PARE
Practical Assessment, Research, and Evaluation
A peer reviewed, open-access electronic journal: ISSN 1531-7714

Applied Shewhart Charts in Education to Understand Chronic Absenteeism Across the State of California: A Methods Application

Erica Geary, *High Tech High Graduate School of Education*
Lloyd Provost, *Associates for Process Improvement*
Brandon Bennett, *Improvement Collective*

Abstract: Chronic absenteeism rates in the State of California doubled from 14.3% to 30% during the COVID-19 Pandemic. This paper introduces the use of Shewhart charts to identify schools exhibiting special cause (good) variation in chronic absenteeism rates across California that can provide insight into practices and policies that can improve attendance. First, an analysis of difference was conducted - a form of time series analysis allowing the investigators to detect unusually (special cause) good/rapid improvement in chronic absenteeism rates by a school from the pandemic high (using I or Individuals charts to detect special causes or bright spots). Second, a cohort analysis was conducted - analyzing the most recent absenteeism data, 2023-2024, in schools across counties (using P charts to detect bright spots). Both datasets were joined, identifying 36 "super" bright spots or schools with both rapid improvement from 2022 to 2024 and lower chronic absenteeism in 2024. A subset of these schools was interviewed to discover common practices. The RAISE Network provides an initial package of change ideas (evidence-based practices) to schools and districts that can be used to inform policy and practice in regards to improving chronic absenteeism.

Keywords: Shewhart charts, Control charts, Chronic absenteeism, Bright spots

Introduction

Important educational outcomes like student achievement, graduation rates or chronic absenteeism are characterized by a wide degree of variation (Brunner et al., 2018; Rahman et al., 2023; Economics of Education Review, 2023). For those interested in improving these outcomes, identifying positive outliers (often called 'bright spots') and understanding what might explain their relative advantage can reveal key insights to help others. The concept of 'Positive Deviance' is a powerful framework that has been used in healthcare improvement (Low et al., 2025). However, there are few pragmatic methodological approaches for identifying bright spots and learning from them. The approach described in this paper is an accessible method for those applying the concept of positive deviance to educational outcomes, offering a framework that informs decision-making for educators, researchers, administrators and policymakers.

The primary purpose of this study is to demonstrate the application of Shewhart charts as an analytical and visual tool to learn from variation in any important outcome. We use a case study illustrating its application to chronic absenteeism rates. Specifically, the study positions Shewhart charts as both a diagnostic and communication tool for identifying bright spots that can then be investigated to learn from and spread best practices.

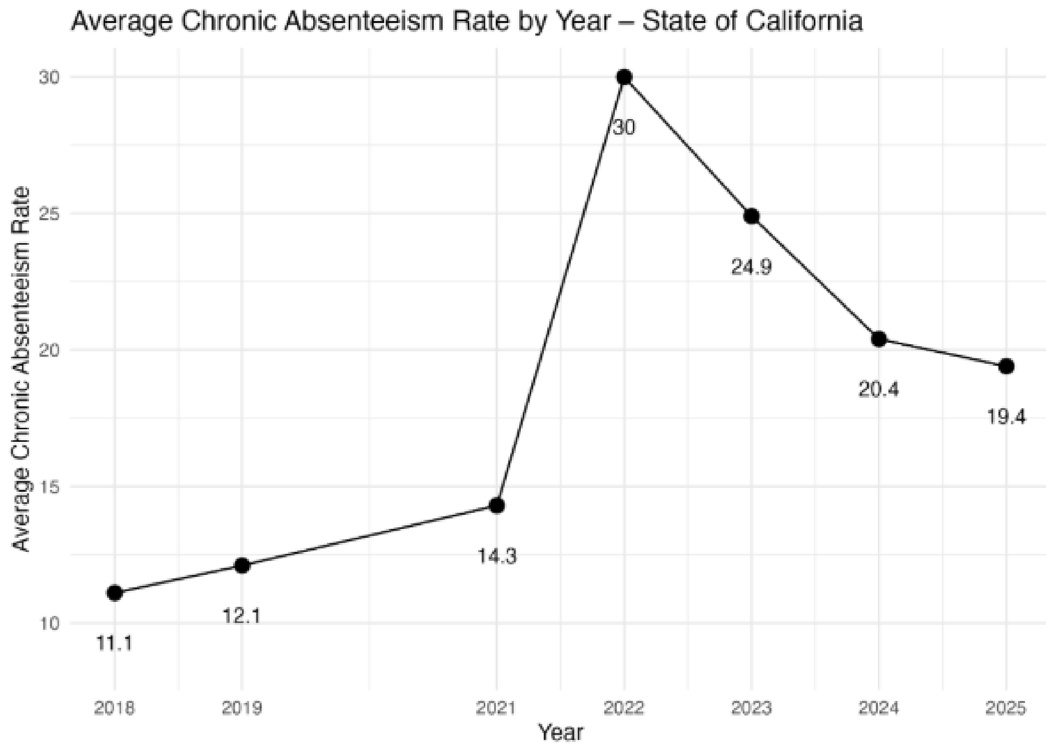
Overview of Shewhart Charts for Understanding Variation

Shewhart charts have been utilized in manufacturing, production and social sector service industries, like healthcare, for decades, but remain an underutilized method for understanding variation in the field of education (Omar, 2010). These charts were originally designed in order to monitor measures of a system to detect and react to special causes (Shewhart, 1926). The use of Shewhart charts has evolved as a method allowing practitioners to both understand their current system and develop a focused strategy to improve it. These charts include a centerline and upper and lower three-sigma limits, with statistical rationale supported by empirical evidence (Provost & Murray, 2022, p. 132-134). Shewhart charts are designed to distinguish between 'common cause' and 'special cause' variation. Common cause variation reflects the inherent variation produced by the design of the process or system when it is stable, while special cause variation signals external influences outside of the system design that warrant investigation. Such analysis is valuable for a number of reasons: assessing the stability of a system, identifying and learning from special causes, diagnosing where or when a system is changing, or, when used in a time series fashion, for understanding when a process or system has changed its performance in some meaningful way. In educational contexts, these charts can reveal when certain schools or districts deviate meaningfully from expected absenteeism, guiding deeper qualitative inquiry. In this study, Shewhart control limits were calculated using the standard three-sigma. A special cause was defined as any data point beyond the upper or lower three-sigma limits (Provost & Murray, 2022). Special causes below the lower control limit indicated bright spots for current performance—schools outperforming expectations in regards to lower rates of chronic absenteeism in 2024. Special causes above the upper control limit indicated bright spots for the change in chronic absenteeism–schools outperforming expectations in the improvement in rates of chronic absenteeism from the 2022 to 2024 school year. Focusing on special cause variation allows for appropriate actions to be taken instead of random actions that may not be productive or could actually increase variability (Deming, 1951).

Chronic Absenteeism as a Case Study

The case study in this paper applies the Shewhart chart method to the timely problem of chronic absenteeism in California. Chronic absenteeism is a key concern in education with negative consequences to students, families and schools. It is defined as the percentage of students absent for 10% or more of the instructional days they were enrolled to attend (California Department of Education, 2025). This rate doubled from 14.3% in 2021 to 30% in 2022 during the pandemic in the State of California as seen in Figure 1. Students who are chronically absent have lower academic achievement (Ansari & Pianta, 2019; García & Weiss, 2018; Gottfried, 2014; Romero & Lee, 2007; Santibañez & Guarino, 2020). In addition, these students face a higher rate of high school drop-out (Alexander et al., 1997; Rumberger, 2015; Rumberger & Larson, 1998). The adverse effects are not just academic. These students also have lowered social skills and increased internalizing problem behaviors (Gottfried, 2014). Therefore, it is paramount to learn how to reduce absenteeism rapidly and return students to school. In the Raising Attendance and Improving Student Engagement Network (RAISE), Shewhart charts were applied in an effort to provide insight into schools or locations that can help achieve this goal (High Tech High Graduate School of Education National Coalition for Improvement in Education, 2025). RAISE is a California networked improvement community created to reduce chronic absenteeism (Bryk, et, al., 2015). Its statewide improvement approach unites students, families, educators, researchers, policymakers, and nonprofits to co-design solutions that foster belonging and engagement in schools. In five years' time the network aims to halve the rate of chronic absenteeism across the state, reducing from the 2023-2024 rate of 20.4% to 10.2% (California Department of Education, 2025). If successful, this equates to 656,788 students who will no longer be chronically absent by the end of the 2028-2029 school year.

Figure 1. Chronic Absenteeism in California has Doubled Since the Pandemic



RAISE applies a multi-year spread strategy to engage students across California. Year one began in May 2025 with 75 schools. If successful, the plan is to spread to 250 schools in year two, 1,000 in year three and around 5,000 California public schools in year four. Early year success is projected to depend on the adoption of foundational practices and innovation by subgroup-specific nodes (High Tech High Graduate School of Education National Coalition for Improvement in Education, 2025). Multiple modalities for discovery were used to produce a quality foundational change package (Bennett, 2020 and Grunow, Park and Bennett, 2024). Literature reviews, interviews with key subject matter experts and bright spot analysis across the state were undertaken. Literature reviews provided ample evidence on the consequences of absenteeism and some direction on specific interventions, but there was not ample research-based evidence to support many of the interventions discussed and the literature was not precise on where schools should initially focus their efforts. Therefore, there was a need to supplement the evidence present in the literature with a different type of analysis aimed at adding practice-based evidence directly from schools and districts. The challenge, addressed through the method we describe in this paper, was in identifying schools that were generating such practice-based evidence. Shewhart charts were used to generate a list of schools across the state that were performing better than what would be expected given the system (county) they found themselves in. These schools were worth learning from because they were likely engaged in interventions beyond the design of their district and county systems and engaged in practices that could be replicated by others. If so, these practices would be important contributions to include in a change package for practice-driven improvements. This paper focuses on the methods and findings of this Shewhart analysis using the 2023-2024 data that was publicly available at the time. School sites identified in this way are referred to as bright spots because their performance is unusually good (i.e., demonstrating special causes of variation in the desired direction), when compared with other schools in their same system.

Literature Review

Multiple types of Shewhart charts have been utilized across a variety of fields to identify special causes of variation present in a system (Provost & Murray, 2022). There are few published examples of the application of Shewhart charts in education and the existing examples mainly focus on CUSUM (Cumulative Sum) charts and some Xbar and S charts (Meijer, 2002; Omar, 2010). This paper extends the body of literature with the application of Shewhart P charts, P-prime (p') charts and Individual (i.e., XmR) charts in educational settings.

Rational subgrouping of data (dividing data into subgroups based on some similar characteristic) has been shown to be important for uncovering special causes within a system (Sefik, 1998). Often data may appear to be all common cause when in fact special causes exist but cannot be identified until the appropriate subgrouping has been performed (Nelson, 1988). This paper expands this body of work with the rational subgrouping of schools in the State of California by county and district.

The limited research in the education field often focuses on the application of Shewhart charts for identifying special cause results of a completed course or assessment (Wan & Keller, 2023). Special causes reveal the instability of the existing system and can be used to improve it (Bakir, 2005). This paper adds to the existing literature by applying Shewhart charts as a means of identifying special causes to learn from variation in an outcome measure (e.g. chronic absenteeism) that can then lead to broader system-wide improvement of that outcome. The scope of this project also adds to the existing literature by offering a novel approach to extract actionable information from a large dataset (e.g. a statewide dataset like that maintained by the State of California). Extant research often focuses on classroom level measures or smaller scale measures (Bi, 2020).

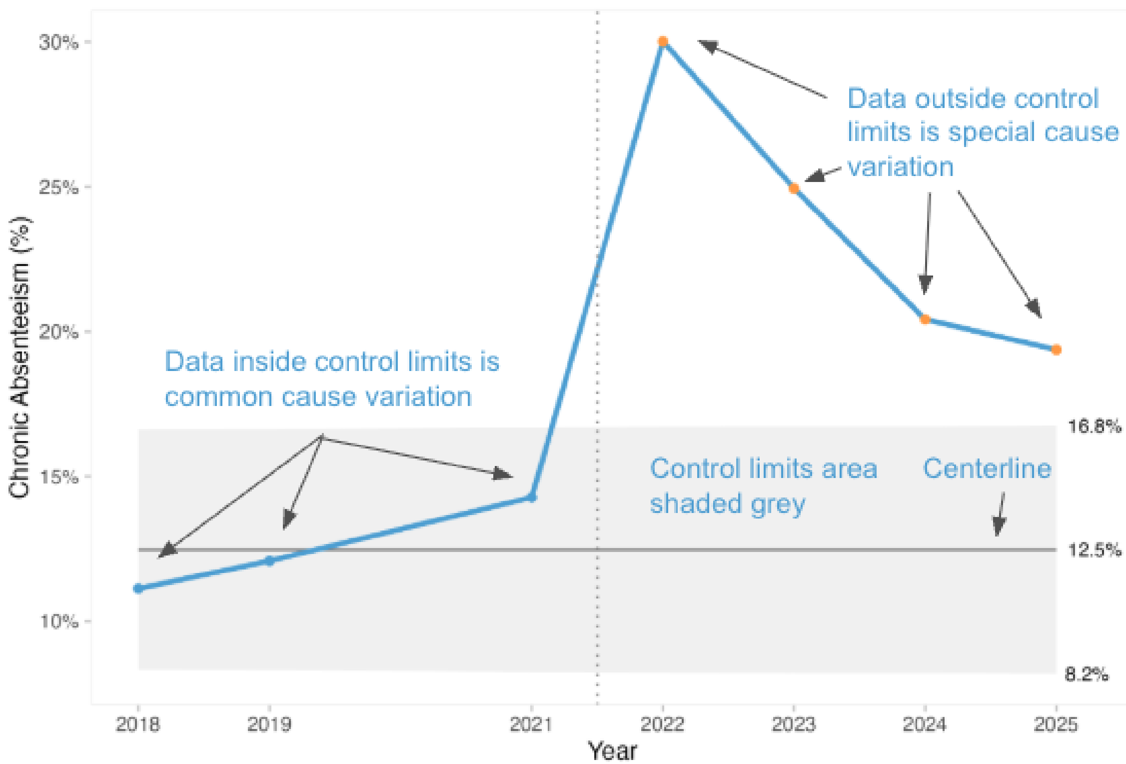
The approach we used is similar to the concept of Positive Deviance, focusing on exceptionally good results within a given system (McKay, 2017). Positive Deviance has been demonstrated to be effective in healthcare settings for identifying specific changes that can be tested and spread to improve systems. The approach follows four steps: identify positive deviants, study positive deviants, test hypotheses, and spread best practice (Low et al., 2025). Other examples in the field of continuous improvement have also shown this method to be beneficial to learn from what is working (Ming & Kennedy, 2020; LeMahieu, Bryk, et al., 2017; LeMahieu, Nordstrum et al., 2017; D. R. Marsh et al., 2004). This paper expands upon the Positive Deviance approach by using data adjusted for income to remove the effect of income on the data and by identifying positive deviants (bright spots) with Shewhart charts using both current performance and recent improvement over time. The joining of special causes from an analysis of difference over a specified time period, combined with the special cause results of the current performance of a measure, can add additional insight and pinpoint even more productive sources of useful ideas.

Method

Overview of Analytical Approach

A key aspect of all systems is that they display variation in outcomes - across locations, and over time and across subgroups of the data. Some of this variation is inherent to the design of the system, meaning the causes of variation are shared by all subgroups, at all locations and at all time periods for the system (common cause variation). Some observable variation is not inherent to the system and only manifests for some subgroups, in some locations and during some time periods (special cause variation). This variation is not causally explained by the design of the overall system, rather it has an assignable cause that is specific to that subgroup, location or time period where it is observed (Grunow, Park, & Bennett, 2024). In the context of chronic absenteeism in K-12 education in the State of California, this reality has profound implications for understanding the performance of the educational system over time and across location (counties, districts and most importantly, schools). We utilize Shewhart analysis to define locations where special causes of variation present themselves in an effort to discover schools and districts where such causes produce unusually good performance in chronic absenteeism (better than can be explained by the system of education in California). In finding these locations, direction was provided to the RAISE Network researchers on where to accomplish in-depth qualitative analysis, including interviews and site visits which could uncover practices unique to those locations. These practices were then incorporated into a replicable change package that other schools could adopt. Figure 2 illustrates how a Shewhart chart displays special and common cause variation.

Figure 2. Chronic Absenteeism – State of California (P' Chart) 2018-2025



Contextual Observations Guiding the Analysis

There are a few key observations that impact our approach to analysis. First, schools in California saw a special cause increase in chronic absenteeism at the start of the COVID-19 Pandemic, as seen in Figure 2. Second, on average, schools in the State of California have seen a three-year trend of improvement in chronic absenteeism but that trend seems to have leveled off at 19.4% in the latest year for which data is available (2024-2025), indicating a plateau. Absenteeism data used in this study are publicly accessible (California Department of Education, 2025). This plateau, along with the time needed for systemic adoption of practices across diverse schools, further supports the need for a network aimed at halving the chronic absenteeism rate in five years.

Research Questions

Two research questions guided this study: (1) Which schools in each county experienced rapid reductions in chronic absenteeism from 2021-2022 to 2023-2024 school years? (2) After accounting for socioeconomic status, within each county, which schools had lower vs. higher rates of chronic absenteeism in the 2023-2024 school year? We downloaded publicly available data from the California Department of Education from the 2017-2018 school year through the 2023-2024 school year and used R software and packages for the analyses (R Core Team, 2024). Data was filtered to exclude Dashboard Alternative Status Schools (DASS) as these schools may have other characteristics not shared by most schools. The pandemic year of 2019-2020 was omitted from the dataset due to incomplete data.

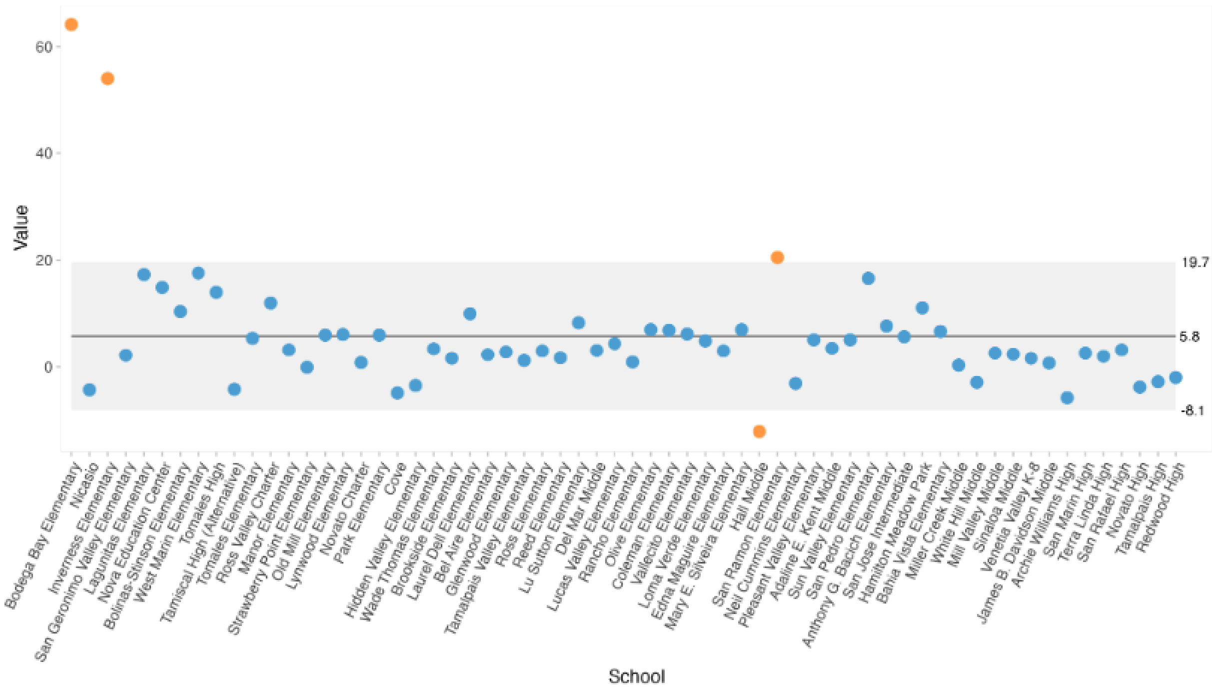
Step 1: Measuring Improvement Using I Charts

To investigate our first research question, we created a measure of difference from a previous year aimed at identifying schools in the state who were improving rapidly on chronic absenteeism. The calculation of difference to assess the rate of recovery was created by subtracting the current known rate for the 23-24 school year from the pandemic year of 21-22, creating an estimate of how quickly schools are recovering from dramatic increases in chronic absenteeism. These differences for each school were then plotted in each of the 58 counties in California on a Shewhart I chart to discover which schools within each county were improving faster than can be explained by the overall improvement taking place across all locations. An I chart was selected due to the nature of the dataset. Only one value was available per school because the value was a calculation of difference between two years. I charts display variation in single, continuous measurements over time or across units. They are uniquely suited to detecting unusually large shifts—either improvements or declines—when only one value per entity is available. For example, a hospital might use an I chart to examine month-to-month changes in average emergency-department wait times; an unusually sharp decrease or increase would signal a potential special cause (Provost & Murray, 2022). In the present study, I charts were used to examine absolute differences in chronic absenteeism between 2021–22 and 2023–24.

For the Individuals (I) charts, also referred to as XmR charts, the centerline (Xbar) was calculated as the mean of all individual observations—in this case, the average difference in chronic absenteeism rates between the 2021–22 and 2023–24 school years. The upper and lower control limits were derived using the standard three-sigma ($\pm 3\sigma$) convention for individual data points, following the formulas: UCL = Xbar + 2.66·MRbar; LCL = Xbar − 2.66·MRbar where MRbar represents the average of the moving ranges between consecutive data points. The constant 2.66 is a statistical multiplier based on the relationship between the average moving range and the standard deviation for individual measurements. This formulation allows for estimation of variability without subgrouping, making I charts particularly useful when only single data values are available per unit of analysis—as in the calculation of differences in absenteeism rates. Points that fall outside the upper or lower control limits are considered special-cause variation, suggesting that a school's improvement or decline exceeds what would be expected due to normal system variation (Provost & Murray, 2022). Points that fall above the upper control limit are bright spots in this case because they show a larger than expected improvement from 2022 to 2024.

These I charts produced 203 special cause schools from the 8,880 schools in the database. 93 special cause schools were above the upper control limit, identifying them as bright spots. These bright spots were then targeted for deeper analysis and learning. As an example, Figure 3 below shows four special cause schools in Marin County, three of which emerge as bright spots, and one school as a 'dim spot' (or point below the lower control limit), indicating that it improved more slowly than would be expected given the system studied.

Figure 3. Delta of Chronic Absenteeism by School in Marin County (2021-22 to 2023-24)



Step 2: Identifying County Current Performance Using P-prime (p') Charts

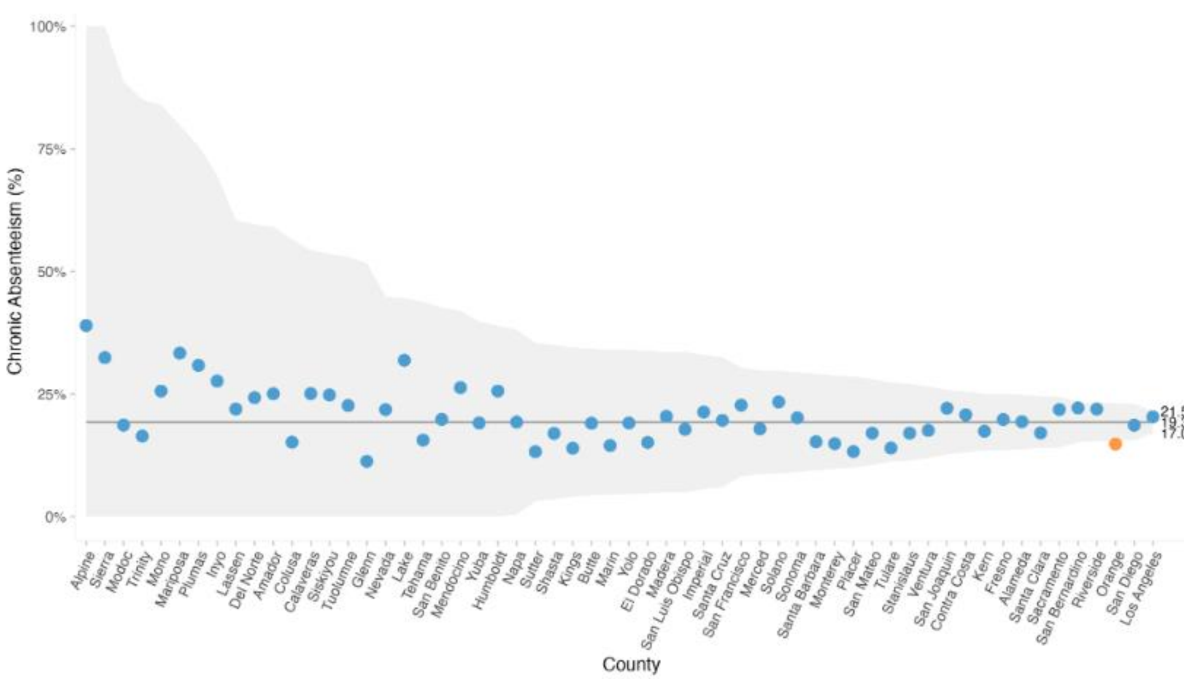
With the first research question answered by the analysis of difference and the I chart, P-prime (p') charts were then applied to investigate who and where in California was particularly good at minimizing chronic absenteeism in the 23-24 school year to answer the second research question. The P and P-prime (p') charts are appropriate for classification data (chronically absent vs not chronically absent). When investigating a system, it is often helpful to zoom out to the big picture and investigate one level of the system at a time. Research from the field has shown most variation in chronic absenteeism rates to be at the student level, rather than the class or school level (Malkus, 2024). Moving from a broad to narrow lens enabled us to see at what level special cause variation was occurring, starting with counties in California, then districts within counties and finally schools within counties. Schools within counties were selected rather than schools within districts to allow for sufficient data to be used in the Shewhart charts (Sefik, 1998). Consistent with prior research, more variation in the data was observed as we moved from the larger to the smaller levels.

The data was adjusted for income status via regression before proceeding to produce a Shewhart chart by county. Research in the field shows that income has an effect on chronic absenteeism and there are large differences in income status across California counties. Adjustments were not done for other factors such as sex and race. Sex was not used to adjust absenteeism rates because there were no notable differences between males and females across contexts studied (Black & Elgaddal, 2024). In addition, other demographic factors were not included in the adjustment since income was used as an umbrella to cover for most of the possible variables. Within the context of the RAISE network learning nodes are planned to solve for specific subgroups of student populations. At the state level with each county represented, there is not a strong correlation between the percentage of low-income students and absenteeism rates in our database (R^2 = 0.02). Stronger correlations were not found until the district within county regressions (e.g., R^2 = 0.37 in San Diego County) and school within county regressions (e.g., R^2 = 0.29 in Los Angeles County), though R^2 values varied region to region.

Counties were plotted on a P-prime (p') chart to identify variation at the county level. The selection of a P-prime (p') chart rather than a P chart was due to the size of the denominator for each subgroup plotted, such as statewide counts, entire hospital systems, or in this case, the number of students in each county. When plotting a Shewhart chart for classification data and the average denominator size per subgroup exceeds 3,000, a P-prime (p') chart is usually the best choice (Provost & Murray, 2022). In this case, P-prime (p') charts were used to compare county-and district-level absenteeism rates, where population sizes were large enough that a standard P chart would exaggerate signals of special-cause variation.

For the P-prime (p') charts, the centerline was established as the overall average proportion (Pbar), representing the mean absenteeism rate across all subgroups. The upper and lower control limits were calculated following the binomial three-sigma (±3σ) convention to define the range of expected common-cause variation. Specifically, for each subgroup $_i$, the limits were determined using the formulas: UCL_i = Pbar + $3 \cdot \sigma_{pi} \cdot \sigma_{zi}$, LCL_i = Pbar − $3 \cdot \sigma_{pi} \cdot \sigma_{zi}$ where σ_p is the standard deviation of the subgroup's proportion and σ_{zi} is the standardized normal deviate. Values outside these limits are interpreted as special cause variation—indicating that a county or district's absenteeism rate differs significantly from what would be expected under normal system variation (Provost & Murray, 2022). The P-prime (p') charts for rates of chronic absenteeism by counties in the state for the 2023-2024 school year produced mostly common cause variation with the exception of one special cause bright spot, Orange County. In this case, P-prime (p') charts were compared for adjusted and unadjusted data, and both revealed one special cause. The adjusted chart can be seen below in Figure 4. The shaded region represents all data falling within normal variation. Points outside of the shaded region represent special cause variation. The x axis represents counties in California and is ordered from smallest to largest. The y axis represents the percent of students who were chronically absent in 2023-2024.

Figure 4. Chronic Absenteeism Adjusted for Income by County in California 2023-2024 Year

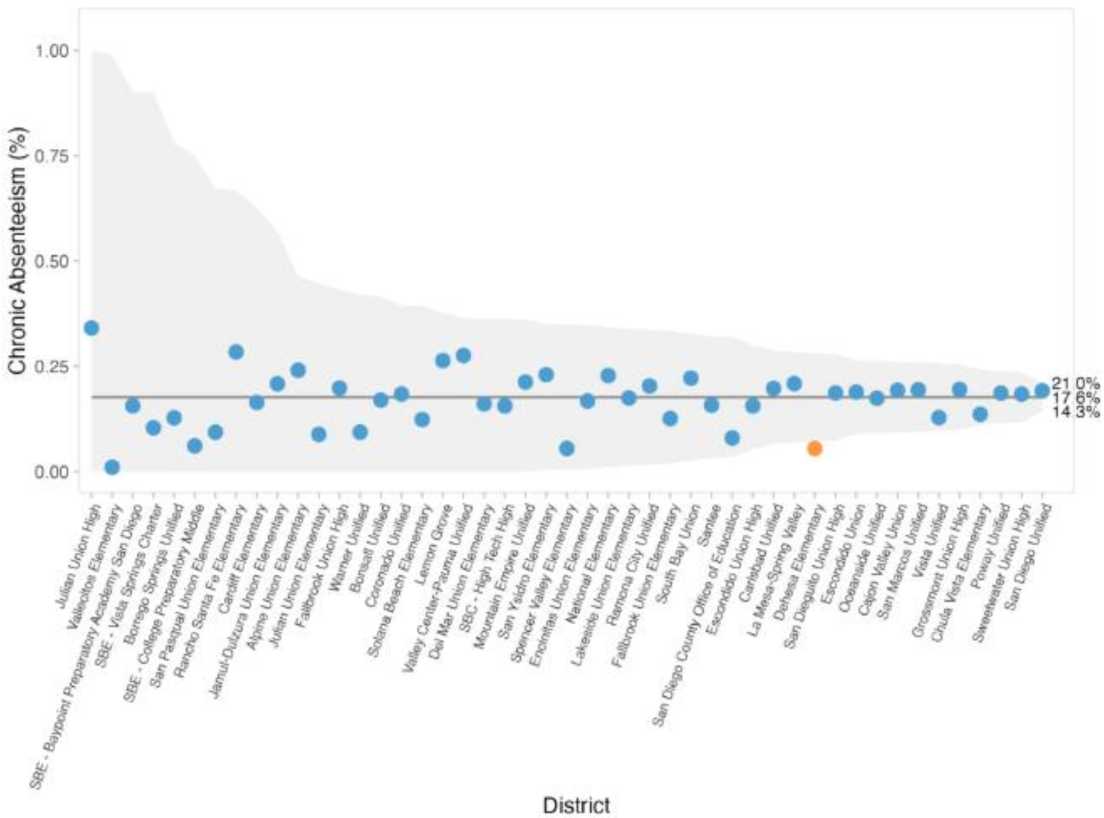


Step 3: Examining District-Level Variation Using P-prime (P') Charts

Drilling down further to the district level, a function was created in R that looped through all districts in each of the 58 counties in California and created P-prime (p') charts of chronic absenteeism rates from the 2023-2024 school year by county. P-prime (p') charts were again selected due to the size of the districts. The

Shewhart charts again, did not reveal much special cause variation. Charts were produced with adjusted and unadjusted data. The unadjusted datasets did not produce many special causes, but a few special causes were detected once the data was adjusted for income as in San Diego County in Figure 5 below. The shaded region represents all data falling within normal variation. Points outside of the shaded region represent special cause variation. The x axis represents districts in California and is ordered from smallest to largest. The y axis represents the percent of students who were chronically absent in 2023-2024.

Figure 5. Chronic Absenteeism Adjusted for Income by District in San Diego County 2023-2024 Year



Step 4: School-Level Analysis Using P Charts

Continuing to the more granular level of schools in a county posed a new decision point. The statewide regression by county did not show a strong correlation between the percentage of low-income students and absenteeism rates. In order to better understand the nature of this relationship, we applied a function to run linear regressions for all of the schools in each county examining the relationship between the percentage of low-income students and absenteeism rates. All of the slopes from each county's regression were then ordered by the magnitude of the standard error, creating a funnel plot at a 99% confidence level to determine if counties should be treated as their own systems or if the state should be treated as one whole system. The x axis was plotted on a log scale to spread the data out. The detection of multiple outliers in combination with a weak correlation by state led to the decision to adjust the data by income for each county individually.

Proceeding to the school level P charts, a function was used to loop through each county in the State of California, adjust the county's data for income, then create P charts of schools within each of the 58 counties to see if any special cause variation could be identified that could give us an opportunity to learn more about specific conditions or changes happening in the field that led to improved rates of absenteeism. P charts were selected over P-prime (p') charts in this case due to the smaller number of students by school rather than district. P charts are used when the outcome is a proportion or percentage derived from relatively small or moderate subgroup sizes (e.g., the percent of products failing quality inspection, the proportion of patients receiving a correct medication, or the proportion of students chronically absent). What P charts uniquely communicate is whether a subgroup's observed proportion is higher or lower than would be expected, given the amount of common-cause variation and subgroup size (Provost & Murray, 2022). In this application, P charts helped identify schools within counties whose 2023–24 absenteeism rates were unusually low (bright spots) or unusually high (dim spots).

The centerlines for the P charts were calculated as the overall average proportion of successes across all subgroups (Pbar). The upper and lower control limits for the P charts were calculated using the standard binomial three-sigma ($\pm 3\sigma$) method to determine the boundaries of expected system variation as follows: $UCL_i = Pbar + 3 \cdot \sqrt{[\, Pbar\,(1 - Pbar)/n_i]}$ and $LCL_i = Pbar - 3 \cdot \sqrt{[\, Pbar\,(1 - Pbar)/n_i]}$ for subgroup i with size n_i (Provost & Murray, 2022). For further clarity, Figure 6 below summarizes the Shewhart charts used in our study.

Figure 6. Comparison of Shewhart Chart Types: P, P', and I Charts

Comparison of Shewhart Chart Types: P, P' and I Charts

Chart Type	Best Used For	Type of Data	Centerline	Control Limit Formula
P Chart	Proportion or percentage data where subgroup sizes are small or moderate (e.g., individual schools)	Categorical (yes/no, absent vs. present).	Pbar is the overall average proportion.	$UCL_i = Pbar + 3 \cdot \sqrt{[\, Pbar\,(1 - Pbar)/n_i]}$; $LCL_i = Pbar - 3 \cdot \sqrt{[\, Pbar\,(1 - Pbar)/n_i]}$
P' Chart	Proportion data where subgroup sizes are very large (e.g., districts or counties).	Categorical (aggregated rates).	Pbar is the overall average proportion.	$UCL_i = Pbar + 3 \cdot \sigma_{p_i} \cdot \sigma_{z_i}$; $LCL_i = Pbar - 3 \cdot \sigma_{p_i} \cdot \sigma_{z_i}$
I Chart	Continuous or single-value data (e.g., year-to-year change in rates).	Continuous (numeric differences).	Xbar is the mean of individual values.	$UCL = Xbar + 2.66 \cdot MRbar$; $LCL = Xbar - 2.66 \cdot MRbar$

*P prime or Laney prime charts utilize a transformation based on the moving range to widen the limits when very large sample sizes (n) artificially narrow the range of the limits calculated by a Shewhart P chart (Provost & Murray, 2022).

Figure 6: A table detailing chart types, use case, type of data, centerline and control limit formulas.

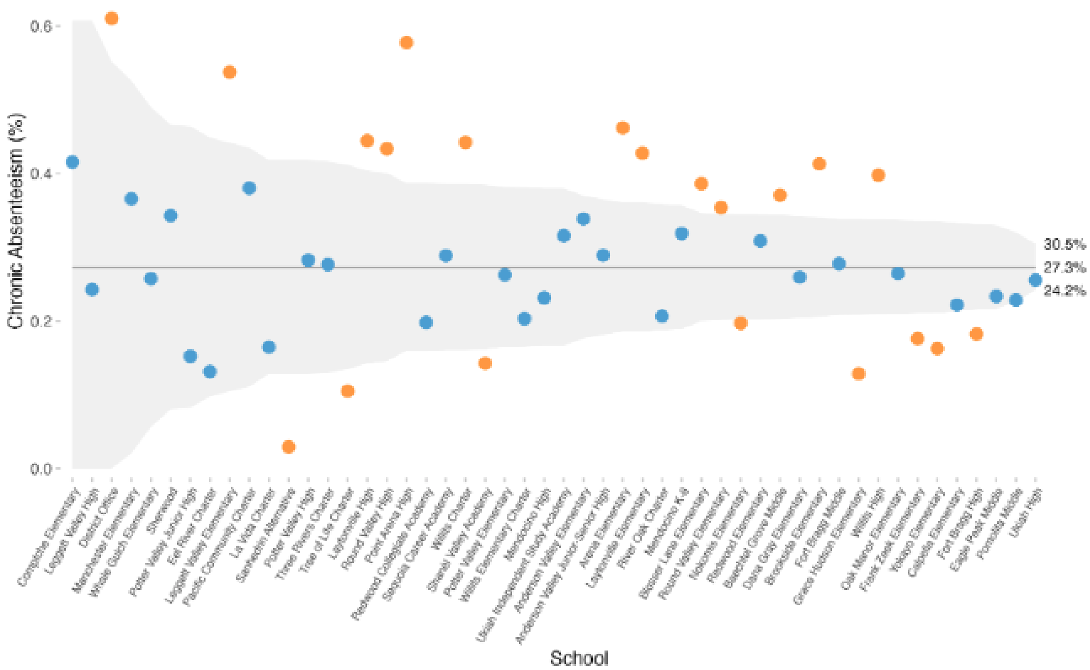
The P charts for each of the 58 counties produced a wealth of special cause variation, with a combined 3,975 outlier schools (schools falling outside the control limits) and 2,017 bright spot schools (schools falling below the lower control limit since lower rates are better when examining chronic absenteeism directly) out of a possible 9,245 schools across the state (slightly more schools in the denominator than the analysis of difference because schools needed data for both 21-22 and 23-24 years to be included). The adjusted P charts were also compared to P charts with unadjusted data to investigate the impact of income on the data. The school level P charts showed many more differences once adjusted. An example of an adjusted P chart is shown in Mendocino County in Figure 7. The shaded region represents all data falling within normal variation. Points outside of the shaded region represent special cause variation. The x axis represents schools in Mendocino County and is ordered from smallest to largest. The y axis represents the percent of students who were chronically absent in 2023-2024 after data was adjusted for income.

Results

Identification of "Super" Bright Spots

In the 23-24 school year, 2,017 bright spot schools out of 9,245 schools were able to achieve special cause lower rates of chronic absenteeism than the other schools in their county. 93 bright spot schools out of

Figure 7. Chronic Absenteeism Adjusted for Income by School in Mendocino County 2023-2024 Year



8,880 schools were able to achieve special cause improvement post-pandemic. This is a large number of schools to interview, and each analysis only answers one of the two questions posed by the team. To address both simultaneously, an inner join of the analysis of difference bright spots with the 23-24 bright spots was done which led to the identification of 36 "super" bright spot schools that were identified as bright spots in both datasets. The inner join of these two tables created a list of 36 schools in the state that knew both how to produce low levels of chronic absenteeism and how to improve their systems rapidly. These schools have achieved special cause results for their students, and we hypothesized that they could be a wealth of information to benefit the entire State of California. The list of schools was provided to the RAISE network.

Qualitative Learning from "Super" Bright Spots

To deepen our understanding of the underlying causes that might explain unusually strong performance, RAISE researchers conducted a qualitative inquiry with a small subset of eight "super" bright spot schools, constituting a judgment sample (Perla & Provost, 2012). These schools were purposefully selected from the list of 36 because they varied in geographic region, school level (elementary, middle, high), and demographic composition, allowing the researchers to explore whether common practices emerged across diverse contexts. The intent of this qualitative component was not to conduct a full-scale qualitative study but rather to generate initial, practice-based insights consistent with the Positive Deviance framework's second step: studying positive deviants (Low et al., 2025).

Researchers from the RAISE Network engaged school leaders and attendance teams in empathy interviews focused on understanding specific practices, routines, or organizational conditions that may have contributed to unusually low absenteeism rates or unusually rapid improvement. Interviews followed a semistructured protocol that included questions about daily attendance routines, communication strategies with families, use of data for early intervention, and schoolwide conditions that promote attendance. When possible, researchers also observed attendance meetings or reviewed relevant artifacts such as outreach logs, staff meeting agendas, or communication tools used with families. Interview notes and artifacts were analyzed and synthesized for the RAISE team.

The purpose of this qualitative work was to complete the third step of the Positive Deviance approach—testing emergent hypotheses—by examining whether patterns repeatedly surfaced across schools flagged by the Shewhart analysis (Low et al., 2025). The qualitative findings were then synthesized into a preliminary change package of five practices that is now being tested and refined within the RAISE Network. While this qualitative component is intentionally limited in scope, it demonstrates how Shewhart-based identification of bright spots can be combined with targeted inquiry to generate actionable, practice-based insights. The foundational package can be requested through the RAISE Network (High Tech High Graduate School of Education National Coalition for Improvement in Education, 2025). In addition, the code to replicate this analysis is included in the supplementary material.

Discussion

Integrating Quantitative and Qualitative Insights

The combined methods of an analysis of difference plotted on I charts and the application of P and P-prime (p') charts can be used as a tool for identifying bright spots or places to investigate for any measure (other types of control charts can be used as well depending on the nature of the data). The joining of the two analyses can reduce the number of special causes and increase the likelihood of discovering evidence-based practices. The implications of this method include opportunities to conduct qualitative analysis to uncover site-specific special causes. This can lead to empathy interviews that are useful for mining specific practices and change ideas for building a theory of improvement. Further field testing will determine the validity of change ideas that arise from the interviews as some ideas may be site-specific or not suitable for replication at scale. The application of these methods can be applied to any field across a wide array of problems of practice.

The four steps of the Positive Deviance approach described earlier align closely with our study: identify positive deviants, study positive deviants, test hypotheses, and spread best practice (Low et al., 2025). First, we were able to identify positive deviants by using I charts and P/P-prime (p') charts to detect schools demonstrating special-cause reductions in chronic absenteeism and unusually low absenteeism relative to their county context, to identify "super" bright spot positive deviants. Second, the RAISE team conducted empathy interviews, site visits, and document reviews to study positive deviants, producing early learning about common practices across "super" bright spot schools. Third, while the step of testing hypotheses is less explicitly represented in our current analysis, it is embedded in the ongoing work of the RAISE Network; the emergent change ideas from bright spot schools are being operationalized and refined as networked improvement teams introduce, monitor, and adjust their change ideas over time. This step is a primary use of Shewhart charts in monitoring a measure over time: to determine if changes tested using Plan-Do-Study-Act cycles have an important impact on the measures (Grunow et al., 2024; Provost & Murray, 2022). Finally, the RAISE change package and statewide dissemination efforts represent the spread of best practice. Together, these elements demonstrate how the Positive Deviance approach can be operationalized within a statewide continuous-improvement initiative aimed at reducing chronic absenteeism.

Learning from Variation Using Both Bright and Dim Spots

Bright spots identified in each county for current performance and/or improvement can be further researched to uncover group-specific causes of success. In addition to bright spots where we focused our study, identifying dim spots—schools with unusually high absenteeism—might reveal contextual barriers and inform targeted support strategies. Dim spots are provided in the supplemental code as well for further opportunities for qualitative analysis. Subgrouping data by school level (elementary, middle, high) could also provide interesting insights into future bright and dim spot analyses.

Implications for Decision-Making and System Improvement

Shewhart charts serve as powerful visual tools that translate complex statistical concepts into accessible insights for district leaders and policymakers, facilitating data-informed decision-making. Shewhart charts can be used in real-time process monitoring to detect both common and special causes of variation as schools and districts begin to test change ideas. The teams in the RAISE Network will be using the charts in this way. Time series control charts are useful to visually detect when a special cause change occurs in the system. As schools test and adapt change ideas to their local contexts, chronic absenteeism data will be monitored biweekly over subsequent years, so progress can be monitored over time. Additionally, when these charts are applied retrospectively to educational data, as in this case, they enable researchers to identify historical instances of systemic improvement or decline. In this context, identifying bright spots reflects detecting schools that have performed outside expected limits, suggesting the presence of effective local practices or interventions. This retrospective use provides a foundation for policy learning and adaptation. Additionally, district and county leaders can identify sources ripe for learning and can conduct interviews to learn from their bright spot schools about what is working and can be shared with other schools.

Comparing Shewhart Charts with Other Analytic Approaches

Other methods for identifying bright spots include regression, residuals, z-scores and clustering methods. The problem with such methods is that they are primarily nonvisual and they are difficult to explain to a practitioner audience and so are less accessible to decision makers. Bright spots using Shewhart charts in this study were also compared with bright spots from z-scores, to evaluate consistency and cross-method validity. Analysis of z-scores produced 19 "super" bright spots with 9 overlapping "super" bright spot schools with the Shewhart analysis, further supporting the case for the use of Shewhart charts as an accessible and generalizable method in educational and social science contexts. P charts also adjust limits for school size, allowing large schools to be flagged more easily, where z-scores provide one standard deviation per county.

Limitations and Considerations

Additionally, there are limitations with Shewhart charts, particularly with P and P-prime (p') charts. In some cases, a P chart with large subgroup sizes (greater than 3,000) may identify everything as special cause variation while a P-prime (p') chart with these same large subgroup sizes may identify a majority of the points as common cause variation. A solution does not currently exist to remedy this dilemma which requires content experts to help in interpretation of charts (Provost & Murray, 2022).

Conclusion

Chronic absenteeism rates in the State of California remain alarmingly high post-pandemic. In a statewide effort to reduce rates by half by 2029, RAISE shared the findings of this analysis in a public change package so schools across the state can improve chronic absenteeism. The application of Shewhart charts for identifying positive deviants, combining bright spots from an analysis of difference and current performance to identify "super" bright spots, and qualitative methods to uncover key changes occurring within a system, is a strong method that can enhance any field interested in learning from variation. The ability to process and narrow down statewide data to an actionable list of schools to focus on can be replicated in other states or locations and applied to various metrics.

Declaration of Conflicting Interests

The authors confirm that there are no conflicts of interest with respect to the research, authorship and/or publication of this paper.

Received: 8/13/2025. **Accepted:** 12/19/2025. **Published:** 1/X/2026.

Citation: Geary, E., Provost, L., & Bennett, B. (2026). Applied Shewhart charts in education to understand chronic absenteeism across the state of California: A methods application. *Practical Assessment, Research, & Evaluation*, 31(1)(3). Available online: https://doi.org/10.7275/pare.3328

Corresponding Author: Erica Geary, High Tech High Graduate School of Education.
Email: egeary@hthgse.edu

References

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*(2), 87–107.

Ansari, A., & Pianta, R. C. (2019). School absenteeism in the first decade of education and outcomes in adolescence. *Journal of School Psychology, 76*, 48–61.

Bakir, S. T., Starnes, B., & Self, R. (2005). A simple tool for quality assessment in education. *Journal of College Teaching & Learning*, *2*(7), 45-52.

Bennett, B. (2020). Special Delivery. Quality Progress, 53(1), 18–24.

Bi, H. H. (2020). A robust approach for assessing teaching performance during regular reviews. *Studies in Higher Education, 45*(11), 2217–2230.

Black, L. I., & Elgaddal, N. (2024). *Chronic school absenteeism for health-related reasons among children ages 5–17 years: United States, 2022.* National Center for Health Statistics.

Brunner, M., Keller, U., Wenger, M., Fischbach, A., & Lüdtke, O. (2018). Between-school variation in students' achievement, motivation, affect, and learning strategies: Results from 81 countries for planning group-randomized trials in education. Journal of Research on Educational Effectiveness, 11(3), 452–478.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). Learning to Improve: How America's Schools Can Get Better at Getting Better. Harvard Education Press.

California Department of Education. (2025). *Chronic absenteeism data.* https://www.cde.ca.gov/ds/ad/filesabd.asp

Deming, W. E. (1951). Elementary principles of the statistical control of quality—series of lectures. Tokyo, Japan: Nippon Kagaku Gijutsu Remnei.

Figlio, D., Loeb, S., Ozek, U., & Strunk, K. (2023). Is the rise in high school graduation rates real? High-stakes school accountability and strategic behavior. Economics of Education Review, 94, 102386.

García, E., & Weiss, E. (2018, September 25). *Student absenteeism: Who misses school and how missing school matters for performance*. Economic Policy Institute.

Gottfried, M. A. (2014). Chronic absenteeism and its effects on students' academic and socioemotional outcomes. *Journal of Education for Students Placed at Risk, 19*(2), 53–75.

Grunow, A., Park, S., & Bennett, B. (2024). *Journey to improvement: A team guide to systems change in education, health care, and social welfare.* Rowman & Littlefield.

High Tech High Graduate School of Education National Coalition for Improvement in Education. (2025, March 6). *RAISE Network*. High Tech High Graduate School of Education. Retrieved June 27, 2025, from https://hthgse.edu/ncie/raise/

LeMahieu, P. G., Nordstrum, L. E., & Gale, D. (2017). Positive deviance: Learning from positive anomalies. Quality Assurance in Education, 25(1), 109–124.

LeMahieu, Paul, Bryk, Anthony, Grunow, Alicia & Gomez, Louis. (2017). Working to improve: seven approaches to improvement science in education. Quality Assurance in Education. 25. 2-4.

Low, A. B., Maresh, J. R., Jarvi, T., Esser, J., Rawson, T., Fernandez, N., Chiem, J. L., Feldman, R. M., Hale, J. V., Merguerian, P. A., Lander, T. A., & Martin, L. D. (2025). Perspectives in perioperative quality improvement: Case studies in positive deviance using real-world data. *Pediatric Anesthesia*. Advance online publication.

Malkus, N. (2024). *Long COVID for public schools: Chronic absenteeism in the aftermath of pandemic school closures.* American Enterprise Institute.

Marsh, D. R., Schroeder, D. G., Dearden, K. A., Sternin, J., & Sternin, M. (2004). The power of positive deviance. BMJ, 329, 1177–1179.

McKay, S. (2017, September 20). *Quality improvement approaches: Positive deviance*. Carnegie Foundation for the Advancement of Teaching. https://www.carnegiefoundation.org/blog/quality-improvement-approaches-positive-deviance/

Meijer, R. R. (2002). Outlier detection in high-stakes certification testing. Journal of Educational Measurement, 39, 219–233

Ming, N., & Kennedy, A. I. (2020). Developing and using indicators for continuous improvement. *Teachers College Record, 122*(14), 1–32. https://doi.org/10.1177/016146812012201407

Nelson, L. S. (1988). Control Charts: Rational Subgroups and Effective Applications, Journal of Quality Technology, 20, 73-75.

Omar, M. H. (2010). Statistical process control charts for measuring and monitoring temporal consistency of ratings. *Journal of Educational Measurement, 47*(1), 18–35.

Perla, R. J., & Provost, L. P. (2012). Judgment sampling: A health care improvement perspective. *Quality Management in Health Care, 21*(3), 169–175.

Provost, L. P., & Murray, S. K. (2022). *The health care data guide: Learning from data for improvement* (2nd ed.). John Wiley & Sons.

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.3.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rahman MA, Renzaho AMN, Kundu S, Awal MA, Ashikuzzaman M, Fan L, et al. (2023) Prevalence and factors associated with chronic school absenteeism among 207,107 in-school adolescents: Findings from cross-sectional studies in 71 low-middle and high-income countries. PLoS ONE 18(5): e0283046.

Romero, M., & Lee, Y. (2007). *A national portrait of chronic absenteeism in the early grades*. National Center for Children in Poverty.

Rumberger, R. W. (2015). *Dropping out: Why students drop out of high school and what can be done about it*. Harvard University Press.

Rumberger, R. W., & Larson, K. A. (1998). Student mobility and the increased risk of high school dropout. *American Journal of Education, 107*(1), 1–35.

Santibañez, L., & Guarino, C. (2020, October). *The effects of absenteeism on academic and social-emotional outcomes: Lessons for COVID-19* [Policy brief]. Policy Analysis for California Education.

Sefik, M. (1998). Importance of the rational subgroups in designing control charts. *Computers & Industrial Engineering, 35*(1–2), 205–208.

Shewhart, W. A. (1926). Quality control charts. Bell System Technical Journal, 5(4), 593–603.

Wackerly, B. (2023). *qicharts2: Quality improvement charts* (Version 0.7.4) [R package]. https://CRAN.R-project.org/package=qicharts2

Wan, S., & Keller, L. A. (2023). Using cumulative sum control chart to detect aberrant responses in educational assessments. *Practical Assessment, Research & Evaluation, 28*, Article 2.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686.

Wickham, H., & Wiernik, M. (2023). *bigrquery: An interface to Google's BigQuery* (Version 1.4.2) [R package]. https://CRAN.R-project.org/package=bigrquery

Appendix A

Supplemental R Code

````
---
title: "RAISE Chronic Absenteeism Notebook"
output: html_notebook
---

```{r}
#1
#Load libraries
library(bigrquery)
library(tidyverse)
library(readxl)
library(qicharts2)
bq_auth()
```

```{r}
#2
#Download the data set from bigquery
project_id <- "calcan-427720"
dataset_id <- "cde_public"
table_id <- "chronic_absenteeism"

# Define the full table reference
table_ref <- bq_table(project_id, dataset_id, table_id)

# Download the entire table as a data frame
df_all_CA_updated <- bq_table_download(table_ref)

# View the first few rows of the data frame
head(df_all_CA_updated)
````
````

```
# Trim DASS
df_all_CA_updated <- df_all_CA_updated %>%
mutate(DASS = trimws(tolower(DASS)))
```

```{r}
#3
# Prepare data to find out which school improved most rapidly post
  pandemic
# Find delta of CA from 2021-2022 to 2023-2024
alldf <- df_all_CA_updated %>%
    filter(Academic_Year == "2023-24" &
             Aggregate_Level == "S" &
             Reporting_Category == "TA" &
             (DASS == "no" | is.na(DASS))) %>%
  mutate(county_district_school = paste(County_Name, District_Name,
  School_Name, sep = " ")) %>%
    select(Academic_Year, County_Name, District_Name, School_Name,
  county_district_school, Charter_School, Aggregate_Level,
           Reporting_Category,
  ChronicAbsenteeismEligibleCumulativeEnrollment,
           ChronicAbsenteeismCount, ChronicAbsenteeismRate) %>%
  distinct(county_district_school, .keep_all = TRUE)

df21 <- df_all_CA_updated %>%
    filter(Academic_Year == "2021-22" &
             Aggregate_Level == "S" &
             Reporting_Category == "TA" &
             (DASS == "no" | is.na(DASS))) %>%
  mutate(county_district_school = paste(County_Name, District_Name,
  School_Name, sep = " ")) %>%
    select(county_district_school, ChronicAbsenteeismRate)%>%
 distinct(county_district_school, .keep_all = TRUE)
```

````
deltajoin <- left_join(df21, alldf, by = "county_district_school")
  %>%
  mutate(delta21_24 = ChronicAbsenteeismRate.x-
  ChronicAbsenteeismRate.y)

```


```{r}
#4
# Function for I charts for deltas in each county
# Define the function for schools within each county in California
delta_function <- function(data) {

  # Filter the data for specified conditions
  filtered_data <- data

  # Initialize an empty dataframe to store outliers
  results <- data.frame(
    County_Name = character(),
    School_Name = character(),
    sigma_signal = numeric(),
    lcl = numeric(),
    ucl = numeric(),
    delta = numeric(),
    stringsAsFactors = FALSE
  )

  # Get unique County_Name values
  unique_county <- unique(filtered_data$County_Name)

  # Iterate over each County_Name
  for (county in unique_county) {

    # Filter data for the current county
````

```
  county_data <- data %>%
    filter(County_Name == county & !is.na(delta21_24))

  # Skip counties with insufficient data
  if (nrow(county_data) < 2) {
    print(paste("Insufficient data for County:", county))
    next
  }

   # Create I chart with deltas
    chart <- qic(
      x = reorder(county_district_school,
ChronicAbsenteeismEligibleCumulativeEnrollment),
      y = delta21_24,
      data = county_data,
      chart = 'i',
      title = paste("Delta of Chronic Absenteeism by School in",
county, "County 2021-2022 to 2023-2024 Year"),
      xlab = "School",
      x.angle = 65,
      point.size = 1,
      show.labels = TRUE
    )

    print(chart)

  # Extract sigma signals from the chart's data
  chart_data <- chart$data
  outlier_data <- chart_data %>%
    filter(sigma.signal == 1 & y > ucl) %>%
    mutate(
      County_Name = county
    ) %>%
    select(
```

```
        County_Name,
        county_district_school = x, delta21_24 = y,
        sigma.signal, lcl, ucl
      )

    # Append to the results dataframe
    results <- bind_rows(results, outlier_data)
  }

  return(results)
}

# Use function on dataset
deltalist <- delta_function(deltajoin)

# View results
View(deltalist)
```

```{r}
#5
# Prepare the data set for counties within California
# Find the number of low income students by county and the total
  number of students, join to create a new variable,
  percent_low_income
alldfc <- df_all_CA_updated %>%
    filter(Academic_Year == "2023-24" &
             Aggregate_Level == "C" &
             Reporting_Category == "TA" &
             (DASS == "no" | is.na(DASS)) &
              Charter_School == "All") %>%
    select(Academic_Year, County_Name, School_Name, Aggregate_Level,
           Reporting_Category,
  ChronicAbsenteeismEligibleCumulativeEnrollment,
```

```
           ChronicAbsenteeismCount, ChronicAbsenteeismRate)


lowincomedfc <- df_all_CA_updated %>%
    filter(Academic_Year == "2023-24" &
             Aggregate_Level == "C" &
             Reporting_Category == "SS" &
             (DASS == "no" | is.na(DASS)) &
              Charter_School == "All") %>%
    select(County_Name,
  ChronicAbsenteeismEligibleCumulativeEnrollment)


incomejoinc <- left_join(alldfc, lowincomedfc, by = "County_Name")
  %>%
  mutate(percent_low_income =
  ChronicAbsenteeismEligibleCumulativeEnrollment.y/ChronicAbsenteeism
  EligibleCumulativeEnrollment.x)



```

```{r}
#6
# Run a regression for all counties in California, adjust chronic
  absenteeism for income, create P prime charts of adjusted
  absenteeism by county
regression_county <- lm(ChronicAbsenteeismRate ~ percent_low_income,
  data = incomejoinc)


regression_county_summary <- summary(regression_county)


 # Calculate the mean of percent_low_income
    rcmean_percent_low_income <-
  mean(incomejoinc$percent_low_income, na.rm = TRUE)


  # Adjust absenteeism rates
    rccoefficient <- coef(regression_county)["percent_low_income"]
```

```
    incomejoinc1 <- incomejoinc %>%
      mutate(
        adjusted_absenteeism = ChronicAbsenteeismRate -
          rccoefficient * (percent_low_income -
  rcmean_percent_low_income),
        number_adjusted = adjusted_absenteeism / 100 *
  ChronicAbsenteeismEligibleCumulativeEnrollment.x
      )

   rcchart <- qic(
      x = reorder(County_Name,
  ChronicAbsenteeismEligibleCumulativeEnrollment.x),
      y = number_adjusted,
      n = ChronicAbsenteeismEligibleCumulativeEnrollment.x,
      data = incomejoinc1,
      chart = 'pp',
      title = paste("Chronic Absenteeism by County in California
  2023-2024 Year"),
      xlab = "County",
      x.angle = 65,
      point.size = 1,
      show.labels = TRUE
    )
    print(rcchart)
```

```
```{r}
#7
# Prepare data to find the percent of low income students by
  district
# Join the dataframes of all students and low income students to
  find the percent of low income students by district
alldfd <- df_all_CA_updated %>%
    filter(Academic_Year == "2023-24" &
             Aggregate_Level == "D" &
```
```

```
            Reporting_Category == "TA" &
            (DASS == "no" | is.na(DASS))) %>%
  mutate(county_district = paste(County_Name, District_Name, sep = "
  ")) %>%
    select(Academic_Year, County_Name, District_Name, School_Name,
  county_district, Aggregate_Level,
          Reporting_Category,
  ChronicAbsenteeismEligibleCumulativeEnrollment,
          ChronicAbsenteeismCount, ChronicAbsenteeismRate) %>%
  distinct(county_district, .keep_all = TRUE)


lowincomedfd <- df_all_CA_updated %>%
    filter(Academic_Year == "2023-24" &
            Aggregate_Level == "D" &
            Reporting_Category == "SS" &
            (DASS == "no" | is.na(DASS))) %>%
  mutate(county_district = paste(County_Name, District_Name, sep = "
  ")) %>%
    select(county_district,
  ChronicAbsenteeismEligibleCumulativeEnrollment)%>%
 distinct(county_district, .keep_all = TRUE)


incomejoind <- left_join(alldfd, lowincomedfd, by =
  "county_district") %>%
  mutate(percent_low_income =
  ChronicAbsenteeismEligibleCumulativeEnrollment.y/ChronicAbsenteeism
  EligibleCumulativeEnrollment.x)
```

```
```{r}
#8
# Create District in County Function to run a regression by county,
  adjust chronic absenteeism for income and create P prime charts,
  returning bright spot districts


# Define the function for districts within each county in California
  to find outlier bright spots
```
```

```
incomefunction_bycounty_district <- function(data) {

  # Filter the data for specified conditions
  filtered_data <- data

  # Initialize an empty dataframe to store outliers and regression
  results
  results <- data.frame(
    County_Name = character(),
    District_Name = character(),
    p_value = numeric(),
    r_squared = numeric(),
    standard_error = numeric(),
    intercept = numeric(),
    slope = numeric(),
    School_Name = character(),
    adjusted_absenteeism = numeric(),
    sigma_signal = numeric(),
    lcl = numeric(),
    ucl = numeric(),
    stringsAsFactors = FALSE
  )

  # Get unique County_Name values
  unique_county <- unique(filtered_data$County_Name)

  # Iterate over each County_Name
  for (county in unique_county) {

    # Filter data for the current county
    county_data <- data %>%
      filter(County_Name == county & !is.na(percent_low_income) &
  !is.na(ChronicAbsenteeismRate))
```

```
    # Skip counties with insufficient data
    if (nrow(county_data) < 2) {
      print(paste("Insufficient data for County:", county))
      next
    }

# Fit the linear regression model
regression <- lm(ChronicAbsenteeismRate ~ percent_low_income, data =
  county_data)
regression_summary <- summary(regression)

 # Extract p-value and r-squared and standard_error and intercept
    p_value <- coef(summary(regression))["percent_low_income",
  "Pr(>|t|)"]
    r_squared <- regression_summary$r.squared
    standard_error <-
  coef(summary(regression))["percent_low_income", "Std. Error"]
    intercept <- coef(regression)["(Intercept)"]
    slope <- coef(regression)["percent_low_income"]

 # Calculate the mean of percent_low_income
    mean_percent_low_income <- mean(county_data$percent_low_income,
  na.rm = TRUE)

  # Adjust absenteeism rates
    coefficient <- coef(regression)["percent_low_income"]
    county_data <- county_data %>%
      mutate(
        adjusted_absenteeism = ChronicAbsenteeismRate -
          coefficient * (percent_low_income -
  mean_percent_low_income),
        number_adjusted = adjusted_absenteeism / 100 *
  ChronicAbsenteeismEligibleCumulativeEnrollment.x
      )
```

```
  # Ensure number_adjusted is non-negative
 county_data$number_adjusted <- pmax(county_data$number_adjusted,
0)



  # Create the P’ control chart
  chart <- qic(
    x = reorder(District_Name,
ChronicAbsenteeismEligibleCumulativeEnrollment.x),
    y = number_adjusted,
    n = ChronicAbsenteeismEligibleCumulativeEnrollment.x,
    data = county_data,
    chart = 'pp',
    title = paste("Chronic Absenteeism by District in", county,
"County 2023-2024 Year"),
    xlab = "District",
    x.angle = 65,
    point.size = 1,
    show.labels = TRUE
  )
  print(chart)

  # Extract sigma signals from the chart's data
  chart_data <- chart$data
  outlier_data <- chart_data %>%
    filter(sigma.signal == 1 & y <lcl) %>%
    mutate(
      County_Name = county,
      p_value = p_value,
      r_squared = r_squared,
      standard_error = standard_error,
      intercept = intercept,
      slope = slope
    ) %>%
```

````
      select(
        County_Name, p_value, r_squared, standard_error, intercept,
  slope,
        District_Name = x, adjusted_absenteeism = y,
        sigma.signal, lcl, ucl
      )

    # Append to the results dataframe
    results <- bind_rows(results, outlier_data)
  }

  return(results)
}

# Use Function
outliers_adjusted_districts <-
  incomefunction_bycounty_district(incomejoind)
View(outliers_adjusted_districts)


```

```{r}
#9
# Prepare data for school level within county
# Join dataframe of all students and low income students to find the
  percent low income students by school
# Create unique keys by adding county_district_school to dataframe
alldf <- df_all_CA_updated %>%
    filter(Academic_Year == "2023-24" &
             Aggregate_Level == "S" &
             Reporting_Category == "TA" &
             (DASS == "no" | is.na(DASS))) %>%
````

```
  mutate(county_district_school = paste(County_Name, District_Name,
  School_Name, sep = " ")) %>%
    select(Academic_Year, County_Name, District_Name, School_Name,
  county_district_school, Charter_School, Aggregate_Level,
           Reporting_Category,
  ChronicAbsenteeismEligibleCumulativeEnrollment,
           ChronicAbsenteeismCount, ChronicAbsenteeismRate) %>%
  distinct(county_district_school, .keep_all = TRUE)


lowincomedf <- df_all_CA_updated %>%
    filter(Academic_Year == "2023-24" &
             Aggregate_Level == "S" &
             Reporting_Category == "SS" &
             (DASS == "no" | is.na(DASS))) %>%
  mutate(county_district_school = paste(County_Name, District_Name,
  School_Name, sep = " ")) %>%
    select(county_district_school,
  ChronicAbsenteeismEligibleCumulativeEnrollment)%>%
 distinct(county_district_school, .keep_all = TRUE)


incomejoin <- left_join(alldf, lowincomedf, by =
  "county_district_school") %>%
  mutate(percent_low_income =
  ChronicAbsenteeismEligibleCumulativeEnrollment.y/ChronicAbsenteeism
  EligibleCumulativeEnrollment.x)

```

```
```{r}
#10
# Create School in County Function to run regressions by county,
  adjust for income, create P charts of adjusted absenteeism by
  school, and return bright spot schools


# Define the function for schools within each county in California
  to find outliers
incomefunction_bycounty <- function(data) {
```
```

```
# Filter the data for specified conditions
filtered_data <- data

# Initialize an empty dataframe to store outliers and regression
results
results <- data.frame(
  County_Name = character(),
  District_Name = character(),
  p_value = numeric(),
  r_squared = numeric(),
  standard_error = numeric(),
  intercept = numeric(),
  slope = numeric(),
  School_Name = character(),
  adjusted_absenteeism = numeric(),
  sigma_signal = numeric(),
  lcl = numeric(),
  ucl = numeric(),
  stringsAsFactors = FALSE
)

# Get unique County_Name values
unique_county <- unique(filtered_data$County_Name)

# Iterate over each County_Name
for (county in unique_county) {

  # Filter data for the current county
  county_data <- data %>%
    filter(County_Name == county & !is.na(percent_low_income) &
!is.na(ChronicAbsenteeismRate))
# Skip counties with insufficient data
  if (nrow(county_data) < 2) {
```

```
      print(paste("Insufficient data for County:", county))
      next
    }

# Fit the linear regression model
regression <- lm(ChronicAbsenteeismRate ~ percent_low_income, data =
  county_data)
regression_summary <- summary(regression)

 # Extract p-value and r-squared and standard_error and intercept
    p_value <- coef(summary(regression))["percent_low_income",
  "Pr(>|t|)"]
    r_squared <- regression_summary$r.squared
    standard_error <-
  coef(summary(regression))["percent_low_income", "Std. Error"]
    intercept <- coef(regression)["(Intercept)"]
    slope <- coef(regression)["percent_low_income"]

 # Calculate the mean of percent_low_income
    mean_percent_low_income <- mean(county_data$percent_low_income,
  na.rm = TRUE)

  # Adjust absenteeism rates
    coefficient <- coef(regression)["percent_low_income"]
    county_data <- county_data %>%
      mutate(
        adjusted_absenteeism = ChronicAbsenteeismRate -
          coefficient * (percent_low_income -
  mean_percent_low_income),
        number_adjusted = adjusted_absenteeism / 100 *
  ChronicAbsenteeismEligibleCumulativeEnrollment.x
      )

    # Ensure number_adjusted is non-negative
```

```
  county_data$number_adjusted <- pmax(county_data$number_adjusted,
0)



  # Create the P control chart
  chart <- qic(
    x = reorder(county_district_school,
ChronicAbsenteeismEligibleCumulativeEnrollment.x),
    y = number_adjusted,
    n = ChronicAbsenteeismEligibleCumulativeEnrollment.x,
    data = county_data,
    chart = 'p',
    title = paste("Chronic Absenteeism Adjusted for Income by
School in", county, "County 2023-2024 Year"),
    xlab = "School",
    x.angle = 65,
    point.size = 1,
    show.labels = TRUE
  )
  print(chart)
  # Extract sigma signals from the chart's data
  chart_data <- chart$data
  outlier_data <- chart_data %>%
    filter(sigma.signal == 1 & y<lcl) %>%
    mutate(
      County_Name = county,
      p_value = p_value,
      r_squared = r_squared,
      standard_error = standard_error,
      intercept = intercept,
      slope = slope
    ) %>%
    select(
```

````
        County_Name, p_value, r_squared, standard_error, intercept,
  slope,
        county_district_school = x, adjusted_absenteeism = y,
        sigma.signal, lcl, ucl
      )

    # Append to the results dataframe
    results <- bind_rows(results, outlier_data)
  }

  return(results)
}

# Use Function
outliers_adjusted <- incomefunction_bycounty(incomejoin)
View(outliers_adjusted)

```
```{r}
#11
# Inner Join deltalist and outliers_adjusted for super bright spot
  dataframe

# Perform the inner join on county_district_school
superbrightspots <- inner_join(deltalist, outliers_adjusted, by =
  "county_district_school")
superbrightspots

```
````
````