In 2001, No Child Left Behind (NCLB) led to a profusion of student achievement data, available for most grades and every year and representative of nearly the entire population of students. Unfortunately, much of these data come in the form of discrete “proficiency”-type categories, such as “Basic,” “Proficient”, and “Advanced”. Because states determine their own definitions of proficiency, making comparisons across states and time using differences in groups’ proficiency rates is problematic. In this paper we describe and test methods of computing achievement gaps from censored, proficiency-type data. In the first part of the paper, we develop a non-parametric framework for estimating achievement gaps, and use simulations and actual student-level NAEP data to assess the validity, bias, and precision of these methods. We show that our methods yield precise and largely unbiased estimates of achievement gaps.

In the second part of the paper, we address two practical concerns: corrections for measurement error and estimation of standard errors. When test scores cannot be assumed to be interval-scaled, the effects of measurement error are not straightforward. Drawing on previous work investigating the scale-sensitivity of reliability, this paper presents alternative strategies for disattenuation depending on the parametric defensibility of the reliability estimate. Likewise, if only censored data are available, the standard formula for the standard error of a standardized gap measure, such as Cohen’s $d$, will not provide appropriate standard errors for our non-parametric gap statistic. We show that standard errors computed under a maximum likelihood algorithm accurately estimate the sampling variance of the gap statistic across a wide range of data generating conditions. We conclude that it is feasible to reliably estimate achievement gaps and their standard errors from NCLB-type proficiency count data.