

# DATA QUALITY: STRATEGIES TO ENSURE DISTRICT DATA ARE ACCURATE

Data quality—the accuracy of your school district’s data—is one of the most important factors supporting the success of any data initiative. The lower the data quality, the higher the risk for eroded stakeholder trust, delayed timelines, or worse yet, erroneous conclusions based on false information.

Yet detecting and addressing errors can be deceptively hard. With thousands of students and hundreds of thousands of data points, where do you start? Here’s what we at Schoolzilla have learned from the dozens of school systems we’ve worked with as they improve and maintain their data quality.

## Cleanup

Improving data quality can feel daunting without a plan. How do you prioritize what to tackle first? Although it depends on your district and the data that’s most important to you, we generally suggest the following:

### Prioritize these types of data in this order:

1. Unique identifiers like student IDs: If these are amiss, any related data point could have gone awry.
2. The data that impact your students the most: Are standards-aligned interim assessments the holy grail at your district? Is reducing chronic absence a big push this year? Start there.
3. Qualifiers: Do grade-level teams need to analyze those interim assessment results? Make sure teacher and grade-level data are correct. Are you concerned about equity when it comes to attendance? Review the quality of each student subgroup carefully.
4. Everything else: When you’ve worked through the data that fall into those categories, then you can move on to the rest with greater peace of mind.

For a checklist of specific data points districts tend to find most useful to check against, see the Audit Checklist included below.

### Brace yourself for what’s likely to be messy.

- ▶ Although not always the case, attendance and enrollment data tend to be fairly clean because districts tend to monitor them closely for compliance purposes. Instead give grade, formative assessment, and behavior (especially minor behavior incidents) data a second look. Because they’re typically entered by many different individuals on flexible platforms, the data can turn out more varied.
- ▶ Unless they are also your state test, the data for college-going assessments like the ACT and SAT, is often entered by students themselves, which can make matching them up with other data even harder. What’s more, these results sometimes come in fixed-width formats and are difficult to edit manually. Set aside extra time for cleaning these data sets, and you won’t be caught by surprise.

### When all else fails, check the dates.

- ▶ School and program enrollment and exit date are likely suspects when you’re stumped during troubleshooting, especially when student counts aren’t what you expect. Check that dates fall within the school year start and end dates and that students’ program enrollment dates fall within their school enrollment dates.

## Maintenance

While cleanups may not be entirely preventable, setting up systems—both technology- and people-based—that help keep data from getting messy quickly will help maintain a reasonable level of accuracy longer. Most of these tips also have the added benefit of saving time, so they are well worth the initial effort.

### Set-up automated systems.

- ▶ Whenever possible, choose a single source of truth so data entry has to be done only once. That might mean using a service like Clever for roster-syncing or an SSO provider like Okta, Active Directory, OneLogin, or Google.
- ▶ Choose live integrations over static ones. Regular data refreshes (whether via API, ODBC, FTP, or some other form) prevent data sets frozen in time from creating version control issues.
- ▶ If you're choosing systems that ingest data from other systems, pick products that match records based on multiple data points. For example, if your data warehouse can match SIS data and assessment data on either a student ID or a state ID, then you don't have to worry as much about missing student IDs in your assessment platform. Although you may want to go back and correct these types of errors, this kind of logic means more accurate data in the meantime.

### Cultivate data neatniks.

- ▶ Align on the reporting level needed before creating data entry norms. Doing so will help you straddle the tension between educator autonomy and organization-wide reporting that often results in either poor data quality or a lack of investment in data entry procedures.
- ▶ Do the people managing data entry utilize those data too? If not, consider how data could help them in their day-to-day work. With a clear understanding of what the data can offer and the consequences of poor data quality, you're likely to see higher engagement in proper data entry.
- ▶ Hire for attention to detail, experience with databases, and a strong sense of ownership.
- ▶ Develop high-quality trainings and documentation, especially if you have high turnover in data entry positions. Be sure to make troubleshooting tips and helpline supports clear alongside data entry instructions.

For a conversation about how Schoolzilla's data quality audit process – a part of implementing our multiple measure dashboard system, [Mosaic District Progress Monitoring](#) – can help you clean up your district's data, save time, and build trust with your stakeholders, contact us at [sales@schoolzilla.com](mailto:sales@schoolzilla.com).

## Audit Checklist

You'll likely have additional metrics to add, but here's our basic checklist when we conduct an audit with a new school partner:

### Student Demographics

- Percent of students missing a value for EL Status
- Percent of students missing a value for Is SPED
- Percent of students missing a value for Foster Care Status
- Percent of students missing a value for FRL Status
- Percent of students missing a value for Gender
- Percent of students missing a value for Grade Level
- Percent of students missing a value for Homeless Status
- Percent of students missing a value for Migrant Status
- Percent of students missing a value for Military Family Status
- Percent of students missing a value for Primary Ethnicity
- Percent of students missing a value for SPED Status

### Enrollment

- Number of students with entry and exit dates outside the first and last day of the school year

### Attendance

- Percent of students missing an attendance value for at least one day (this year)
- Percent of students missing an attendance value for at least one day (last year)
- Number of missing FTE IDs
- Number of entry and/or exit dates outside school term windows
- Number of mismatched attendance conversion items
- Number of FTE ID mismatch between student and school
- Number of FTE ID mismatch between student and school year

### Behavior

- Percent of students with behavior records (this year)
- Percent of students with behavior records (last year)
- Percent of records with Behavior Response Name populated

### Grades

- Percent of high-school students with grades records (current year)
- Percent of high-school students with grades records (last year)
- Percent of middle-school students with grades records (current year)
- Percent of middle-school students with grades records (last year)
- Percent of elementary-school students with grades records (current year)
- Percent of elementary-school students with grades records (last year)
- Percent of records with Course Subject populated