

# National Center Brief

## The Importance of Disaggregating Student Data

April 2012

Disaggregating data means breaking down information into smaller subpopulations. For instance, breaking data down into grade level within school aged students, country of origin within racial/ethnic categories, or gender among student populations are all ways of disaggregating data.

Disaggregating student data into subpopulations can help schools and communities plan appropriate programs, decide which evidence-based interventions to select (i.e. have they been evaluated with the target population), use limited resources where they are needed most, and see important trends in behavior and achievement. Collecting and analyzing data can seem intimidating to someone without a strong statistics background, however, many of the tools you need are readily available. This brief provides:

- An overview of the value of disaggregating data
- Common areas of data to disaggregate
- Examples of how disaggregated data has been used
- Limitations of disaggregating data, particularly data describing students

## The Importance of Disaggregating Data

As Safe Schools/Healthy Students sites, you are already collecting important information about students in your district. In addition to the federally required GPRA measures, many states also use the Youth Risk Behaviors Survey (YRBS), in addition to other smaller student information surveys. These data are incredibly valuable; however, much of it is combined, or aggregated, to represent the student population generally. Disaggregating data can show where aggregate data are masking discrepancies. For example, many schools look at student data separated by race/ethnic group. By looking at these data among smaller subpopulations (disaggregating the data), you can see if outcomes vary by subpopulation and if some subpopulations' strong results are masking others' poorer results.

The American Community Survey of 2006, for example, reported a relatively low rate (14 percent) of Asians achieving less than a high school level education. However, disaggregating the data showed discrepancies. Specifically, among Hmong, Laotian, and Cambodian populations, the rates of achieving less than a high school level education were more than double the 14 percent national average: 39 percent (Hmong), 38 percent (Laotian), and 35 percent (Cambodian) (Khan & Ro, 2009). This information could be used for targeted outreach programs and to better inform teachers and other youth-serving providers about which students are at higher risk for lower academic success, information that could easily be missed by only looking at the broader Asian totals. This information

could also be used to inform any needed adaptations to evidence-based programs used with these populations.

Disaggregating data can also valuably inform *program implementation and monitoring*. For example, if student survey results show a gender divide in truancy rates, it might be efficient and useful to have gender specific targeted drop-out prevention and attendance programs. This could ensure that resources are spent where they are needed most.

Disaggregated data can also provide measures of the *effectiveness and equity of a program or ways to view achievement measures*. For example,

- Is there a gender or racial/ethnic outcome difference among students who participate in a particular evidence based intervention?
- Are students in particular grades or with certain teachers performing better, on average, than other grades?
- Are high socio-economic status students overrepresented in accessing and receiving services?

In this way, disaggregated data can confirm perceptions of what is really occurring (i.e. teachers have noticed that ninth grade students consistently perform better on standardized math tests than their eleventh grade counterparts) or debunk stereotypes (i.e. students of lower SES abuse alcohol and other drugs more than their higher SES classmates).

One area where this type of information is commonly used is to show disproportionate minority contact, the number of times a youth is involved with the court system. In fact, the [Office of Juvenile Justice and Delinquency Prevention](#) (OJJDP) uses a specific indicator to collect this information, called the Relative Rate Index (RRI). The RRI compares rates of contact with the juvenile justice and law enforcement systems at various stages among different groups of youth. It can show if there are differences in arrest rates or court sentences, for example, between racial/ethnic groups that are not explained by simple differences in population numbers.

A similar step was taken by the Department of Health and Human Services (HHS) as part of the Affordable Care Act. In the recently released, *HHS Action Plan to Reduce Racial and Ethnic Health Disparities*, a priority was placed on "ensuring that data collection standards for race, ethnicity, sex, primarily language, and disability status are implemented throughout HHS-supported programs, activities, and surveys" (HHS, 2012). Disaggregated data can be used to see if there are meaningful differences by subpopulations in who is accessing mental services and what treatments are successful. This can inform evidence-based programs focusing on mental health as well as documenting a possibly overlooked need for mental health providers.

Disaggregated data can also be used *to advocate for specific policy changes*, to provide evidence for *targeted funding opportunities*, and to look for *patterns over time* and see if similarities or differences within and among subpopulations are emerging. For example, a 1998 Canadian study found that over 90 percent of suicides in First Nation populations were occurring in just 10 percent of First Nation communities in British Columbia (Chandler & Lalonde, 1998). Without disaggregating the data by community, this critical piece of information could have easily been missed. Resources could have been spent too broadly or not focused on the root causes of this discrepancy. Instead, this information

allowed researchers to obtain specific funding to look into what factors were contributing to these substantial population differences. Their results showed that a high level of self-determination was found to significantly reduce a communities' suicide rate (National Collaborating Centre for Aboriginal Health, 2009).

An evaluation specialist is also a valuable resource in this area. They can help to determine what data sets are important, the best way to collect data, and then can assist in analysis and disaggregation.

## Common Areas to Disaggregate

Choosing what data to disaggregate largely depends on the question you are trying to answer about your population and the type of data you have collected. Common characteristics used to disaggregate data include (Boeke, 2012):

- Race/ethnicity (country of origin)
- Generation status (i.e. first, second, etc. generation or recently arrived)
- Immigrant/ refugee status (refugee status often means people are eligible for certain services)
- Age group
- Gender
- Grade
- Geographic (within a state there is often enough data to compare school district data versus a state comparison to a national average)
- Sexual orientation
- Free or reduced lunch status (as a SES indicator)
- Insurance status

## Limitations of Data Disaggregation

Beyond the budgetary and expertise constrictions that many schools now face, there are limitations to what data can be collected, and thus, how data can be analyzed. A big limitation is low statistical power related to small sample sizes when you start disaggregating data. Statisticians from the National Evaluation Team caution that power analyses should be conducted on sample-based data sets, and in the absence of such analyses these data should not be disaggregated further than a cell size of 20 (e.g., if data from a sample size of 70 are disaggregated by race, and there are 20 nonwhites and 50 whites, then that might be okay; but if there are 10 nonwhites and 60 whites, then any conclusions may be misleading. The chance that those 10 nonwhites over-represent a variable of interest compared to the true value of that variable in the nonwhite population is too great). Common limitations to disaggregating data include :

## To protect individual student privacy

Example: A school administers a student survey that collects demographic information on race/ethnicity. The survey items also ask about previous contact with child welfare. If there are two white students in fourth grade and one reported case of previous contact with child welfare by a student who self-reported as non-Hispanic white, it would be easy for someone reviewing those results to identify the student, thus violating the student's privacy.

## Small numbers make it hard to view trends

Example: When evaluating a five-year grant program it would be hard to see true trends when combining three of the five years as a subpopulation. The differences in years could be big enough to misguide what is actually happening by chance or due to program implementation.

## Different data sources do not use the same definitions or break downs

Example: One survey may identify youth by ages 18-24, whereas another would include 18-25 year olds. This could also result from a lack of awareness of visibility of potential significant sub-populations.

## Conclusion

Disaggregating data is important to reveal patterns that can be masked by larger, aggregate data. Looking specifically at sub-populations can help make sure that resources are spent on the areas and students where they are most needed and can have the biggest impact. Perhaps most importantly, disaggregated data can help to make wiser future implementation decisions and secure targeted funding as you work to sustain SS/HS practices.

## Resources

### [Data-Driven High School Reform: The Breaking Ranks Model](#)

Chapter Five: Data-Driven Reform in Low-Performing High Schools

### [Improving School Board Decision-Making: The Data Connection.](#)

Chapter Three: [What is disaggregated data?](#)

Sample Resource Mapping Websites:

[Mapping the Measure of America](#)

[Diversity Data](#)

[Data Resource Center for Child & Adolescent Health](#)

[U.S. Census Interactive Map](#)

[Kids Count Data Center](#)

## [Children's Defense Funds Fact Sheets](#)

Safe Schools/Healthy Students Resources by Topic: [Demographic Data](#)

Safe Schools/Healthy Students Resources by Topic: [Cultural and Linguistic Competence](#)

Safe Schools/Healthy Students Resources by Topic: [Evaluation](#)

Safe Schools/Healthy Students Resources by Topic: [Sustainability and Financing](#)

## References:

1. Asian & Pacific Islander American Health Forum. *Disaggregation of Data: Needs of Challenges for Collecting and Reporting Race/Ethnicity Data*. (August 20, 2009). Webinar. Suhaila Khan and Marguerite Ro.
2. United States Department of Health and Human Services. HHS Action Plan to Reduce Racial and Ethnic Health Disparities: A Nation Free of Disparities in Health and Health Care. Washington, 2012. Retrieved from [http://minorityhealth.hhs.gov/npa/files/Plans/HHS/HHS\\_Plan\\_complete.pdf](http://minorityhealth.hhs.gov/npa/files/Plans/HHS/HHS_Plan_complete.pdf).
3. Chandler, M., Lalonde, C. (1998). *Cultural continuity as a hedge against suicide in Canada's First Nations*. *Transcultural Psychiatry*, 35(2): 191-219.
4. National Collaborating Centre for Aboriginal Health. (2009). *The Importance of Disaggregated Data*. Child & Youth Health.
5. Melissa Boeke. Personal communication, January 30, 2012.