



WIDATM

RESEARCH BRIEF

JUNE 2015

Evaluating Teacher Effectiveness Using ACCESS for ELLs

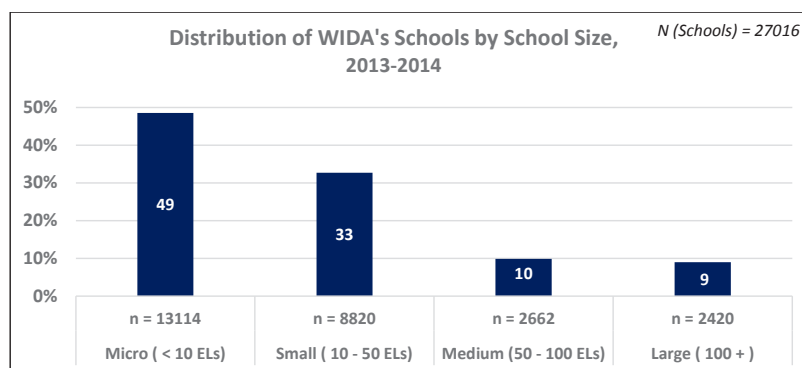
Introduction

Since the adoption of the Growth Model Pilot Program (GMPP) nearly a decade ago, estimation of growth models for evaluation and accountability purposes has been gaining traction across the country.¹ Advances in statistical methodology, coupled with modern computing power, allow for rich and flexible modeling of many factors potentially underlying students' performance on standardized tests. Among some of the more popular growth models currently used by districts and states to estimate growth are Value-Added Models (VAMs) and Student Growth Percentiles (SGPs). Both models use sophisticated statistical techniques and large samples of test score data to estimate individual students' growth. The output of VAMs and SGPs are then used for a variety of evaluative purposes. Despite the increasing popularity of these models, the literature on educational measurement and evaluation continues to suggest using caution when "high-stakes" accountability decisions are based on these measures of student growth. The issue is, regardless of the growth model, aggregate test-score-based models of student growth require large and longitudinally connected samples of student data. When sample sizes are small it becomes impossible to reliably estimate and disentangle district, school, and teacher effects from student growth data.

Small Samples

Small samples are of particular concern in the context of special student subgroups like English Learners (ELs). High student mobility, frequently interrupted education, and complex teacher–student linkage are among some of the EL-specific factors exacerbating the problem of insufficient sample sizes. Figure 1 presents the distribution of WIDA's 27,016 schools by size (classified as "Micro," "Small," "Medium," and "Large") as determined by the number of ELs with connected ACCESS for ELLs (ACCESS) Composite Scale Scores across the 2013–2014 ACCESS test administrations.

FIGURE 1: WIDA's schools by size category, 2013–2014.



¹ The GMPP report is available online at: <http://www2.ed.gov/rschstat/eval/disadv/growth-model-pilot/gmpp-final.pdf>

Figure 1 shows that about half of WIDA's schools had fewer than 10 enrolled ELs with connected ACCESS for ELLs test growth scores. An additional third of the schools that we label as "Small" had fewer than 50 enrolled ELs, and many of these ELs are further nested across multiple teachers. Regardless of the selected growth model, for a large majority of WIDA's schools (about 8 out of 10), the statistical estimation of both teacher- and school-effects would be based on small samples of ELs, resulting in high imprecision and cross-year instability in the growth estimates.

Attribution

In addition to small sample sizes, the attribution of growth effects becomes challenging with respect to assigning EL teachers' contributions to students' growth. Difficulties arise in defining how a single teacher affects student growth, given the complex and often varying assignment schedule of teachers to ELs. The inherent diversity within the EL population that places many students in unique paths to proficiency further complicates model estimation. ELs vary in their starting English language proficiency and age, prior education, home language proficiency, and types of the instructional programs they are enrolled in. Even when it is technically possible to match a group of EL students with a single EL teacher, it is overly simplistic to ignore the effect of all these potentially confounding factors, which in many cases are outside of the schools' or teachers' control. Further, is it plausible to assume that only one teacher contributes to students' academic English language development, dismissing influence of other educators or student involvement in other school or academically related activities? Attributing EL students' growth in language proficiency development to an individual teacher is fraught with complexities, which can only add to the imprecision and instability of growth effects.

Conclusion

Issues with small samples and teacher-student attribution introduce a multitude of problems, casting doubt on the accuracy and stability of growth estimates based on aggregated EL student test scores. WIDA recommends that educational agencies NOT make high-stakes decisions using growth models based on ACCESS test scores unless the issue of sample size and attribution has been overcome.

Prepared by Narek Sahkryan, MS, Wisconsin Center for Education Research, University of Wisconsin-Madison
For more information contact: help@wida.us

Suggested citation:

Sahkryan, N. (2015). *Evaluating teacher effectiveness using ACCESS for ELLs* (WIDA Research Brief). Madison, WI: Board of Regents of the University of Wisconsin System. Retrieved from WIDA website: www.wida.us/downloadLibrary.aspx

To learn more about various growth models and the details behind their application for evaluation and accountability purposes, we recommend:

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., ... & Shepard, L. A. (2010).

Problems with the use of student test scores to evaluate teachers. (EPI Briefing Paper #278). Washington, DC: Economic Policy Institute.

Castellano, K. E. & Ho, A. D. (2013). *A practitioner's guide to growth models*. Washington, DC: CCSSO.

Goldhaber, D., Walch, J., & Gabele, B. (2014). Does the model matter? Exploring the relationship between different student achievement-based teacher assessments. *Statistics and Public Policy*, 1(1), 28–39.

Holdheide, L. R., Goe, L., Croft, A., & Reschly, D. J. (2010). *Challenges in evaluating special education teachers and English language learner specialists*. (Research & Policy Brief). Denver, CO: National Comprehensive Center for Teacher Quality.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. (Monograph). Santa Monica, CA: RAND Corporation.