| Medium | 51% | 22 | 50 51 52 53 54 55 | 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 | |
| Medium-High | 16% | 28 | 50 51 52 53 54 55 | 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 | |
| High | 11% | 33 | 50 51 52 53 54 55 | 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 | |

= Mean
 = 1st Standard Deviation

Establishing Cuts - Effective

Examples of a minimum Effective score = 56

- Minimum Teacher Growth Points = 11
- Teacher Self Reflection = 1
- Teacher is Proficient = 44
 - Proficient ($22 * 2 = 44$)

- Average Teacher Growth Points = 22
- Teacher Self Reflection = 1
- Teacher is Equally Basic and Proficient = 33
 - Basic ($11 * 1 = 11$)
 - Proficient ($11 * 2 = 22$)

...less than these, and a teacher is rated Developing

Establishing Cuts - Developing

Examples of a minimum Developing score = 40

- Minimum Teacher Growth Points = 11
- Teacher Self Reflection = 1
- Teacher is more than one quarter Proficient = 28
 - Basic ($16*1=16$)
 - Proficient ($6*2=12$)

- Average Teacher Growth Points = 22
- Teacher Self Reflection = 1
- Teacher is more than three quarters Basic = 17
 - Unsatisfactory ($5*0=0$)
 - Basic ($17*1=17$)

...less than these, and a teacher is rated Ineffective

Establishing Cuts – Highly Effective

Examples of a minimum Highly Effective score = 74

- Minimum Teacher Growth Points = 11
- Teacher Self Reflection = 1
- Teacher is greater than 80% Distinguished = 62
 - Proficient ($4*2=8$)
 - Distinguished ($18*3=54$)

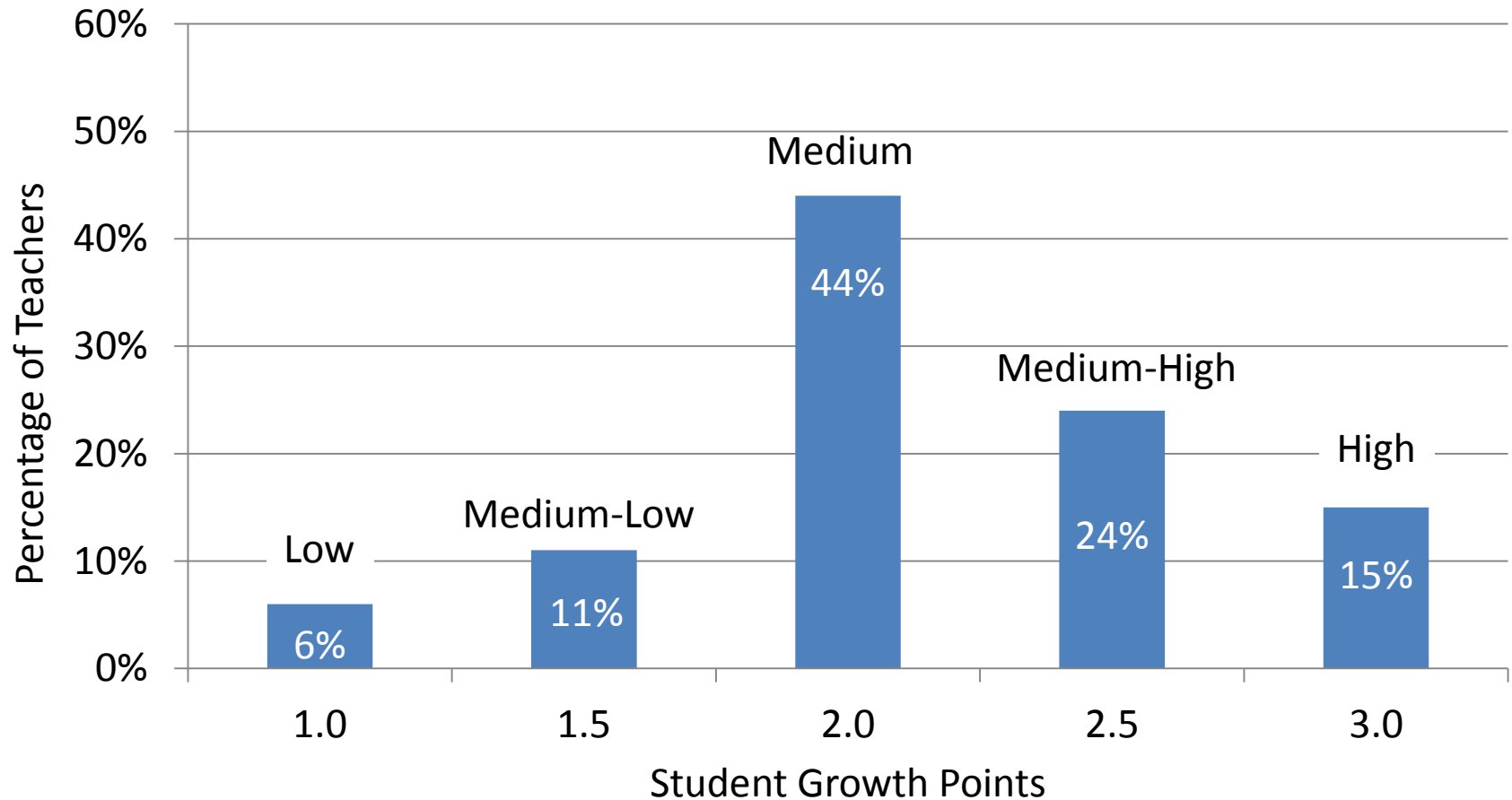
- Average Teacher Growth Points = 22
- Teacher Self Reflection = 1
- Teacher is greater than 30% Distinguished = 51
 - Proficient ($15*2=30$)
 - Distinguished ($7*3=21$)

...less than these, and a teacher is rated Effective

Preliminary Results

TUSD Preliminary 2013-14 Results

Distribution of Growth Points (Group A and Group B Combined)



TUSD Preliminary 2013-14 Results

Danielson Framework - Domain 2: Classroom Environment

| Item | Item Description | Unsatisfactory | Basic | Proficient | Distinguished | Total | Average Score |
|------|--|----------------|-------|------------|---------------|-------|---------------|
| | | % | % | % | % | N | |
| 2a | Creating an Environment of Respect and Rapport | <1% | 5% | 58% | 36% | 1786 | 3.3 |
| 2b | Establishing a Culture for Learning | <1% | 8% | 68% | 23% | 1787 | 3.1 |
| 2c | Managing Classroom Procedures | 1% | 8% | 68% | 23% | 1786 | 3.1 |
| 2d | Managing Student Behavior | 1% | 8% | 61% | 31% | 1789 | 3.2 |
| 2e | Organizing Physical Spaces | <1% | 3% | 74% | 23% | 1766 | 3.2 |
| | Total | <1% | 7% | 66% | 27% | 8914 | 3.2 |

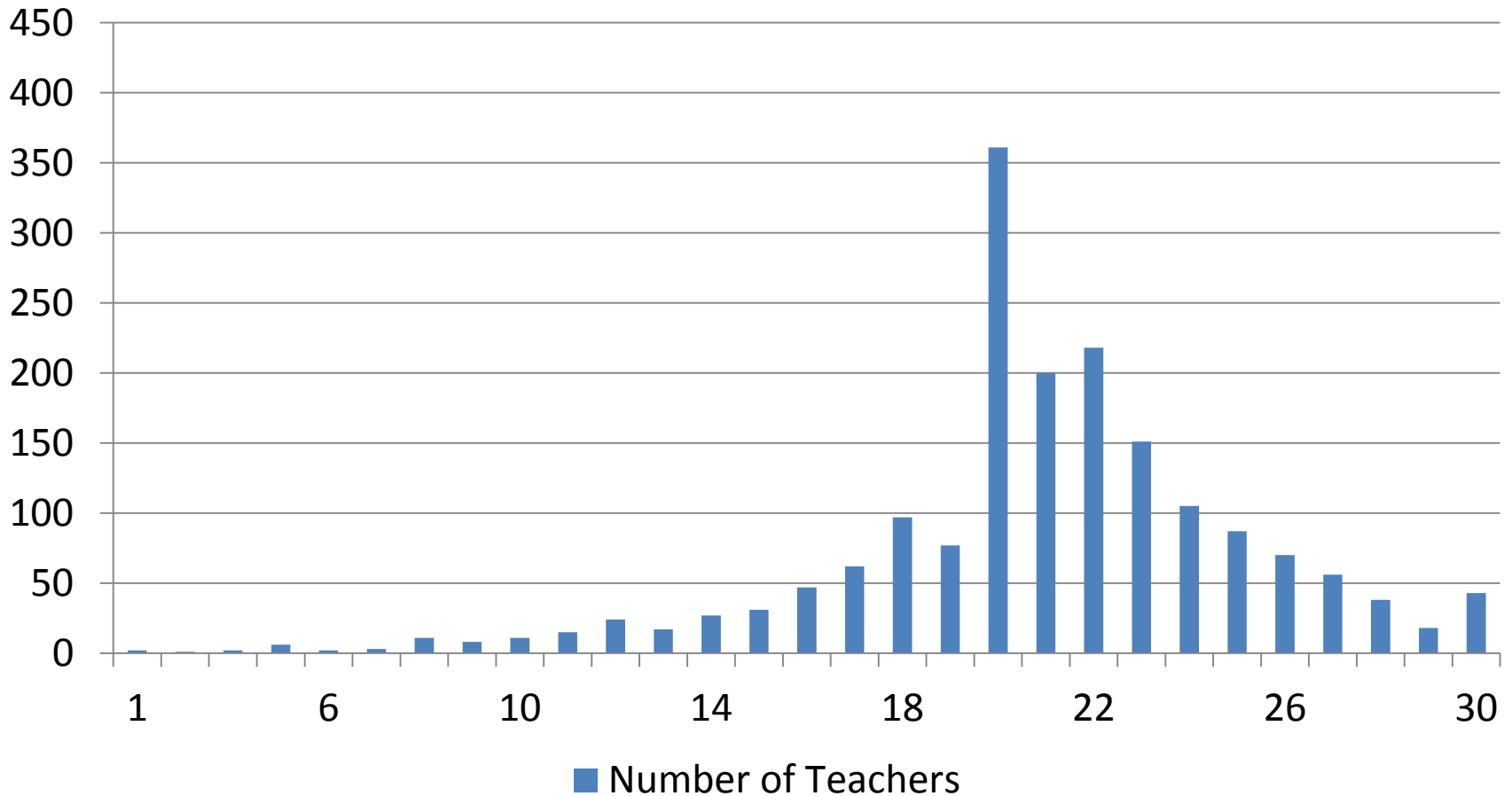
TUSD Preliminary 2013-14 Results

Danielson Framework - Domain 3: Instruction

| Item | Item Description | Unsatisfactory | Basic | Proficient | Distinguished | Total | Average Score |
|------|--|----------------|-------|------------|---------------|-------|---------------|
| | | % | % | % | % | N | |
| 3a | Communicating With Students | <1% | 8% | 68% | 24% | 1790 | 3.2 |
| 3b | Using Questioning and Discussion Techniques | 2% | 17% | 71% | 10% | 1781 | 2.9 |
| 3c | Engaging Students in Learning | 1% | 11% | 70% | 18% | 1790 | 3.1 |
| 3d | Using Assessments in Instruction | 1% | 11% | 76% | 12% | 1766 | 3.0 |
| 3e | Demonstrating Flexibility and Responsiveness | 1% | 6% | 73% | 19% | 1708 | 3.1 |
| | Total | 1% | 11% | 72% | 17% | 8835 | 3.0 |

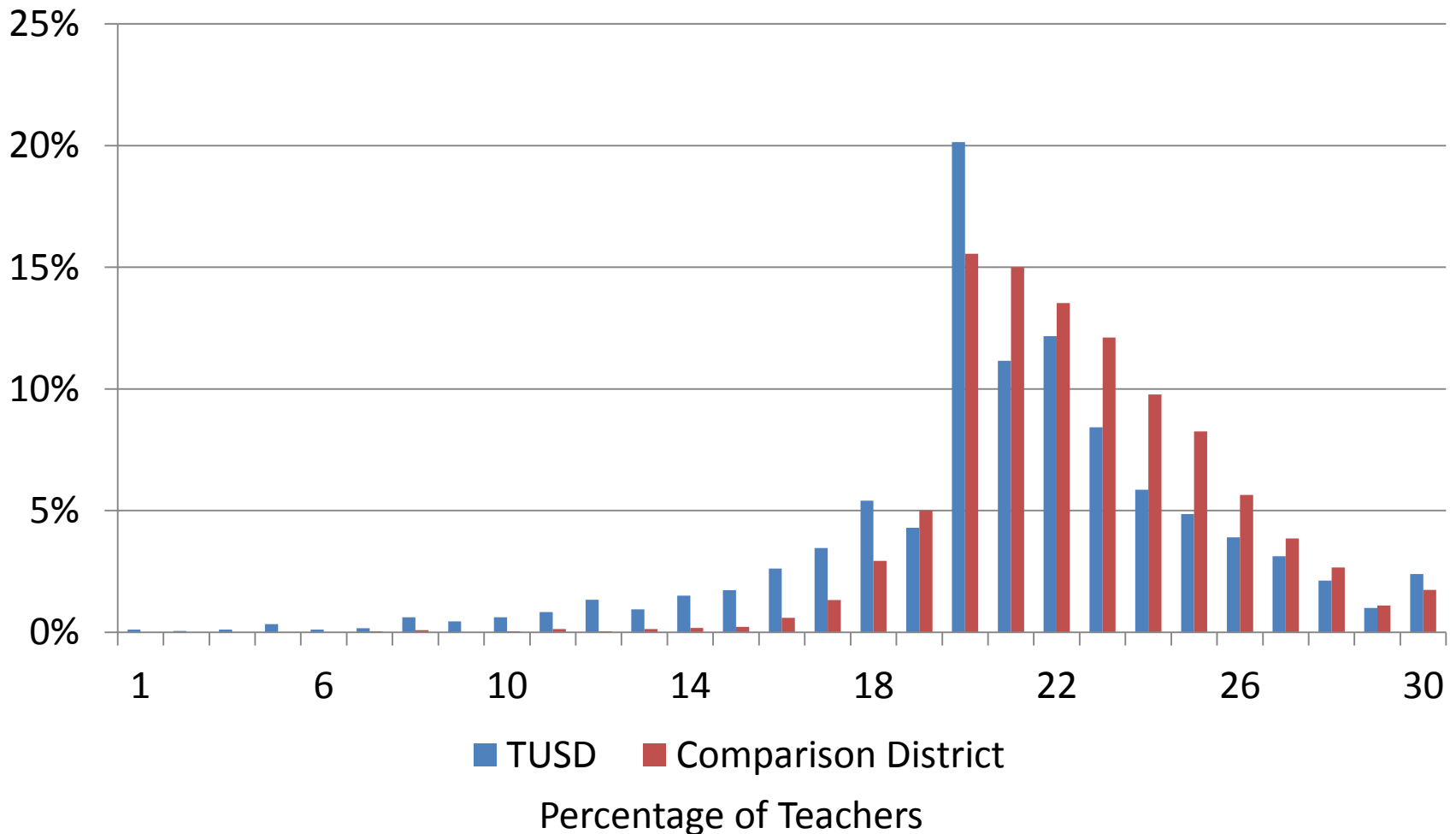
TUSD Preliminary 2013-14 Results

Domains 2 & 3 Combined



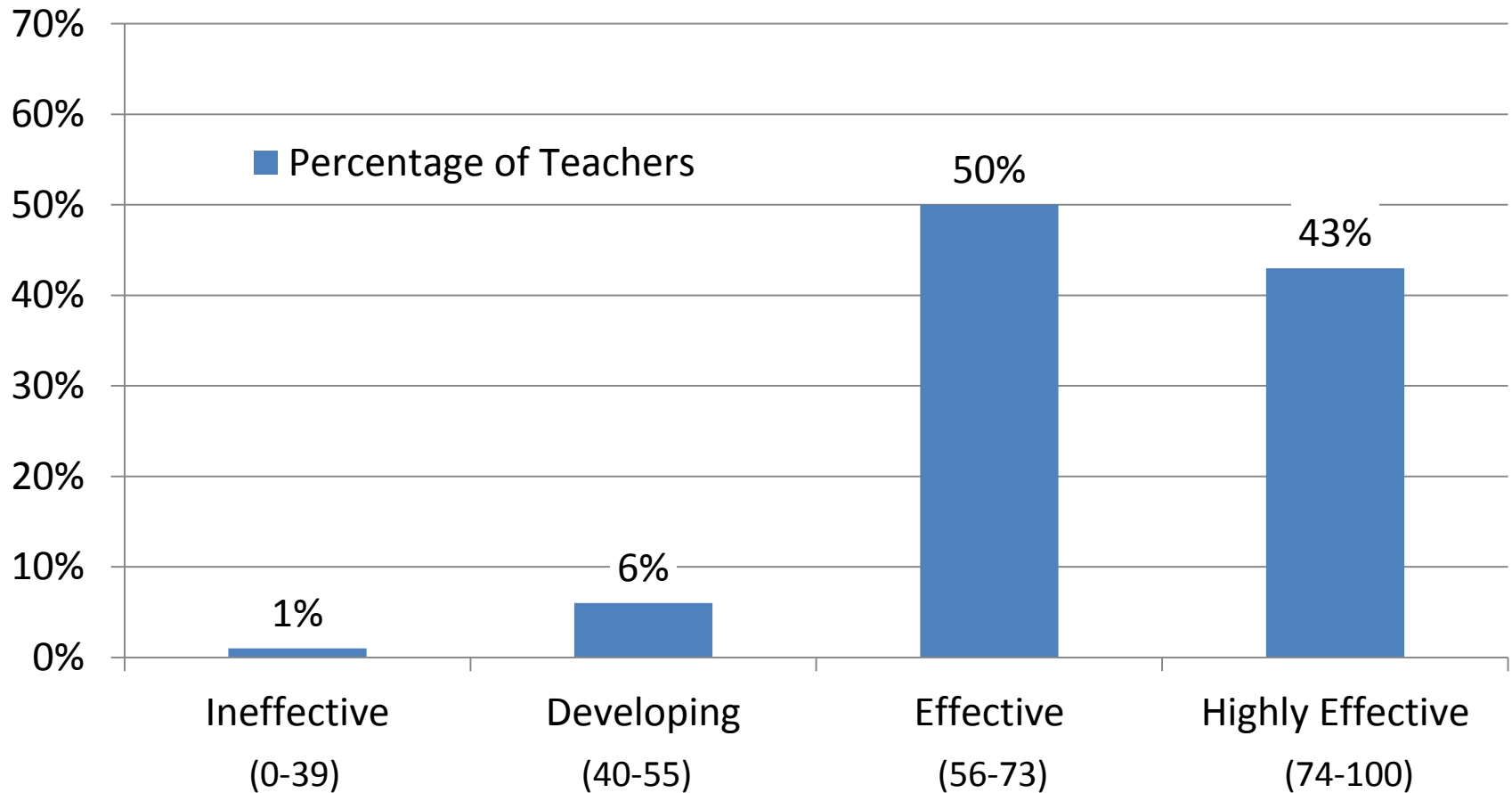
TUSD Preliminary 2013-14 Results

TUSD vs. Comparison District - Domains 2 & 3 Combined



TUSD Preliminary 2013-14 Results

*Estimated Distribution – Growth Points + Self Reflection Point + 22*Average of Domains 2&3*



TUSD Preliminary 2013-14 Results

*Estimated Distribution – Growth Points + Self Reflection Point + 22*Average of Domains 2&3*

