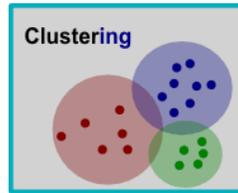Using Cluster Analysis
By Rhonda Williams, Ph.D.

I am Rhonda Williams, a Research and Evaluation Analyst for Region 10 Education Service Center, where I provide internal program evaluations in addition to external research and evaluation services for our partners. As a new evaluator, I search for ways to utilize different analysis beyond what I was exposed to in my Ph.D. program. While reading a report from the Brookings Institute entitled *Meet the Out-of-work*, I began to think about this idea of clustering to understand and describe data could be used in my work.

Most social science research is situated in terms of demographic data, such as socio-economic levels, census tract and zip code and more recently GIS data that helps to identify metrics like the distance to major hospitals or grocery stores. In education research, I chose the following variables: student achievement levels, ethnicity, district size and socio-economic levels. These parameters were publicly available about school districts and could be used to create clusters to better understand school districts.

For each variable that determines your clusters, you must first identify cut scores or boundaries. These decisions can be based on literature that supports the classifications or best practices of your field. You, as the expert, know and understand that there are natural categories and levels from experience. Decisions can also be based on your data which naturally illuminates breaking points; those points can be used to describe above average, below average, quartiles or even quintiles.

When examining large sets of data, software programs can help you find these cut scores per variable. After you have made the decisions to support your clusters you can begin to assign labels for each parameter; your clusters are then formed by examining your cases with similar markers. You will also notice the commonalities between your cases. For example, a cluster was formed from the following: school districts whose enrollment was less than 10K, SES levels between 30% and 50%, with student achievement averages higher than 75%. You can also use an additional variable to determine differences within clusters to help identify minor cluster or subgroups of the major clusters.

By using clusters, it allows you a way to examine groups in context with others that may be similar across the state or nation who might have similar programmatic needs. You can use clusters to define and describe patterns in relation to other programs and services evaluated. It is a more sophisticated way of looking at data; without cluster analysis, you might be prone to place your school district, cases or groups into one area using one at most two variables. Clustering allows you to describe and explain groups using multi-dimensions as a more real-world way to make stronger recommendations.