



School Climate
IMPROVEMENT

Resource Package

Data Interpretation Guide

on Making School Climate Improvements

Second Edition
2018

National Center on Safe Supportive Learning Environments



Safe Supportive Learning
Engagement | Safety | Environment

Disclaimer

This *Data Interpretation Guide* was designed and written under the U.S. Department of Education (Department) Contract Number EDESE12O0035 by American Institutes for Research, the National Center on Safe Supportive Learning Environments (NCSSLE). Rita Foy Moss served as the contracting officer's representative (COR) for the NCSSLE technical assistance center. This document contains resource materials that are provided for the user's convenience. The inclusion of these materials is not intended to reflect their importance, nor is it intended to endorse any views expressed, or products or services offered. These materials may contain the views and recommendations of various subject matter experts as well as hypertext links, contact addresses, and websites to information created and maintained by other public and private organizations. The opinions expressed in any of these materials do not necessarily reflect the positions or policies of the Department. The Department does not control or guarantee the accuracy, relevance, timeliness, or completeness of any outside information included in these materials.

Second Edition
2018

This document is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this product is not necessary, the following citation is preferred:

U.S. Department of Education, Office of Safe and Healthy Students. (2018). *Data interpretation guide on making school climate improvements*. Washington, DC: Author.

This document is available free of charge at <https://safesupportivelearning.ed.gov/sites/default/files/SCIRP/datainterpretationguide.pdf>.

Availability of Alternate Formats

Requests for documents in alternate formats such as Braille or large print should be submitted to the Alternate Format Center by calling 202.260.0852 or by contacting the 504 coordinator via e-mail at om_eeos@ed.gov.

Notice to Limited-English-Proficient Persons


If you have difficulty understanding English, you may request language assistance services for Department information that is available to the public. These language assistance services are available free of charge. If you need more information about interpretation or translation services, please call 1-800-USA-LEARN (1.800.872.5327) (TTY: 1.800.877.8339) or e-mail us at ED.Language.Assistance@ed.gov. You also can write to U.S. Department of Education, Information Resource Center, LBJ Education Building, 400 Maryland Ave. SW, Washington, DC, 20202.



Contents

Introduction	1.
Interpreting Survey Results	2
Discussion Guides and Additional Resources	3
A Multi-Tiered Approach to Improving School Climate	3
Interpreting Your Data	4
Contents of This Guide	4
Scale Scores	6.
Benchmarked Scale Scores	6
Overall Benchmarked Scale Scores	7
Benchmarked Scale Scores by Respondent Characteristics	7
Interpreting Benchmarked Scale Scores	8
Discussion Guides for Interpreting Topic Area Benchmarked Scale Scores	9
Item-Level Data	10.
Comparing Item Averages (Means)	10
Reverse Coding Negatively Valenced Items	11
Interpreting Item-Level Data	11
Conclusion	12.
Appendix A. Comparisons for Districts	13.
Interpreting Benchmarked Scale Scores: Direction, Support, and Resources for EDSCLS Districts	13
Interpreting Item-Level Data: Direction, Support, and Resources for Districts	15
Appendix B. Comparisons for Schools	19.
Interpreting Benchmarked Scale Scores: Direction, Support, and Resources for EDSCLS Schools	19
Interpreting Item-Level Data: Direction, Support, and Resources for Schools	20
Appendix C. Average (Mean) Topic Area Values	25.
Average (Mean) Topic Area Values	25
Overall Average (Mean) Topic Area Values	25
Average (Mean) Topic Area Values Broken Out by Respondent Characteristics	25
Interpreting Average (Mean) Topic Area Values	26
Instructions for Calculating Average (Mean) Topic Area Values or Item- Content Groups (Mean of Means)	26
Interpreting Average (Mean) Topic Area Values: Direction, Support, and Resources for Districts	27
Interpreting Average (Mean) Topic Area Values: Direction, Support, and Resources for Schools	29
Appendix D. Glossary of Terms	31.
Appendix E. What Not to Compare Using Legacy Scale Scores	32

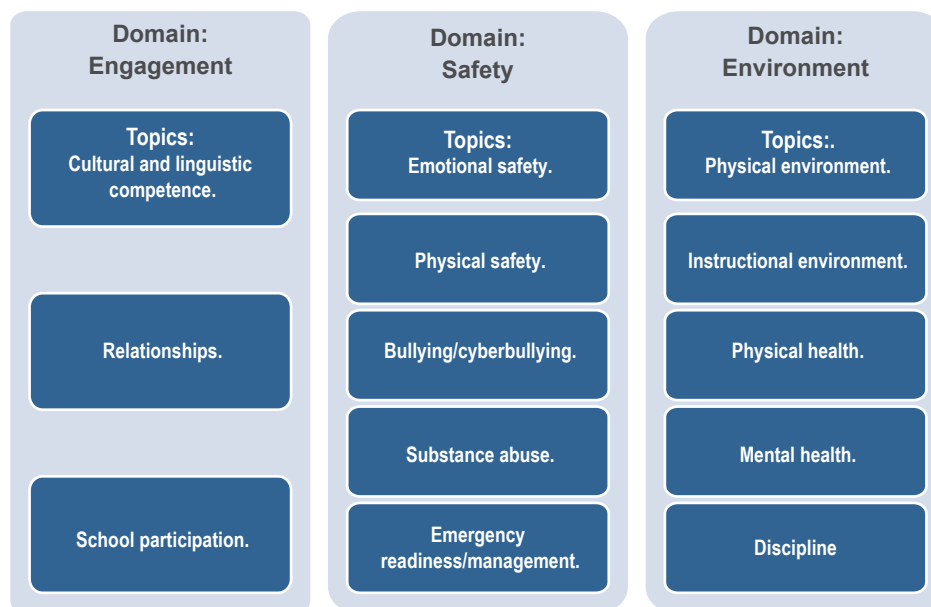
Introduction

Congratulations! You have made an important commitment to improve the climate in your state, district, or school.¹ Measuring and understanding how students, staff, and parents/guardians perceive the climate in their schools are key steps in making wise decisions on how to use resources to focus on areas in need of improvement. This document contains detailed information and resources to help you interpret and use results from the school climate surveys you have administered. It is intended for users of any school climate survey, but sections intended specifically for users of the U.S. Department of Education's (ED's) School Climate Surveys (EDSCLS) are marked with .


Who should read this guide? The person(s) at your state, district, or school who will be most actively involved in the interpretation of the survey findings, who will translate the data into action, and who will communicate with those who will promote data-driven decision making for school climate improvement should read this guide. The information is technical in nature. Thus, if you or your team does not have technical expertise in interpreting data, you may wish to use this document as a reference while using the [EDSCLS Data Analysis Worksheet](#) instead

We used the school climate model developed by ED to develop the EDSCLS.² The model includes three key school climate domains composed of 13 subdomains or topic areas, as shown in Figure 1. Whether you use the EDSCLS or another survey that includes constructs similar to those in this model, this guide is for you! If you use the EDSCLS surveys, look for the EDSCLS logo throughout to see information specific to them.

Figure 1. Model of School Climate: Domains and Topic Areas.



¹ This document provides strategies applicable to public schools and districts (including charter authorizers, charter management organizations, education management organizations, individual charter schools, and charter local educational agencies), as well as private schools.

²  The EDSCLS surveys and Web-based platform were developed by the National Center for Education Statistics in collaboration with the Office of Safe and Healthy Students (OSHS), both part of ED.



As shown in Figure 1, the domain of Engagement comprises several actionable topic areas—Cultural and Linguistic Competence, Relationships, and School Participation—while the other two domains, Safety and Environment, comprise actionable topic areas of their own.

Sites using the EDSCLS should note three things about this model and related results:

1. **Physical Health** is shown in the model as one of the topic areas within the Environment domain. However, although the data for instructional and non-instructional staff formed a Physical Health scale that can be used for analytical purposes, they did not form a Physical Health scale for students.
2. The data for the **parent/guardian survey** did not form a scale at the topic area or domain level because of the brevity of the survey. Thus, parent data should be examined at only the item level (i.e., analysis of individual survey questions), explained more fully below.
3. Although **Emergency Readiness/Management (ERM)** is shown as one of the topic areas within the Safety domain, it was not designed to form a scale; thus, ERM data should be examined at only the item level.

Interpreting Survey Results



If you use the EDSCLS platform, it will generate scale scores and item-level data for you overnight after the close of data collection. Beginning with the release of the EDSCLS platform Virtual Machine (VM) 3.0 in December 2017, the way the platform reports scale scores changed. Based on psychometric benchmarking, the scale scores are now set at 300 and 400 cut scores to separate “performance” levels 1, 2, and 3. The EDSCLS platform generates graphs that show your scale score results embedded into the following three “performance” levels:

- Level 1: Least Favorable (scale scores below 300);
- Level 2: Favorable (scale scores 300–400); and
- Level 3: Most Favorable (scale scores above 400–500).

The benchmarking allows users to compare the levels into which scale scores fall no matter what the topic or domain is. For example, if students in a school are in benchmark level 1 (Least Favorable) for Engagement and benchmark level 2 (Favorable) for Safety, it suggests they perceive the school is not doing as well in engaging students as it is in providing for safety. Please note, although the performance levels are comparable across the three EDSCLS domains (Engagement, Safety, and Environment), scale scores themselves are not.

This guide has been updated to reflect benchmarking in the sections discussing scale scores. See [Benchmark Performance Levels](#) for further information on the EDSCLS benchmarking.

If you administered a survey other than the EDSCLS, you can use this guide as you interpret and use your survey results. Although the survey you administered may not use the same labels



- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

to identify topics as does the EDSCLS model (e.g., “emotional safety” or “cultural and linguistic competence”), many school climate surveys include constructs that are similar in composition and, most importantly, are actionable. You will find this guide and its link to a [Discussion Guide](#) for each topic area (along with other helpful resources) especially useful when you are considering appropriate actions your educational agency can take to improve climate based on your data.

Using EDSCLS data for multiple purposes: Schools and districts are also required to report information about school climate pursuant to ED’s Civil Rights Data Collection (CRDC). Information collected by this survey may help schools and districts prepare their responses to the CRDC survey. More information about the CRDC can be found at ocrdata.ed.gov.

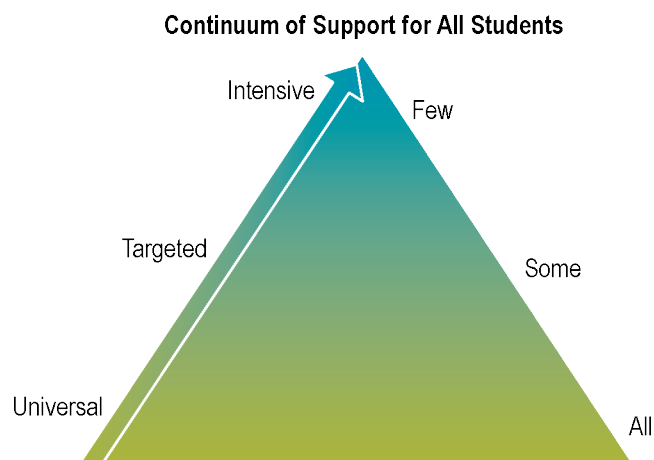
Discussion Guides and Additional Resources

At the end of each of the following sections, you will find links to a [Discussion Guide](#) for each topic area. It includes a brief description of the topic area and initial and deeper guiding questions that you can use to examine and interpret your results for that topic. Each [Discussion Guide](#) also contains links to webpages with interventions that you can begin to implement immediately and use as a foundation for further work, as well as links to longer term intervention resources.

A Multi-Tiered Approach to Improving School Climate

The directions and resources in this document are presented using a “multi-tiered” approach. A multi-tiered system comprising various levels of support (universal, targeted, and intensive) can be conceptualized as a triangle, or a continuum within a triangle, as shown in Figure 2.³

Figure 2 Multi-Tiered System of Support: A Continuum Within a Triangle.



As illustrated in Figure 2, all students are served in a multi-tiered system, with more general support at the base, becoming gradually more intensive advancing up the triangle. The base, which is the largest tier, illustrates that universal interventions and supports are provided schoolwide, to

³ Office of Special Education Programs and OSHS Technical Assistance Center on Positive Behavior Interventions and Supports. (2014). Multi-tiered behavior frameworks: Developing implementation capacity [PowerPoint slides]. Retrieved from https://www.pbis.org/Common/Cms/files/pbisresources/mtbf_overview_sctg_28oct2014_GSugai.pptx



all students and other stakeholders in the school community. The middle tier, which comprises a smaller area, suggests that interventions and supports are provided to a subset of those in need of more targeted help (often provided in small-group sessions). The top tier, which comprises the tip of the triangle, illustrates that interventions and supports are delivered to those in need of more intensive help (usually provided on a one-to-one basis).

Following this section, you will find information about interpreting your data as well as links to [Discussion Guides](#) that will help you use data in a universal or targeted way to improve school climate in specific topic areas. Click on the School Climate Improvement Resource Package (SCIRP) [Reference Manual](#) for additional information about multi-tiered systems of support. More information also is available on websites such as www.pbis.org/ and www.rtinetwork.org/.

Supports should be designed to improve school climate for the students who are most in need regardless of the subgroup(s) to which they belong. Targeting supports based on need as opposed to membership in a subgroup will support compliance with relevant civil rights laws.

Interpreting Your Data

As you begin to review your survey results, you may feel overwhelmed and like you are “drowning in data.” If so, you may want to look at your survey results one step at a time, perhaps in four or five “chunks,” by respondent group (students, instructional staff, noninstructional staff, parents/guardians)⁴ or by school climate topic areas (e.g., physical safety, relationships). For example, EDSCLS users can use the [EDSCLS Data Analysis Worksheet](#) for this purpose.

Contents of This Guide

In this guide, you will find descriptions of and suggestions for interpreting results from different types of data:


- Scale scores (overall and by respondent characteristics),
- Item-level data, and
- Average (mean) item grouping values (overall and by respondent characteristics).⁵



For EDSCLS users, we provide examples when examining the results and making comparisons, and how those comparisons can help you improve climate in your district or school.



If you are a **district or local educational agency (LEA)** hosting the survey administration, you will find direction on interpreting your district data, comparing across schools in your district, and comparing individual school data to that of your district.

⁴  The EDSCLS parent/guardian survey includes some items that have a Don't Know response. Analyzing those data can help districts and schools identify areas in need of effective communication with families.

⁵ Some users who administer surveys may find it helpful to calculate and use average (mean) item grouping values. See [Appendix C](#) for further information on how to calculate and use these values.




- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores



If you are one of a number of **schools** in a district that is hosting the survey administration, you will find suggestions for interpreting your school data, and how to compare them to your district if those data are available.

- If you are a **state educational agency (SEA)** hosting the survey administration, you can follow the support provided to districts in this guide, plus you can examine district- to-district comparisons as well as school-to-school comparisons.

At the end of each section, you will be able to link to universal/schoolwide or targeted topic area [Discussion Guides](#) that provide links to webpages and resources for interventions and strategies for each of the school climate topic areas.



ARE YOU A SINGLE SCHOOL HOSTING A SCHOOL CLIMATE SURVEY?

Comparisons throughout this document focus on school-to-school and school-to-district comparisons. If you are a single school hosting the EDSCLS or another survey, you will not have other school and district data for comparison purposes. Non-EDSCLS sites may wish to focus on comparing data over time and on item-level data comparisons and respondent subgroup comparisons (see the section [Interpreting Item-Level Data: Direction, Support, and Resources for Schools](#) in Appendix B).

Scale Scores

Scale scores are the premier way that EDSCLS as well as many other school climate surveys measure school climate. A scale score, which combines multiple survey items related to different aspects of a topic area, is a more robust measure than attempting to measure that topic by asking about it with a single item.



For sites administering EDSCLS, scale scores are produced for each topic area or domain as a number on a scale of 100 to 500. These scores can be thought of in the same way you might think about an SAT score or a credit score, in that higher numbers represent more positive results, but the real meaning is in how a scale score compares with others. Scale scores are produced by the platform overall for each respondent group and by respondent characteristics (e.g., male students and female students).

Benchmarked Scale Scores



Beginning in December 2017, the way the EDSCLS platform (VM 3.0 and higher) reports scale scores changed. Three perceived school climate levels—Least Favorable, Favorable, and Most Favorable—were created based on the psychometric property of the survey items. These levels allow users to compare respondents' perceptions of school climate across domains, topics, and respondent groups (e.g., you can compare topic areas across the three EDSCLS domains: Engagement, Safety, and Environment). To set the cutoff scores at the fixed values, a slightly different transformation was used in converting raw estimates to the new scale scores. The three perceived school climate levels and their definitions are listed here:

- Level 1: Least Favorable (scale scores below 300): The most likely answer to each positively valenced question in the scale is *disagree* or *strongly disagree*; the most likely answer to each negatively valenced question in the scale is *agree* or *strongly agree*.⁶
- Level 2: Favorable (scale scores 300–400): The most likely answer to each positively valenced question in the scale is *agree*; the most likely answer to each negatively valenced question in the scale is *disagree*.
- Level 3: Most Favorable (scale scores above 400–500): The most likely answer to each positively valenced question in the scale is *strongly agree*; the most likely answer to each negatively valenced question in the scale is *strongly disagree*.

In the graphs that the platform produces, scale score results are embedded into the three benchmark levels (see Figure 3).



CAUTIONS: You cannot compare the scale scores themselves across domains; only the levels in which they fall can be compared in this way. (However, you can compare the scores themselves within domains.) Also, you cannot compare older “legacy” scale scores to newer benchmarked scale scores!

⁶ Valenced in a positive direction means that agreement with the statement is a positive perception (e.g., “Teachers understand my problems.”). Valenced in a negative direction means that agreement with the statement is a negative perception (e.g., “I sometimes stay home because I don’t feel safe at this school.”).



- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

Because any EDSCS scale scores produced by platforms lower than VM 3.0 are not recalibrated, the scale scores they produced, called “legacy” scores, cannot be compared to the new benchmarked scale scores. If you want to preserve a trend line, you can convert your older legacy scores from prior years to benchmarked scale scores and then compare across years. You will find information on how to convert legacy scores to benchmarked scores at <https://safesupportivelearning.ed.gov/edscls/benchmarks>.



Districts will be able to view their scale scores at the district level or at the school level for any school that participated in the district’s survey administration.



Schools will be able to see their own scale scores as well as those from their district, if the district makes them available to schools.

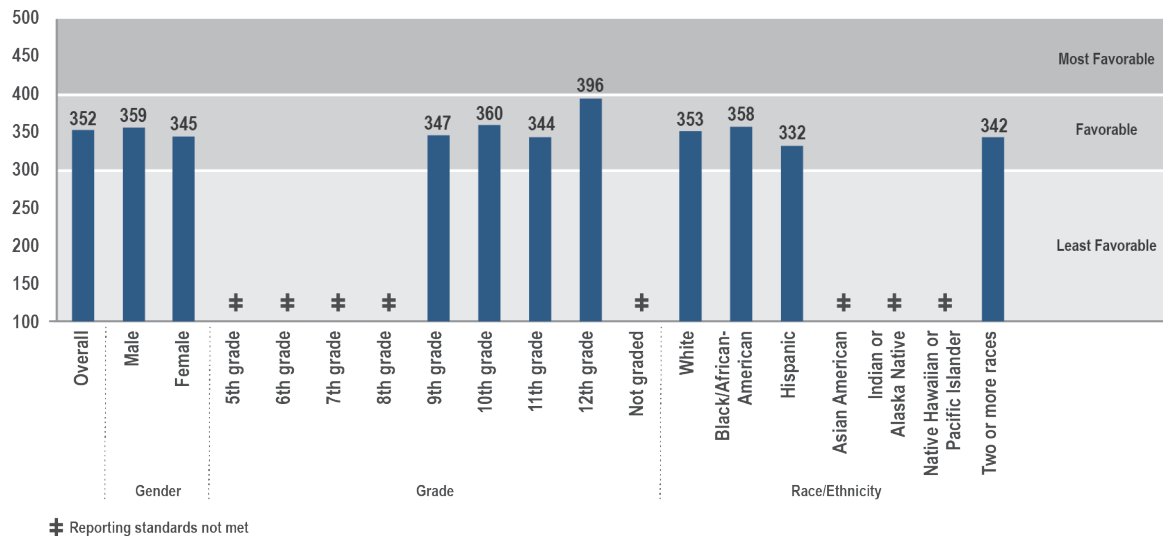
Overall Benchmarked Scale Scores



For each respondent group, overall benchmarked scale scores are presented by the EDSCS platform in a chart similar to the example in Figure 3.

This example shows that the overall benchmarked scale score for students in this school for Engagement is 352 (on a scale of 100–500), within the Favorable performance level.

Figure 3 Average Engagement by Gender, Race/Ethnicity, Grade, and Benchmark Level



Benchmarked Scale Scores by Respondent Characteristics



Benchmarked scale scores broken out by respondent subgroups provide a richer set of data, a way to see how perceptions of a topic area differ across subgroups of students and staff.

- Introduction
- Scale Scores**
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

Scale scores are produced for EDSCLS users for the following subgroups⁷:

- Student scale scores per topic area can be examined by:
 - Gender,
 - Race/ethnicity, and
 - Grade.
- Staff (instructional and noninstructional) scale scores per topic area can be examined by:
 - Gender and
 - Race/ethnicity.

In the event of a possible disclosure risk that would allow a sole respondent or fewer than 10 of a subgroup of respondents to be identified (e.g., if there is only one Asian teacher in the school), the EDSCLS platform will suppress the results for that subgroup (i.e., results for that subgroup will not be shown). If the data are suppressed (and thus, not shown) for a respondent subgroup that is important to your district or school, there are other ways that you can collect overall school climate or specific topic area data. For example, you may want to convene focus groups or conduct targeted interviews or student “fishbowls.”⁸ Click on the SCIRP [Reference Manual](#) to go to additional information about these methods of collecting data. (You also will have a chance to click on *Discussion Guides* to go to specific resources in topic areas of interest in Figure 4.)

For each respondent group, benchmarked scale scores for subgroups by respondent characteristics can be viewed in a chart similar to the example in Figure 3.

This example shows that the overall benchmarked scale score for Engagement for students in this school is 352 (on a scale of 100–500), while the scale score for male students is 359 and the score for female students is 345. All are within the Favorable benchmark level, with male students scoring higher than female students or the students overall in this school. The figure also shows that 12th-grade students have a more positive perception of the Engagement domain (396) than do students from other grades.

Interpreting Benchmarked Scale Scores



Appendices A (for districts) and B (for schools) include suggestions for interpreting overall benchmarked scale scores and by respondent characteristics. They also include suggested comparisons that can be made using these scores and how those comparisons can help you improve school climate.



If you are a **district**, click on [Appendix A, Table A-1](#), to go to interpretation of benchmarked scale scores, overall and by respondent characteristics.



If you are a **school**, click on [Appendix B, Table B-1](#), to go to interpretation of benchmarked scale scores, overall and by respondent characteristics.

⁷ Note that the EDSCLS platform does not produce data by crossed demographics (e.g., Asian female students).

⁸ A “fishbowl” is a discussion style in which one group (students) has a dialogue, and another group (school staff) listens to their discussion. Afterwards, both groups come together to discuss themes to work on together.



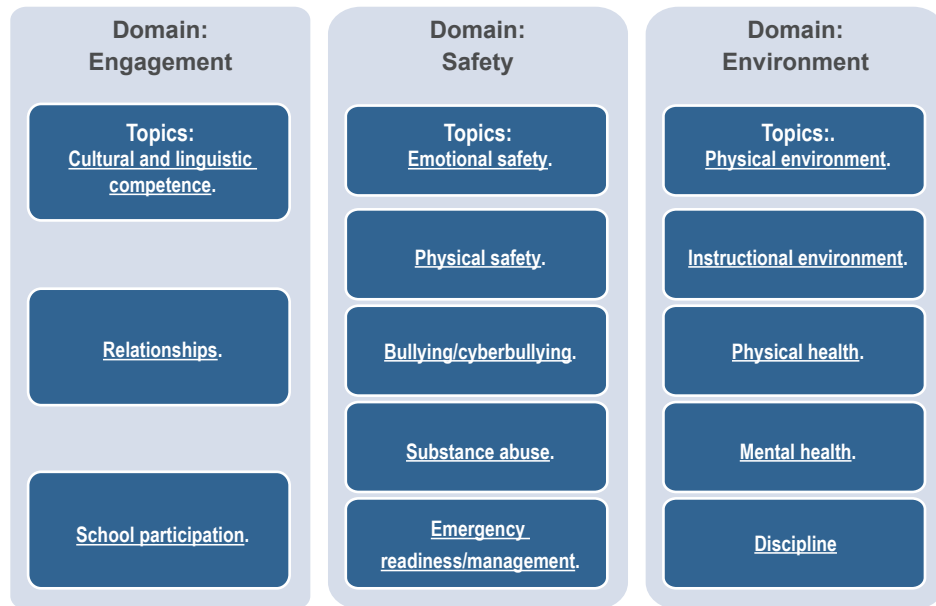
Discussion Guides for Interpreting Topic Area Scale Scores


You can click on any of the topic area labels in Figure 4 to go to that topic area's [Discussion Guide](#) for using scale scores, whether benchmarked or not. Each [Discussion Guide](#) provides *universal/schoolwide* guidance through interpreting the overall scale score for that topic area. Each also includes *targeted* support in the “By respondent characteristics” section, which guides you through interpreting scale scores by respondent characteristics for the topic area.

? **WHAT IS IN EACH DISCUSSION GUIDE?**

You will find a brief description of the topic area as well as a set of guiding questions and links to topic area webpages with shorter and longer term intervention resources pertinent to that topic area.

Figure 4 Links to Topic Area *Discussion Guides*^{9, 10}.



⁹  For EDSCLS users, no scale is produced for the emergency readiness and management topic. However, percentage distributions and item averages (means) are produced for individual items by the Web-based platform.

¹⁰  For EDSCLS users, no physical health scale score is produced for students; however, it is produced for instructional and noninstructional staff.

- Introduction
- Scale Scores**
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

Item-Level Data

In addition to examining scale scores, looking at item-level data may help a district or school dig deeper into the data to target specific areas or issues. Item-level results can often provide schools with concrete examples of the underlying topic area construct that may be more actionable, warranting more immediate implementation of interventions included in the topic area webpages. Item-level results can also suggest areas in which to begin planning and preparation for longer term interventions and strategies.



Item-level values reflect the EDSCLS response options (values of 1–4). Sites can produce the percentage distribution of those response options and their average (mean) for each item in the survey that is included in topic area scales produced by the platform. Item-level results can also be produced for important items that are not included in the final topic area scales. They are marked for you as “stand-alone” items in the questionnaires and codebooks.

Item-level values reflecting the EDSCLS response options (values of 1–4) can be viewed in a chart like the example in Figure 5.

Figure 5 Example of Item Percentage Distribution and Mean Chart.

Relationships

Hide details		PERCENT				
	SURVEY ITEM	STRONGLY AGREE (4)	AGREE (3)	DISAGREE (2)	STRONGLY DISAGREE (1)	MEAN
	Teachers understand my problems.	13.64	50	27.27	9.09	2.68

This example shows that when presented with the item “Teachers understand my problems,” 13.64 percent of students reported that they strongly agree (value 4), 50 percent reported that they agree (value 3), 27.27 percent reported that they disagree (value 2), and 9.09 percent reported that they strongly disagree (value 1). The average (mean) of this item’s values, calculated as the average of all responses to the item (2.68), also is provided in the right column of the chart.

Comparing Item Averages (Means)

The example in the chart is a positively valenced item (“Teachers understand my problems”). “Valenced” in a positive direction means that agreement with the statement is a positive perception. Out of all of the students who responded to this item, the average response was 2.68 (where *strongly agree* = 4 and *strongly disagree* = 1), indicating that students’ perceptions, on average, were more positive than negative. (Click on [Figure C-1](#) for further information on interpreting item favorability.)

However, some items may be negatively valenced (e.g., “I sometimes stay home because I don’t feel safe at this school.”). “Valenced” in a negative direction means that agreement with

the statement is a negative perception. Because the numeric values of both negative and positive items are on the same scale of *strongly agree* = 4 and *strongly disagree* = 1, the item means for negatively and positively valenced items should not be directly compared without first reverse coding the negative items.



For EDSCLS users, these negatively valenced items are marked for you in the questionnaires and codebooks.

Reverse Coding Negatively Valenced Items



Negatively valenced items have been reverse-coded for the purpose of calculating scale scores. *However, the item-level results that the platform produces are not reverse-coded.* If you want to compare item averages (means) across all items or a group of items (both positively and negatively valenced), you should first reverse-code the values of negatively valenced items so that they are directed in the same way as positively valenced items. To reverse-code those items, use the following formula:

$$\text{Reverse-Coded Item Mean} = [5 - \text{item mean}]$$

For example, if the mean of a negatively valenced item is 3.4, the reverse-coded mean would be calculated as $[5 - 3.4 = 1.6]$.

After negatively valenced items are reverse-coded, you can compare them with means of positively valenced items.

Interpreting Item-Level Data

Appendices A (for districts) and B (for schools) provide suggestions for using item-level data to assess how favorably (positively) or unfavorably (negatively) respondents perceive items. This information can be used by both EDSCLS and other survey users.



If you are a **district**, click [Appendix A, Table A-2](#), to go to interpretation of item-level data.



If you are a **school**, click [Appendix B, Table B-2](#), to go to interpretation of item-level data.

These appendices include information to help you interpret item-level data, including how:

1. An item is perceived by individual respondent groups;
2. An item is perceived across respondent groups, but only for items worded exactly the same for each group (called “comparable items”); and
3. Items organized by content within a topic area (called “item-content groups”) are perceived by individual respondent groups.

The appendices also provide suggestions for districts and schools on how item-level data can be used to improve school climate, and limitations to be aware of when using item-level data.

Conclusion

Your commitment to using a data-driven approach to improving school climate is a key first step in your journey toward a positive educational environment for students, staff, and parents. In this document, we have presented information and resources for interpreting and using results of survey data to identify areas for improvement and to work toward providing universal/schoolwide support (using overall scale scores and item-level data). We also have provided assistance in interpreting and using data to help you identify specific groups within the school community who may need more targeted support using scale scores parsed out by respondent characteristics. In addition, we have included links to a [Discussion Guide](#) for each of 13 school climate topic areas that include (1) an introduction; (2) guiding questions for you to consider as you work through your data; and (3) links to webpages with resources that you can implement immediately and in the longer term. In [Appendix C](#), we present additional information about calculating, using, and interpreting average (mean) values for topic areas and other item groupings for users who are interested in using these types of data.

Introduction

Scale Scores

Item-Level Data

Conclusion

Appendix A:
Comparisons
for Districts

Appendix B:
Comparisons
for Schools

Appendix C:
Average (Mean)
Topic Area Values

Appendix D:
Glossary of Terms

Appendix E:
What Not to
Compare Using
Legacy Scale
Scores

Appendix A. Comparisons for Districts



Interpreting Benchmarked Scale Scores: Direction, Support, and Resources for EDSCLS Districts

Table A-1 provides direction, support, and resources to EDSCLS **districts** on interpreting benchmarked scale scores overall and by student and staff characteristics.¹¹ It includes example comparisons that can be made and how they can help districts improve school climate. The benchmarked scale scores have been anchored to fixed cut scores across the EDSCLS domains and topic areas. Thus, although the topic area scale scores themselves can be compared *within* domains, the perceived school climate levels in which they fall can be compared *across* domains and topic areas and across respondent groups and subgroups.

Table A-1 Benchmarked Scale Scores: Comparisons for Student and Staff Data.



You Can Compare:

- The district's benchmarked scale scores within domains and "performance" levels across topic areas and domains.

Example: Within the Engagement domain, you can compare benchmarked scale scores themselves for its topic areas: Cultural and Linguistic Competence (CLC), Relationships, and Participation. Across domains, you can compare the levels in which the scores fall; for example, you can compare the level for Relationships (in the Engagement domain) with the level for Physical Safety (in the Safety domain).

- Each school's overall benchmarked scale score for each domain or each topic area with the district's overall score.

Example: You can compare a school's benchmarked scale score on Engagement with the district's benchmarked scale score on that same domain.

- Benchmarked scale scores produced for each domain or topic area across schools in the district.

Example: You can compare benchmarked scale scores for all district schools with each other for each of the 12 topic areas.

- The district's benchmarked scale scores for each domain or topic area between categories of respondent characteristics within the same set of respondents.

Example: You can compare district benchmarked scale scores for female students versus male students on Physical Safety (a topic area) or on Safety (a domain).

¹¹ Readers should note that you cannot compare older "legacy" scale scores to newer benchmarked scale scores. For those who wish to preserve a trend line started with legacy scale scores, you will find information on how to convert legacy scores to benchmarked scores at <https://safesupportivelearning.ed.gov/edscls/benchmarks>. If you are working only with older legacy data, see [Appendix E](#) for information on comparisons that cannot be made using legacy scale scores.

- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts**
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

- Benchmarked scale scores by respondent characteristics within the same set of respondents for specific domains or topic areas across schools in the district.

Example: You can compare male instructional staff's benchmarked scale score on CLC with that of other schools.

- Benchmarked scale scores within domains for subgroups within the same set of respondents, and levels for topic areas across domains and subgroups across sets of respondents.

Example: You can compare district benchmarked scale scores for Black or African-American students on Emotional Safety, Physical Safety, Bullying/Cyberbullying, and Substance Abuse within the Safety domain. Across domains and respondent subgroups, you can compare the levels in which scores fall for Hispanic students on Discipline (in the Environment domain) with White teachers on CLC (within the Engagement domain).

- The levels in which benchmarked scale scores fall across respondent groups or subgroups and across domains or topic areas.

Example: Benchmark scale score level for Hispanic noninstructional staff on the Physical Safety topic area (in the Safety domain) compared with the level for White instructional staff on the CLC topic area (in the Engagement domain).



Use These Comparisons To:

- Provide universal/schoolwide support across the district to students and staff for topic areas that have low benchmarked scale scores relative to those for other topic areas within the same domain.

Example: If the district's benchmarked scale score on student perceptions of CLC is low compared with student benchmarked scale scores on Relationships or Participation (all within the Engagement domain), the district can focus supports on improving CLC for students in schools across the district.

- Provide universal/schoolwide supports to schools for topic areas that have low benchmarked scale scores relative to the district score.

Example: For schools that have staff benchmarked scale scores on CLC that are low compared with the district's benchmarked scale score on CLC, the district can work on improving CLC with staff in those schools.

- Provide universal/schoolwide supports to schools for topic areas that have low benchmarked scale scores relative to other schools in the district.

Example: For schools that have lower benchmarked scale scores for staff on Discipline than other schools in the district, the district can help those schools improve conditions related to Discipline for staff.

- Provide targeted support across the district to subgroups of students and staff for topic areas that have low benchmarked scale scores relative to other subgroups within that set of respondents on the same or other topics within domains.

Example: If the district's benchmarked scale score for male students on Instructional Environment is low compared with those of female students' benchmarked scale scores, the district can focus supports on improving Instructional Environment for boys in the district.

- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts**
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

- Provide targeted support to schools that have low benchmarked scale scores for specific subgroups on topic areas relative to their district benchmarked scale scores on the topic area.

Example: For schools whose female staff’s benchmarked scale scores on Emotional Safety are low compared with those of the districts’ female staff’s benchmarked scale scores, the district can help those schools improve Emotional Safety–related issues for their female staff.

- Provide targeted supports on topic areas to schools that have low subgroup benchmarked scale scores relative to those of other subgroups within that set of respondents in other schools in the district.

Example: For schools where Hispanic noninstructional staff report lower benchmarked scale scores on CLC than in other schools in the district, the district can focus help in those schools to improve the CLC climate for Hispanic noninstructional staff.

Interpreting Item-Level Data: Direction, Support, and Resources for Districts

This section provides suggestions for **districts** when using item-level data to assess how favorably (positively) or unfavorably (negatively) respondents perceive items. This information can be used by both EDSCS and other survey users.

Table A-2 provides direction, support, and resources for **districts** on interpreting item-level data overall and by respondent characteristics. It includes comparisons that can be made using item-level data, how these comparisons can help districts improve school climate, and cautions or limitations when interpreting the data.

Note that parent/guardian item-level results can be interpreted in the same way as for other respondent groups (students, instructional staff, and noninstructional staff¹²). That is, parent/guardian data can be examined using item-level percentage distributions and averages (means), across respondent groups, if items are worded the same (“comparable items”), and they can be included in item-content groupings (Items organized by content within a topic area).

Table A-2 Using Item-Level Data.




You can examine:

1. **For items per individual respondent group (students, instructional staff, noninstructional staff, and parents/guardians), you can examine:**

- How favorably a single item is perceived, on average (rated 1–4), across the district per respondent group.

Example: A district can see how favorably students across the district perceive the single item, “This school provides instructional materials (e.g., textbooks, handouts) that reflect my cultural background, ethnicity, and identity.”

¹²  The EDSCS parent/guardian survey includes some items that have a Don’t Know response. Analyzing those data can help districts and schools identify areas in need of effective communication with families.

- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts**
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

- How favorably an item is perceived, on average (rated 1–4), in each school compared with the district and with other schools in the district, per respondent group.

Example: A district can see how favorably, on average, students in a single school perceive the item, “This school provides instructional materials (e.g., textbooks, handouts) that reflect my cultural background, ethnicity, and identity,” compared with other schools in the district or with the district average.

2. For individual items that are worded exactly the same across respondent groups of interest (called “comparable items”), you can examine:

- How a comparable item is perceived, on average, across the district, across respondent groups.

Example: All four respondent groups (students, instructional staff, noninstructional staff, and parents) are asked, “At this school, how much of a problem is student drug use?” Districts can examine how much the responses to this item might differ across these groups.

- The extent of the difference in ratings (1–4) of a comparable item between respondent groups in a specific school, compared with other schools in the district and with the district average (mean).

Example: Do any of the district schools show a substantially greater difference in the percentage of students and staff who reported they strongly agree on an item, compared with the extent of that difference in other district schools or compared with the district average (mean)?

3. For items within a topic area grouped by content (“item-content group”).

- It can sometimes be daunting to examine item-level data because of the sheer number of items and response options; therefore, it is useful to organize items by content area and examine these groupings. (Suggested item-content groupings can be found in each topic area’s *Discussion Guide*, and instructions for calculating the average [mean] of the items can be found in [Appendix C](#)). In this way, specific content areas within topics with particularly low favorability ratings may be more readily identified; districts and schools may want to target these content areas for immediate interventions.

For example, items in the Bullying/Cyberbullying topic area might be grouped by content areas such as bullying is a problem; bullying related to race, ethnicity, and religion; cyberbullying; bullying related to sexuality; bullying related to physical and mental disability; bullying prevention: student reporting and stopping of bullying; and bullying knowledge and formal prevention strategies.

The average (mean) of the item ratings (1–4) for each content group can be compared to examine:

- How favorably an item-content group is perceived overall across the district, compared with other item-content groups in the same topic area.

Example: A district can see how favorably respondent groups in the district perceive the item-content group called “cyberbullying,” compared with the other bullying item-content groups.

- How favorably an item-content group is perceived overall in each school, compared with the district and with other schools in the district.

Example: A district can see how favorably respondent groups in a specific school perceive the cultural diversity of materials compared with the other schools in the district and with the district average.

- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A:
Comparisons
for Districts**
- Appendix B:
Comparisons
for Schools
- Appendix C:
Average (Mean)
Topic Area Values
- Appendix D:
Glossary of Terms
- Appendix E:
What Not to
Compare Using
Legacy Scale
Scores



Cautions/Limitations:

1. **For items per individual respondent group (students, instructional staff, noninstructional staff, and parents/guardians).**
 - It is important to consider that items represent various unique aspects of the overall topic area that is being measured; each has been selected to reflect only one specific aspect of respondent perceptions.

Example: If examining Environment items suggests that bathroom cleanliness in this school is an area for improvement, this information might help contextualize perceptions of the environment in the school. However, although improving bathroom cleanliness might lead to more favorable responses about that particular problem, it may have less of an effect on the important underlying perception of the overall environment in the school.
 - You can think of excessive focus on improving responses to a single item as analogous to a physician attempting to alleviate symptoms without curing the underlying disease or condition.
 - Perhaps as important, when measuring any topic area, the indicators (items) should reflect a representative sample of behaviors and attitudes about the topic. If schools focus change solely on behaviors and attitudes as defined by specific items, they risk a form of unintentional “teaching to the test” in which schools are able to show growth with respect to specific items even though perceptions of the underlying topic area have not changed.
2. **For individual items that are worded exactly the same across respondent groups (called “comparable items”).**
 - Item-level data cannot be compared across respondent groups or subgroups of interest unless they are worded exactly the same. (Also see cautions detailed above.)
3. **For items within a topic area grouped by content (“item-content group”).**
 - In addition to the cautions and limitations described for item-level results above, data users should recognize that item groupings are intended solely as a way to organize a large amount of information and do not represent psychometrically validated scales.



Use These Results To:

1. **For items per individual respondent group (students, instructional staff, noninstructional staff, and parents/guardians).**

In general, if the rating on a specific item is less favorable than you think acceptable and is an important issue to your district and community, you may consider this a specific area to target interventions.

 - Look at items that have low favorability ratings compared with other items in the same domain, within specific respondent groups, to target interventions.

Example: If, on average, students rate the single item, “This school provides instructional materials (e.g., textbooks, handouts) that reflect my cultural background, ethnicity, and identity,” lower than other items in the CLC topic area, a district can, for instance, support its teaching staff to start including more culturally diverse examples in class materials and lessons.
 - Look at items that have low favorability ratings compared with other items across domains, within specific respondent groups, to target interventions.

- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts**
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

Example: If examining Safety domain items in the student survey suggests that there is a problem with students being teased or picked on about race or ethnicity, the district may want to examine other items from other domains to see if related issues are being reported there. For instance, examining Engagement domain items may show that students report few examples of different racial, ethnic, or cultural backgrounds in class lessons. Targeting a quick adjustment on the latter issue, such as instructing teachers to incorporate more culturally diverse examples in materials and lessons, may help to bring a sense of cultural sensitivity to the district.

2. For individual items that are worded exactly the same across respondent groups (called “comparable items”).

- In general, if the rating on a comparable item is more unfavorable for a specific respondent group than for other respondent groups, or if it is more unfavorable for a respondent group in an individual school than that group in the district, and this issue is important to your district and community, you may consider this area one to specifically target for interventions.
- Analyze percentage distributions on comparable items across respondent groups to investigate differences in perceptions of a potential problem or specific area.

Example: Look at the percentage distribution across response options for the student drug-use item (not a problem, small problem, somewhat a problem, large problem). If most students say that it is somewhat or a large problem, but instructional staff are saying it is not a problem or a small problem, you can infer that teachers are not seeing what the students are seeing in school. You can then begin targeting support for staff in regard to this issue.

- Compare averages (means) of comparable items across respondent groups to investigate differences in perceptions of a potential problem or specific area.

Example: Look at the reported average (mean) for the item “I feel safe at this school” for students and for instructional staff. The average (mean) that is reported for this comparable item will range from 1 (*strongly disagree*, or Most Unfavorable) to 4 (*strongly agree*, or Most Favorable). If the average (mean) of student reports is 3.5 and the average (mean) for instructional staff is 2, you can see that teachers view this more unfavorably than do students. Again, you can use this information to target supports concerning this issue to appropriate stakeholders in the school (in this case, teachers).

3. For items within a topic area grouped by content (“item-content group”).

- Looking at groups of items that pertain to similar substantive content is another way of looking at item-level data. This approach may be especially helpful when digging deeper into the data to see which areas might be ripe for immediate intervention.
- For each school climate topic area, we have consulted with content experts to arrive at helpful ways of grouping items. When you click on links to the [Discussion Guides](#), you will see these item-content groupings for each specific topic area.
- Using this approach, districts can see where they can immediately target interventions and where they might be able to wait to implement longer term interventions.

Example: In digging deeper into the CLC topic area data, it may become evident that the item-content group around culturally diverse materials is wanting for students and staff. A quick adjustment can immediately be implemented, such as instructing teachers to incorporate more culturally diverse examples in materials and lessons. If, conversely, a district’s respondents report that the item-content group around equal treatment/respect in the district is favorable, efforts to improve respect and equality may wait until more urgent CLC needs are taken care of.



Appendix B. Comparisons for Schools



Interpreting Benchmarked Scale Scores: Direction, Support, and Resources for EDSCLS Schools

Table B-1 provides direction, support, and resources to EDSCLS schools on interpreting benchmarked scale scores overall and by student and staff characteristics.¹³ It includes comparisons that can be made using benchmarked scale scores, benchmark levels, and how these comparisons can help schools improve school climate. The benchmarked scale scores have been anchored to fixed cut scores of 300 and 400 across the EDSCLS domains and topic areas to separate the three “performance” levels into which scale scores fall. Scale scores can be compared across scales and respondent and subgroups *within* a domain. *Across* domains, comparison can be made using the “performance” levels into which scale scores fall.

Table B-1 Benchmarked Scale Scores: Comparisons for Student and Staff Data.



You Can Compare:

- The school's benchmarked scale scores within and across topic areas within each domain.
Example: Within the Engagement domain, you can compare scores for its topic areas: Cultural and Linguistic Competence, Relationships, and Participation.
- The school's benchmarked scale scores for each domain or topic area between subgroups within the same set of respondents.
Example: You can compare a school's benchmarked scale scores for female students versus male students on Physical Safety (a topic area) or on Safety (a domain).
- The school's benchmarked scale scores for topic areas within domains for subgroups within the same set of respondents or benchmarked scale score levels across domains and across respondents.
Example: You can compare benchmarked scale scores for Black or African-American students on Emotional Safety, Physical Safety, Bullying/Cyberbullying, and Substance Abuse within the Safety domain. Across domains and respondent groups, you can compare benchmarked scale score levels for Hispanic students on Discipline (in the Environment domain) with those of White teachers on CLC (within the Engagement domain).
- The school's benchmarked scale scores with those of the district, if available, for each domain or topic area by respondent characteristics.
Example: You can compare your school's benchmarked scale score for Black or African-American students on CLC with that of the district for Black or African-American students on CLC.
- Benchmarked scale score levels by respondent characteristics for topic areas across domains.
Example: Benchmarked scale score levels for Hispanic noninstructional staff on the Physical Safety topic area (within the Safety domain) compared with Hispanic noninstructional staff's benchmarked scale score levels for the Relationships topic area (within the Engagement domain).

¹³ Readers should note that you cannot compare older “legacy” scale scores to newer benchmarked scale scores. For those who wish to preserve a trend line started with legacy scale scores, you will find information on how to convert legacy scores to benchmarked scores at <https://safesupportivelearning.ed.gov/edscls/benchmarks>. If you are working only with older legacy data, see Appendix E for information on comparisons that cannot be made using legacy scale scores.



Introduction

Scale Scores

Item-Level Data

Conclusion

Appendix A:
Comparisons
for Districts

**Appendix B:
Comparisons
for Schools**

Appendix C:
Average (Mean)
Topic Area Values

Appendix D:
Glossary of Terms

Appendix E:
What Not to
Compare Using
Legacy Scale
Scores



Use These Comparisons To:

- Provide universal/schoolwide support to students and staff for topic areas that have relatively low benchmarked scale scores compared to those of other topic areas.

Example: If a school's student benchmarked scale score on CLC is low compared with the student benchmarked scale scores on Relationships, you can focus supports on improving CLC for students in the school.
- Provide universal/schoolwide support to students and staff for topic areas that have relatively low benchmarked scale scores compared with those of the district, if available.

Example: If your school's benchmarked scale score for teachers on Physical Safety is low compared to that of the district, you can focus support on Physical Safety issues for teachers in your school.
- Provide targeted support on topic areas to subgroups of students and staff whose benchmarked scale scores fall into less favorable levels compared to their levels on other topic areas across domains.

Example: If a school's benchmarked scale score level for Asian students on Physical Safety falls into the least favorable category, but the levels for other topic areas across domains are in the Favorable category for Asian students, the school can target supports on improving Physical Safety for Asian students in the school.
- Provide targeted support to subgroups of students and staff for topic areas that have relatively low benchmarked scale scores compared with the district benchmarked scale scores, if available.

Example: If a school's benchmarked scale score on Relationships is lower for Black or African-American instructional staff than the district's benchmarked scale score on Relationships for Black or African-American instructional staff, the school can target efforts on improving issues related to Relationships for those teachers.
- Provide targeted support to subgroups of students and staff for topic areas that have relatively low benchmarked scale scores for one subgroup compared with another subgroup within the same set of respondents.

Example: If a school's benchmarked scale score for female students on Emotional Safety is lower than that for male students on Emotional Safety, the school can target supports on improving Emotional Safety for girls in the school.



Interpreting Item-Level Data: Direction, Support, and Resources for Schools

This section provides suggestions for schools when using item-level data to assess how favorably (positively) or unfavorably (negatively) respondents perceive items. This information can be used by both EDSCLS and other survey users.

Table B-2 provides direction, support, and resources for schools on interpreting item-level data overall and by respondent characteristics. It includes comparisons that can be made, how these comparisons can help schools improve school climate, and cautions or limitations when interpreting the data.

Introduction

Scale Scores

Item-Level Data

Conclusion

Appendix A:
Comparisons
for Districts

**Appendix B:
Comparisons
for Schools**

Appendix C:
Average (Mean)
Topic Area Values

Appendix D:
Glossary of Terms

Appendix E:
What Not to
Compare Using
Legacy Scale
Scores

- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts
- Appendix B: Comparisons for Schools**
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

Note that parent/guardian item-level results can be interpreted in the same way as for other respondent groups (students, instructional staff, and noninstructional staff¹⁴). That is, parent/guardian data can be examined using item-level percentage distributions and averages (means), across respondent groups if items are worded the same (“comparable items”), and can be included in item-content groupings (items organized by content within a topic area).

Table B-2 Using Item-Level Data.

You Can Examine:

1. For items per individual respondent group (students, instructional staff, noninstructional staff, and parents/guardians), you can examine:

- How favorably an item is perceived, on average (rated 1–4), in the school per respondent group.

Example: A school can see how favorably students perceive the single item, “This school provides instructional materials (e.g., textbooks, handouts) that reflect my cultural background, ethnicity, and identity.”
- How favorably an item is perceived, on average (rated 1–4), in the school compared with the district, if available, per respondent group.

Example: A school can see how favorably, on average, students perceive the item, “This school provides instructional materials (e.g., textbooks, handouts) that reflect my cultural background, ethnicity, and identity” compared with the district average.

2. For individual items that are worded exactly the same across respondent groups (called “comparable items”), you can examine:

- How a comparable item is perceived across respondent groups in the school.

Example: Instructional staff, noninstructional staff, and parents/guardians are asked, “At this school, how much of a problem is student drug use?” Schools can examine how much responses to this item might differ across the groups.
- The extent of the difference in ratings (1–4) to a comparable item between respondent groups in the school, compared with the extent of the difference in the district, if data are available.

Example: Does your school show a substantially greater difference in the percentage of students and staff who reported they *strongly agree* with an item, compared with the extent of that difference in the district?

- 3. For items within a topic area grouped by content (“item-content group”).**
- It can sometimes be daunting to examine item-level data because of the sheer number of items and response options; therefore, it is useful to organize items by content area and examine these groupings. (Suggested item-content groupings can be found in each topic area’s [Discussion Guide](#) and instructions for calculating the average [mean] of the items can be found in Appendix C.) In this way, specific content areas within topics with particularly low favorability ratings may be more readily identified; schools may want to target these content areas for immediate interventions.

¹⁴ The EDSCLS parent/guardian survey includes some items that have a Don’t Know response. Analyzing those data can help districts and schools identify areas in need of effective communication with families.

For example, items in the Bullying/Cyberbullying topic area can be grouped by content areas such as bullying is a problem; bullying related to race, ethnicity, and religion; cyberbullying; bullying related to sexuality; bullying related to physical and mental disability; bullying prevention: student reporting and stopping of bullying; and bullying prevention: bullying knowledge and formal prevention strategies.

The average (mean) of the item rating (1–4) for a content group can be compared to examine:

- How favorably an item-content group is perceived by all respondent groups in the school, compared with other item-content groups in the same topic area.

Example: A school can see how favorably respondent groups in the school perceive the item-content group called “cyberbullying,” compared with the other bullying item-content groups.

- How favorably an item-content group is perceived by respondent groups in the school, compared with the district, if data are available.

Example: A school can see if its respondent groups perceive the cultural diversity of materials more or less favorably than do the district respondent groups.



Cautions/Limitations:

1. **For items per individual respondent group (students, instructional staff, noninstructional staff, and parents/guardians).**

- It is important to consider that items represent various unique aspects of the overall topic area that is being measured; each has been selected to reflect only one specific aspect of respondent perceptions.

Example: If examining Environment items suggests that bathroom cleanliness in this school is an area for improvement, this information might help contextualize perceptions of the environment in the school. However, although improving bathroom cleanliness might lead to more favorable responses about that particular problem, it may have less of an effect on the important underlying perception of the overall environment in the school.

- You can think of excessive focus on improving responses to a single item as analogous to a physician attempting to alleviate symptoms without curing the underlying disease or condition.
- Perhaps as important, when measuring any topic area, the indicators (items) should reflect a representative sample of behaviors and attitudes about the topic. If schools focus change solely on behaviors and attitudes as defined by specific items, they risk a form of unintentional “teaching to the test” in which schools are able to show growth with respect to specific items even though perceptions of the underlying topic area have not changed.

2. **For individual items that are worded *exactly the same* across respondent groups (called “comparable items”).**

- Item-level data cannot be compared across respondent groups or subgroups of interest unless they are worded exactly the same. (Also see cautions detailed above.)

3. **For items within a topic area grouped by content (“item-content group”).**

- In addition to the cautions and limitations described for item-level results above, data users should recognize that item groupings are intended solely as a way to organize a large amount of information and do not represent psychometrically validated scales.





Use These Results To:

1. For items per individual respondent group (students, instructional staff, noninstructional staff, and parents/guardians).

In general, if the rating on a specific item is less favorable than you think acceptable and is an important issue to your school, district, and community, you may consider this a specific area to target interventions.

- Look at items that have low favorability ratings compared to other items in the same domain, within specific respondent groups, to target interventions.

Example: If, on average, students rate the single item, “This school provides instructional materials (e.g., textbooks, handouts) that reflect my cultural background, ethnicity, and identity,” lower than other items in the CLC topic area, a school can, for instance, support its teaching staff to start including more culturally diverse examples in class materials and lessons.

- Look at items that have low favorability ratings compared with other items across domains, within specific respondent groups, to target interventions.

Example: If examining Safety domain items suggests that there is a problem with students being teased or picked on about race or ethnicity, the school may want to examine other items from other domains to see if related issues are being reported there. For instance, examining Engagement domain items may show that students report few examples of different racial, ethnic, or cultural backgrounds in instructional materials. Targeting a quick adjustment on the latter issue, such as instructing teachers to incorporate more culturally diverse examples in materials and lessons, may help to bring a sense of cultural sensitivity to the school.

2. For individual items that are worded exactly the same across respondent groups (called “comparable items”).

In general, if the rating on a comparable item is more unfavorable for a specific respondent group than for other respondent groups, or if it is more unfavorable for a respondent group in your school than that group in the district, and this issue is important to your school and community, you may consider this an area to specifically target interventions.

- Analyze percentage distributions on comparable items across respondent groups to investigate differences in perceptions of a potential problem or specific area.

Example: Look at the percentage distribution across response options for the student drug-use item (not a problem, small problem, somewhat a problem, large problem). If most students say that it is somewhat or a large problem, but instructional staff are saying it is not a problem or a small problem, you can infer that teachers are not seeing what the students are seeing in school. You can then begin targeting support for staff in regard to this issue.

- Compare averages (means) of comparable items across respondent groups to investigate differences in perceptions of a potential problem or specific area.

Example: Look at the reported average (mean) for the item “I feel safe at this school” for students and for instructional staff. The average (mean) that is reported for this comparable item will range from 1 (*strongly disagree*, or Most Unfavorable) to 4 (*strongly agree*, or Most Favorable). If the average (mean) of student reports is 3.5 and the average (mean) for instructional staff is 2, you can see that teachers view this more unfavorably than do students. Again, you can use this information to target supports concerning this issue to appropriate stakeholders in the school (in this case, teachers).

Introduction

Scale Scores

Item-Level Data

Conclusion

Appendix A:
Comparisons
for Districts

**Appendix B:
Comparisons
for Schools**

Appendix C:
Average (Mean)
Topic Area Values

Appendix D:
Glossary of Terms

Appendix E:
What Not to
Compare Using
Legacy Scale
Scores

- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts
- Appendix B: Comparisons for Schools**
- Appendix C: Average (Mean) Topic Area Values
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

3. For items within a topic area grouped by content (“item-content group”).

Looking at groups of items that pertain to similar substantive content is another way of looking at item-level data. This approach may be especially helpful when digging deeper into the data to see which areas might be ripe for interventions.

For each school climate topic area, we have consulted with content experts to arrive at helpful ways of grouping items. When you click on links to the [Discussion Guides](#), you will see these item-content groupings for each specific topic area.

- Using this approach, schools can see where they can immediately target interventions and where they might be able to wait to implement longer term interventions.

Example: In digging deeper into the CLC topic area data, it may become evident that the cultural diversity reflected in learning materials is wanting for students and staff. A quick adjustment can immediately be implemented, such as instructing teachers to incorporate more culturally diverse examples in materials and lessons. If, conversely, respondents report that the climate of respect and equality in the school is favorable, efforts to improve respect and equality may wait until more urgent CLC needs are taken care of.

Appendix C. Average (Mean) Topic Area Values

This appendix is intended for users of the EDSCLS as well as of other surveys who are interested in using average (mean) values for interpreting their survey results. This includes (1) EDSCLS users who want to calculate the average (mean) of item content groupings as presented in the topic area *Discussion Guides*; (2) those who use the EDSCLS instrument, but not the online platform, and wish to calculate average (mean) topic area values; and (3) users of other surveys who employ items with a 1–4 response option scale.

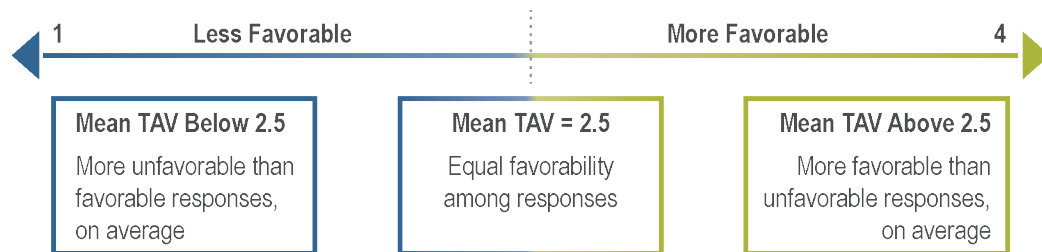
Average (Mean) Topic Area Values

Overall Average (Mean) Topic Area Values

Average (mean) topic area values can be used to gauge the degree to which respondents agree with the items that make up a content group or topic area (i.e., how favorably they perceive the topic area). (See Figure 1 for the topic areas in the EDSCLS.)

Assuming that all of the response options in the items to be included in the average (mean) calculation are on a scale of 1–4 (where 1 is less favorable and 4 is more favorable), you can think of the favorability of the item compared with the values as shown in Figure C-1.

Figure C-1 Item Favorability Interpretation.



Note: Topic area value is abbreviated TAV.

As illustrated in Figure C-1, an average (mean) topic area value of less than 2.5 would indicate that there are more unfavorable than favorable responses, on average, to the associated items, suggesting that this topic area is perceived relatively unfavorably. Conversely, an average (mean) value of more than 2.5 indicates more favorable than unfavorable responses, on average, suggesting that the topic area is perceived favorably. An average (mean) value of 2.5 indicates equal favorability among responses, suggesting that perception of the topic area is “neutral” (neither favorable nor unfavorable).

Average (Mean) Topic Area Values Broken Out by Respondent Characteristics

Average (mean) topic area values can be broken out by respondent subgroups to provide a deeper understanding of the data. For example, districts and schools can see how the favorability of a topic area differs across subgroups of students and staff.





For sites using the EDSCLS, average (mean) topic area values can be examined by the following respondent characteristics:¹⁵

- Student average (mean) topic area values by:
 - Gender,
 - Race/ethnicity, and
 - Grade.
- Staff (instructional and noninstructional) average (mean) topic area values by:
 - Gender and
 - Race/ethnicity.

Average (mean) topic area values by categories of respondent characteristics are not directly available from the EDSCLS platform but can be calculated from raw survey data. See the [EDSCLS User Guide](#) for information on exporting raw survey data.

Interpreting Average (Mean) Topic Area Values

Tables C-1 (for districts) and C-2 (for schools) provide suggestions for EDSCLS and other survey users for interpreting average (mean) topic area values overall and by respondent characteristics. They also include comparisons that can be made using average (mean) topic area values, how these comparisons can help you improve school climate, and cautions and limitations.



If you are a **district**, click [Appendix C, Table C-1](#) to go to interpretation of average (mean) topic area values overall and by respondent characteristics.



If you are a **school**, click [Appendix C, Table C-2](#) to go to interpretation of average (mean) topic area values overall and by respondent characteristics.

Instructions for Calculating Average (Mean) Topic Area Values or Item- Content Groups (Mean of Means)¹⁶

Averages (means) of groups of items, whether they are topic area values, item-content groups, or any other item groupings, can be calculated as the mean of the means of the associated items. For example, if a topic area or item-content group has four associated items, the mean would be calculated as follows:

Example Items	Item 1	Item 2	Item 3	Item 4
Item Mean	2.5	2.5	2.4	2.6

In this example, the averages (means) for items 1–4 are 2.5, 2.5, 2.4, and 2.6. The average (mean) of these averages (means) is calculated as the sum of the item average (mean) divided by the number of items:

$$\text{Average (mean) topic area value} = (2.5+2.5+2.4+2.6)/4 = 2.5$$

¹⁵ Note that the EDSCLS platform does not produce crossed demographics (e.g., Asian female students).

¹⁶ For more information on calculating mean (average) topic area values with EDSCLS data, please contact the EDSCLS Help Desk at edscls@air.org.

- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values**
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

Here, the average (mean) of the items 1–4 is 2.5. Before calculating an average (mean) value for a group of items, be sure to reverse code any negatively valenced items in the group, as described in the section [Reverse Coding Negatively Valenced Items](#).



For EDSCLS users, the reverse coding of negative items has already been done for you in the scale scores produced by the Web-based platform. However, you will have to reverse code negative items yourself if you want to calculate the average (mean) of item-content groups (see Table 1 of the [Discussion Guides](#) for your specific topic of interest) or any other groups of items.



Interpreting Average (Mean) Topic Area Values: Direction, Support, and Resources for Districts

Table C-1 provides direction, support, and resources to **districts** on what comparisons can be made with average (mean) topic area values overall and by respondent characteristics, cautions and limitations when interpreting these data, and how these comparisons can help schools and districts improve school climate.

Table C-1 Average (Mean) Topic Area Values: Comparisons, Cautions, and Uses.



You Can Compare:

- The favorability of a topic area (rated as 1–4) within one domain versus a topic area in another domain, as perceived by a single respondent group.

Example: You can compare your district's average (mean) topic area values on the Relationship topic area (within the Engagement domain) with the Mental Health topic area (within the Environment domain) to determine how favorably respondents perceive each of these topic areas.

- The favorability of a topic area (rated as 1–4) within a single domain as perceived by a respondent group versus the favorability perceived by another respondent group.

Example: You can compare your district's average (mean) topic area values on Emotional Safety and Physical Safety (both within the Safety domain) across students and staff.

- The favorability of a topic area (rated as 1–4) within one domain versus a topic area in another domain, as perceived by subgroups of a single respondent group.

Example: You can compare your district's average (mean) topic area value on the Relationship topic area (within the Engagement domain) with the Mental Health topic area (within the Environment domain) to determine how favorably female students versus male students perceive each area.

- The favorability of a topic area (rated as 1–4) within a single domain as perceived by a specific subgroup of respondents versus the favorability perceived by that same subgroup of a different respondent group.

Example: You can compare your district's average (mean) topic area value on the Emotional Safety versus Physical Safety topic areas (both within the Safety domain) to determine how favorably female students versus female instructional staff perceive each area.

- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values**
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores



Cautions and Limitations:

- These average (mean) item groupings will provide information about how positively or negatively respondents perceive a topic area or domain or item content group. However, even though topic areas may have similar averages (means), they still may not be “equal” (i.e., respondents may feel more or less favorably about various topics to start with or have more or less difficulty answering questions about them). Readers should therefore use caution when interpreting comparisons across respondent groups.
- Similarly, when data are disaggregated by subgroup, they will indicate how positively or negatively subgroups of respondents perceive the topic area or domain or item content group. However, even though topic areas may have similar averages (means), they still may not be “equal” (i.e., the importance of a topic or how intense the desire is to achieve excellence in that topic area may vary across respondent subgroups). Readers should therefore use caution when interpreting comparisons across respondent subgroups.



Use These Comparisons To:

- Focus immediate universal/districtwide or school-specific support pertaining to those topic areas perceived as the most unfavorable across all domains, for a specific respondent group.

Example: You can compare your district’s average (mean) values on each topic area in your student survey and prioritize supports based on the topic area(s) with the lowest average (mean) values.
- Focus immediate targeted support to a specific respondent group that perceives a topic area more unfavorably than do other respondents.

Example: If instructional staff report highly favorable perceptions of CLC in the district but students report less favorable perceptions, this difference could indicate that staff’s intended attention to cultural sensitivity and inclusion is not being executed in a way that speaks to students, and improvement could be focused on staff execution.
- Focus immediate universal/districtwide or school-specific support pertaining to topic areas perceived as the most unfavorable across all domains, for a specific subgroup of respondents.

Example: You can compare your district’s average (mean) topic area values for Hispanic students and prioritize supports based on the topic area(s) with the lowest average (mean) values.
- Focus immediate targeted support to a specific subgroup of respondents that perceives a topic area more unfavorably than does the same subgroup of other respondents.

Example: If Black or African-American teachers report highly favorable perceptions of CLC in the district, but Black or African-American students report less favorable perceptions, this difference could indicate that these teachers’ intended attention to cultural sensitivity and inclusion is not being executed in a way that speaks to these students, and improvement could be focused on staff execution.



Interpreting Average (Mean) Topic Area Values: Direction, Support, and Resources for Schools

Table C-2 provides direction, support, and resources for schools on what comparisons can be made with average (mean) topic area values overall and by respondent characteristics, cautions and limitations when interpreting these data, and how these comparisons can help schools to improve school climate.

Table C-2 Average (Mean) Topic Area Values: Comparisons, Cautions, and Uses.



You Can Compare:

- The favorability of a topic area (rated as 1–4) within one domain versus a topic area in another domain, as perceived by a single respondent group.

Example: You can compare your school's average (mean) value on the Relationship topic area (within the Engagement domain) with the Mental Health topic area (within the Environment domain) to determine how favorably respondents perceive each of these topic areas.

- The favorability of a topic area (rated as 1–4) within a single domain as perceived by a respondent group versus the favorability perceived by another respondent group.

Example: You can compare your school's average (mean) topic area values on Emotional Safety and Physical Safety (both within the Safety domain) across students and staff.

- The favorability of a topic area (rated as 1–4) within one domain versus a topic area in another domain, as perceived by subgroups of a single respondent group.

Example: You can compare your school's average (mean) topic area value on the Relationship topic area (within the Engagement domain) with the Mental Health topic area (within the Environment domain) to determine how favorably female students versus male students perceive each area.

- The favorability of a topic area (rated as 1–4) within a single domain as perceived by a specific subgroup of respondents versus the favorability perceived by that same subgroup of a different respondent group.

Example: You can compare your school's average (mean) topic area value on the Emotional Safety versus Physical Safety topic areas (both within the Safety domain) to determine how favorably female students versus female instructional staff perceive each area.



Cautions and Limitations:

- These average (mean) values will provide information about how positively or negatively respondents perceive a topic area or domain or item content group. However, even though topic areas may have similar averages (means), they still may not be "equal" (i.e., respondents may feel more or less favorably about various topics to start with or have more or less difficulty answering questions about them). Readers should therefore use caution when interpreting comparisons across respondent groups.
- Similarly, when data are disaggregated by respondent subgroup, they will indicate how positively or negatively those subgroups perceive a topic area or domain or content group. However, even though topic

Introduction

Scale Scores

Item-Level Data

Conclusion

Appendix A:
Comparisons
for Districts

Appendix B:
Comparisons
for Schools

**Appendix C:
Average (Mean)
Topic Area Values**

Appendix D:
Glossary of Terms

Appendix E:
What Not to
Compare Using
Legacy Scale
Scores

- Introduction
- Scale Scores
- Item-Level Data
- Conclusion
- Appendix A: Comparisons for Districts
- Appendix B: Comparisons for Schools
- Appendix C: Average (Mean) Topic Area Values**
- Appendix D: Glossary of Terms
- Appendix E: What Not to Compare Using Legacy Scale Scores

areas have similar averages (means), they still may not be “equal” (i.e., the importance of a topic or how intense the desire to achieve excellence in that topic area may vary across respondent subgroups). Readers should therefore use caution when interpreting comparisons across respondent subgroups.



Use These Comparisons To:

- Focus immediate universal/schoolwide support pertaining to those topic areas perceived as the most unfavorable across all domains, for a specific respondent group.

Example: You can compare your school’s average (mean) values on each topic area in your student survey and prioritize supports based on the topic area(s) with the lowest average (mean) values.

- Focus immediate targeted support to a specific respondent group that perceives a topic area more unfavorably than do other respondents.

Example: If instructional staff report highly favorable perceptions of CLC in the school building but students report less favorable perceptions, this difference could indicate that staff’s intended attention to cultural sensitivity and inclusion is not being executed in a way that speaks to students, and improvement could be focused on staff execution.

- Focus immediate universal/schoolwide support pertaining to topic areas perceived as the most unfavorable across all domains, for a specific subgroup of respondents.

Example: You can compare your school’s average (mean) topic area values for Hispanic students and prioritize supports based on the topic area(s) with the lowest average (mean) values.

- Focus immediate targeted support to a specific subgroup of respondents that perceives a topic area more unfavorably than does the same subgroup of other respondents.

Example: If Black or African-American teachers report highly favorable perceptions of CLC in the school but Black or African-American students report less favorable perceptions, this difference could indicate that these teachers’ intended attention to cultural sensitivity and inclusion is not being executed in a way that speaks to these students, and improvement could be focused on staff execution.



Appendix D. Glossary of Terms

Domain: Refers to a major overarching school climate scale, constructed psychometrically, often containing subdomains. The ED School Climate Surveys model of school climate contains three domains—Engagement, Safety, and Environment.

Mean: Refers to the average value of a group of scores. Means are calculated by adding all of the scores and dividing by the number of scores. For example, the mean of three scores (2, 3, and 4) would be calculated as: *Mean:* $(2 + 3 + 4) / 3 = 3$.

Mean topic area value: Refers to the average value of a group of items that comprise a topic area. Mean topic area values are presented on a scale of 1–4 to reflect the response options in the EDSCLS (see [Appendix C](#) for more information).

Scale score: Refers to a numeric measurement that represents a survey respondent’s answers to a group of related survey questions. Psychometric analysis is often used to produce scale scores. One common type of scale score is called a “theta” value, which ranges from –3 to +3 but is often transformed. For example, scale scores for the EDSCLS are derived from theta values but are transformed to a scale of 100–500.

Benchmarked Scale score: Starting with the EDSCLS VM version 3.0, scale score results are benchmarked into three “performance” levels.

- Level 1: Least Favorable (scale scores below 300);
- Level 2: Favorable (scale scores 300–400), and;
- Level 3: Most Favorable (scale scores above 400–500).

The platform produces graphs showing scale scores embedded into the three “performance” levels. The benchmarking allows users to compare the “performance” levels across three EDSCLS domains.

Subdomain: Refers to a concept of school climate that is part of a larger domain. For example, in the EDSCLS model of school climate, there are three subdomains within the Engagement domain. These subdomains are Cultural and Linguistic Competence, Relationships, and Participation. Subdomains also are referred to topics or topic areas in the EDSCLS.

Topic area: Refers to a concept of school climate that is part of a larger domain. For example, in the EDSCLS model of school climate, there are three topics or topic areas within the Engagement domain. These topics are Cultural and Linguistic Competence, Relationships, and Participation. Topics also are referred to as subdomains.

Valenced: Refers to the intrinsic direction of a term or phrase. For example, “I feel safe” is considered positively valenced because agreeing with the statement indicates good or positive perceptions. In contrast, “I don’t feel safe” is considered a negatively valenced phrase because agreeing with the statement indicates bad or negative perceptions.

Introduction

Scale Scores

Item-Level Data

Conclusion

Appendix A:
Comparisons
for Districts

Appendix B:
Comparisons
for Schools

Appendix C:
Average (Mean)
Topic Area Values

**Appendix D:
Glossary of Terms**

Appendix E:
What Not to
Compare Using
Legacy Scale
Scores

Appendix E. What Not to Compare Using Legacy Scale Scores

Legacy scale scores are scale scores that were produced by EDSCLS platforms prior to VM 3.0, which was released in December 2017. Legacy scale scores should not be compared to the scale scores produced by EDSCLS VM 3.0 and higher.

Legacy scale scores should not be used to make the following comparisons:

- Scale scores for domains/topic areas across respondent groups.

Example: Scores on CLC for students should not be compared with CLC scores for staff.

- Scale scores for respondents on topic areas across domains.

Example: Scores for students on the Physical Safety topic area (within the Safety domain) should not be compared with the student score for the Relationship topic area (within the Engagement domain).

- Scale scores by respondent characteristics for domains/topic areas for subgroups within different sets of respondents.

Example: Scores on CLC for White Students should not be compared with CLC scores for White staff.

- Scale scores by respondent characteristics for topic areas across domains.

Example: Scores for Hispanic noninstructional staff on the Physical Safety topic area (within the Safety domain) should not be compared with Hispanic noninstructional staff's score for the Relationships topic area (within the Engagement domain).

Introduction

Scale Scores

Item-Level Data

Conclusion

Appendix A:
Comparisons
for Districts

Appendix B:
Comparisons
for Schools

Appendix C:
Average (Mean)
Topic Area Values

Appendix D:
Glossary of Terms

**Appendix E:
What Not to
Compare Using
Legacy Scale
Scores**

