Use of observational measures to monitor preschool quality is growing rapidly. Although a large body of research has examined the validity of classroom observation tools within the context of researcher-conducted studies, little research to date has examined the extent to which the observations conducted as a part of state accountability efforts correspond to observations collected by research teams. This paper examines the degree of agreement between local and research rater teams using an observational measure of preschool classroom quality. It also explores the extent to which ratings predicted gains in children’s literacy, math, and self-regulation skills. Local ratings were conducted as a part of Louisiana’s quality rating and improvement system. Both rating teams observed 85 classrooms offering publicly funded preschool programs using the Pre-K CLASS, and 820 children from these classrooms (average age = 52.6 months, SD = 3.6 months) were directly assessed in the fall and spring. Results indicated correlations between local and research teams’ scores on corresponding domains, ranging from $r = .21$ to $.43$. Both teams’ scores were significantly but modestly related to children’s learning gains, although patterns of association differed. Results are discussed in the context of policies that require observational measures at scale.