



Introduction and Purpose.

With NCLB, educators are expected to back decisions with evidence and are held responsible for results. Ideally, decisions are based on scientific research, yet most principals lack the facilities or staff for rigorous studies. MeasureResults is a web-based system designed to assist school and district administrators to <u>design and execute scientific</u> evaluations. It builds the appropriate design and analytical techniques into a simple framework that includes a web-based interface, automatically generated reports, and technical support featuring expert, in-house review of all analyses and reports.

Overview of the Web Site.

MeasureResults is an interactive web site whose basic flow of operation is shown below.

Figure 1: MeasureResults System Flow



Generate report

- After initial interaction with us to qualify the study for MeasureResults, the user signs on and creates a <u>project</u> representing the desired evaluation.
- In the setup phase, the user answers several questions about the study.
- MeasureResults then selects a study type based on the responses. The study type specifies the kinds of data that must be obtained and the method(s) of analysis that may be appropriate.
- Next, MeasureResults makes a data request for the user to fulfill. The contents of the data request depend on the selected study type.
- After the user fulfills the request, MeasureResults verifies the dataset for correctness.
- The engine performs the analysis determined to be appropriate.
- MeasureResults writes a report and delivers it to the user.

with interaction typical MeasureResults appears to the right. The entire process is scripted through the use of project templates written in XML. Thus we can easily change the behavior of the system and provide customized processing for classes of users.

Next Help

MeasureResults: A Tool for Data-Driven Decision Making for Schools Presenter: Robert Smith

Data Collection and Statistical Analyses. Data Collection

Typically we request data at the level of individual students assigned to classes or teachers. Required data generally include identification, standard NCLB demographics, and standardized test results.

MeasureResults generates data requests in the form of Excel documents with accompanying instructions and explanations to help district staff generate the data. After uploading the completed response, MeasureResults verifies the dataset and passes it to the analytical engine.

Statistical Analyses

We are currently implementing several study types based on specific experimental or quasi-experimental designs. Each of these several parametric variations depending on the exact information obtained in the setup process.

- how many variables, and what kind of matching process to use with multiple possible comparison groups. We presently us propensity score, Mahalanobis distance, and optimal matching for this purpose (Caliendo & Kopeinig, 2005; Gu & Rosenbaum, 1993; Michalopolous, Bloom, & Hill, for small control groups and optimal matching with non-replacement for large control groups.
- study and their interest in the results.

Reports.

The MeasureResults report processor module automatically generates study reports using a report template together with output from interactions with the user and output from the statistical analysis. This process creates graphs and tables and incorporates both template text and graphics together with generated data. The figure below illustrates the process of filling out templates into a page of the completed report. During the QC process, a reviewer checks the report and makes decisions about template selection and the need for additional text, selecting among different proposed formulations of conclusions and cautions.

		ROSS SCHO	MOR UNIFI	ED E1 CT PI	fectivene ogram fo	ss of <i>On</i> r Elemen	<i>Our Way</i> tary Engli	<i>to English</i> a ish Learners
	_		antiala (OLA/E) a		ala di la casa uni	Ma	ay 2005	
		to help elementary schools. It was co	y students learr nducted during	a single schoo	gby/Harcourt, peak English. I year: 2003-2	This is the firs	t study of OW	/E in Rossmor USI
<heading text=""></heading>				St	udy Descri	ption		
<date></date>		Research question	ons.					
		1) Did eleme California programs	entary school E English Langu already in plac	nglish learners age Developme e?	in classrooms ent Test than	using OWE a comparable st	achieve higher udents who co	r scores on the ontinued using the
Study Description		2) Did the impact of OWE vary according to whether students were in bilingual programs or in immer programs?					rams or in immersi	
		3) Did the im	pact of OWE v	ary by students	s' level of Eng	ish proficiency	y as measured	d by the pretest?
EXT>		Research design groups by coin tos approximately bala	 Randomized ss via a mechar anced between 	experiment. Te nism that assur conditions.	eachers who v ed that grade	olunteered we levels and bili	ere assigned t ngual and imr	o <i>OWE</i> or control mersion classes we
		Participants. The second- and fourth were missing for 6 <i>OWE</i> and control 9 <i>OWE</i> . Two teache were not used in th sample divided be	e selection and h-grade classro 38 students. The groups. Two te ers (one <i>OWE</i> to he analysis beo stween Bilingua	randomization oms, including e number of stu achers left the eacher in the ir cause none of t I and Immersio	procedure res six bilingual o udents with mi district or were nmersion grou heir students n classrooms	sulted in an ini lasses and 21 ssing pretests e reassigned. Ip and one con had pretest so	tial sample of immersion cl were roughly Both replacen ntrol teacher in cores. Table 1	379 students in 29 lasses. Pretest sco v equal between the nent teachers used n the bilingual grou shows the analysi
		Table 1. Particip	ants in Califor	nia OWE and	and Control Groups		Total	
		implementation	Number of teachers	Number of students	Number of teachers	Number of students	Number of teachers	Number of students
		OWE	5	60	9	88	14	148
		Control	3	49	10	119	13	168
		Total	8	109	19	207	27	316
<ack text=""></ack>								
		This research	ch was condue nat provided a	cted in 2003-2 Is a MeasureR	004. The resu esults™ serv	ilts are prese vice by Empir	nted to illust	rate the
			and the reason of	www.empir	caleducation	.com		
	I. I							

Funding for this project is provided by the Institute of Education Sciences, US Department of Education, SBIR contract# ED-08-CO-0046

References

Caliendo, M., & and Kopeinig, S., (May, 2005). Some Practical Guidance for the Implementation of Propensity Score Matching. Discussion Paper No. 1588. The Institute for the Study of Labor (IZA) in Bonn. Gu, X. S., & Rosenbaum, P. R., (1993). Comparison of multivariate matching methods: Structures, distances and algorithms. Journal of Computational and Graphical Statistics, 2, 405-420. Holdzkom, D. et al (2008) Extending Value-Added Analyses to Promote Effective Teaching Grades 3-12. Symposium presented at the annual meeting of the American Educational Research Association, New York. Michalopolous, C., Bloom, H. S., & Hill, C. J., (2004). Can propensity-score methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs? The Review of Economics and Statistics, 86, 156-179. Shadish, W. R., Cook, T. D., & Campbell, D. T., (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

• Interrupted time series (ITS): The first pilots will employ ITS. Following Bloom (2004), this model spans several years, including data from before the intervention was added, and data from the period. ITS is particularly useful in cases where an intervention is put into effect throughout a given school, and the question is "Do we see a change at this school?" after the fact. Subtasks include the development of various types of content: project template, data request, report template, and associated technical specs. (Appendix 2 of our first-year report contains a detailed discussion of the ITS study type.)

• Comparison group study: An increasing focus of our work consists of cases where we must find a well matched comparison group from other available units (Shadish, Cook & Campbell, 2002). In such cases it is critical to eliminate biases. Questions we face are what variables we use to make the match,

2004; Shadish, Cook, & Campbell, 2002). We calculate propensity scores and Mahalanobis distances, and then use optimal matching with replacement

• Randomized control trial (RCT): Experiments in which units such as teachers are assigned to a "treatment" condition using a randomization scheme are considered the best means for eliminating selection bias. We have conducted dozens of such experiments, automating many of the statistical processes called for. We recognize that, while RCTs entail a level of advanced planning that is unusual for schools (in that most evaluations begin after the participants have been identified) we also know from experience that the process of random assignment itself can improve educators' commitment to a





Deployment / Evaluation.

Several evaluations are scheduled for the next few months.

Texas Instruments Math Forward

We are currently implementing a customized version of MeasureResults in the schools of four districts that are evaluating a math technology developed by Texas Instruments. Within the next few weeks, customers will be uploading their data to MeasureResults. We will monitor the steps carefully both to make certain that the process goes smoothly for these users and to detect issues with the MeasureResults system.

Poway School District

We have also begun a partnership with Poway Unified School District (PUSD), in which we are granting them a one-year subscription license to MeasureResults. PUSD has agreed to pilot MeasureResults and to provide feedback on user interface, product features and capabilities, report content and clarity, and overall functionality and usability. We will collect feedback via surveys, email, phone, and/or in-person interviews.

We are planning to conduct an RCT to test the effectiveness of MeasureResults itself at the end of the study.

Marketing Outreach.

A key task in our messaging is to educate school administrators about the kind of evidence that can be provided through statistical processes generally lacking from current school data systems. With the rapid expansion of data systems, driven in large part by mandated reporting requirements, many technology-savvy educators can already produce descriptive reports and graphs depicting raw achievement trends.

MeasureResults can go well beyond the group average achievement outcomes provided by most data systems as standard queries. With MeasureResults, educators can examine program impacts at the school and classroom levels, controlling for student and teacher characteristics. Making full use of longitudinal data systems goes beyond producing descriptive graphs. Statistical processes allow for the projection of trends, the determination of impact, and the measurement of teacher performance, controlling for environmental factors and background characteristics.

Next Steps

- Complete systems to accommodate all study types.
- Create new project template "skins" and report templates.
- Deploy on-line teacher surveys with automatic integration of information about implementation integrity.
- Evaluate effectiveness of the product using an RCT.