Using Microanalytical Simulation Methods in Educational Evaluation: An Exploratory Study

Toni A. Sondergeld
Bowling Green State University

Svetlana A. Beltyukova, Christine M. Fox, and Gregory E. Stone
The University of Toledo

Scientifically based research used to inform evidence based school reform efforts has been required by the federal government in order to receive grant funding since the reenactment of No Child Left Behind (2002). Educational evaluators are thus faced with the challenge to use rigorous research designs to establish causal relationships. However, access to student-level longitudinal or comparison group data is often scarce, which significantly restricts researchers' choice of research design. Although most state departments of education have school- and district-level data available to the public, the individual student-level data that are often needed to perform appropriate statistical analyses are unavailable. This exploratory study demonstrates the process of and provides evidence that microanalytical simulation methods, conducted using Microsoft Excel, may be a useful research tool in the field of education if adequate school-level modeling information is available. These simulation methods may assist in providing greater opportunities to execute more rigorous methodological designs.

With the onset of No Child Left Behind, the bar has been raised for evaluations of federally funded educational programs. Mandates have been created for schools "who depend on federal funding to select and implement programs that are based on scientific research" (Beghetto, 2003, p. 1). This means that researchers cannot validly claim any educational reform program effective if its evaluation is not based on experimental or quasi-experimental methods (Crowley & Hauser, 2007). For this reason, the evidence of effectiveness of educational programs, especially Comprehensive School Reform (CSR) programs, has been highly scrutinized by educational researchers who argue that the lack of rigorous research designs has made it virtually impossible to infer program effects on student achievement (see, for example, Borman, Hewes, & Overman, 2002; Crowley & Hauser, 2007; Slavin, 2002).

As a result of the call for scientifically based educational research and the lack of consensus as to what scientifically based research means, the Comprehensive School Reform Quality (CSRQ) Center established standards for evaluating CSR programs. These standards are aligned with the American Psychological Association's recommendations for evaluation of research design and analysis (APA, 2002) and American Educational Research Association's process for reporting quantitative analyses (AERA, 2006), and focus on six primary research criteria: design, assessment, timing, sampling, program implementation, and data analysis (Crowley & Hauser, 2007). To meet the CSRQ Center's research and evaluation criteria, designs should either be randomized controlled trials or quasi-experimental (pre- and post-test nonequivalent group; regression discontinuity; and cohort or single-group longitudinal designs). Assessment refers to establishing face validity of student achievement outcomes. Timing indicates that researchers must have a baseline measure of student achievement when using a comparison group and in addition to this use at least two follow-up measures if no viable comparison group exists. With regard to sampling, control groups must either be "business as usual" type schools or schools at least not undergoing the same CSR program as the treatment group. Further, adequate group equivalence on pre-test assessment measures with matching of key demographic variables is important. Implementation information (e.g., how long the program has been running and fidelity to the program curricula and goals) is imperative. Finally, data analysis should employ any statistical technique necessary to correct for group nonequivalence. Although these rigorous evaluation standards have been established, implementing them is not as straightforward as following the guidelines listed above because the necessary data to conduct such evaluations are not always available to researchers.

Availability of Data

To conduct rigorous evaluations of school reform efforts as suggested by the CSRQ Center, great amounts of data are needed on numerous variables across time. Expansion and ease of school, district, and state level data access are providing researchers with new and improved means of obtaining these necessary data. Today, researchers can easily retrieve longitudinal or potential comparison group data at the school or district level. Recently created state databases of longitudinal achievement, attendance, discipline, and enrollment data at the school, district, and state levels are readily available to the public at most state department of education websites.

With collaborative efforts such as the Data Quality Campaign (DQC), managed by the National Center for Educational Achievement (NCEA) and funded by the Bill and Melinda Gates Foundation, states are being provided with support for creating high quality longitudinal educational data systems. The mission of the DQC (2008) is to encourage and support state policymakers to advance the collection, availability, and use of high-quality education data and to implement state longitudinal data systems to improve student achievement. This campaign aims to provide tools and resources that will assist development of state-level quality longitudinal data systems, while providing a national forum for reducing duplication of effort and promoting greater coordination and consensus among the organizations focusing on improving data quality, access and use (About Us section, ¶ 1).

The DCQ (2010) generated 10 elements critical to the establishment of a quality PK – 12 longitudinal educational data system. They are as follows:

1. A unique statewide student identifier that connects student data across key databases across years;
2. Student-level enrollment, demographic, and program participation information;
3. The ability to match individual students' test records from year to year to measure

academic growth;
4. Information on untested students and the reasons they were not tested;
5. A teacher identifier system with the ability to match teachers to students;
6. Student-level transcript information, including information on courses completed and grades earned;
7. Student-level college readiness test scores;
8. Student-level graduation and dropout data;
9. The ability to match student records between the PK–12 and higher education systems; and
10. A state data audit system assessing data quality, validity, and reliability.

From a recent survey of all 50 states, Washington D.C., and Puerto Rico, the DCQ (2010) found that 12 states met all 10 of the above listed elements critical to the establishment of a quality PK – 12 longitudinal educational data system, 22 states met eight to nine elements, 17 states met six to seven elements, and 1 state met only four to five elements. These numbers have dramatically increased since the DCQ began surveying states in 2005. Table 1 illustrates the movement toward more robust longitudinal educational data systems by showing the number of states (including D.C. and Puerto Rico) that have implemented each of the essential elements for a longitudinal data system in 2005 and in 2009. Florida is one state leading the way in not only collecting longitudinal student and teacher level data, but also in making them available to the public as well (Hood, 2007). However, states such as Ohio more commonly will collect student and teacher level data but only provide school, district, and state level data to researchers (or the public), even though student and teacher identifiers do not identify individual students (M. Mottley, personal communication, September 12, 2008).

Table 1

Number of States Implementing DCQ's 10 Essential Elements of a Longitudinal Data System Comparison from 2005 – 2009

	Year	
Element	2005	2009
1. Statewide student identifier	36	50
2. Student-level enrollment data	38	51
3. Student-level test data	32	50
4. Information on untested students	25	47
5. Statewide teacher identifier	13	24
6. Student-level course-completion/transcript data	7	23
7. Student-level SAT, ACT, and Advanced Placement data	7	36
8. Student-level graduation and dropout data	34	51
9. Ability to match student-level P-12 and higher education data	12	33
10. State data audit system	19	51

Unavailability of student-level data becomes problematic for those researchers who need individual-level data to answer their research questions and make their desired statistical analyses possible. All educational evaluators/researchers certainly do not face this problem of lacking student-level data because they plan for collection of or have funding to collect baseline or comparison group data built into their evaluation/research plan. However, for those who do not have access to student-level data, aggregate school or district level data become useful for looking at the overall picture or answering descriptive research questions. These data may also be helpful if a researcher is using the school or district as the unit of analysis (Slavin, 2008). Yet such data do not serve researchers well if they intend to address research questions in one or a few schools at the student-level. For instance, aggregate data provided by most states allow evaluators to describe general trends over time or in comparison to other schools. However statistical comparisons over time cannot be made if student-level data are not possessed. More specifically, if an evaluator is not working in conjunction with a school's data center, the evaluator may not have access to and thus not be able to investigate statistical differences between program participants and non-participants to study a program's impact statistically or quantify effect sizes. Even if working with a school district's data center, there are often difficulties in acquiring the needed student-level data because the district does not want to release detailed data (School Communities That Work, 2008). Or if dated data (pre-2000) are being requested, obtaining them may pose a great challenge or be impossible because older data in some school districts are only available on tapes in antiquated data storage systems. Therefore, alternative methods for obtaining student-level data need to be sought in order to answer research questions requiring data at the student-level.

Simulating Student-Level Data

Simulation as a research methodology is a type of modeling that has been used in the social sciences to better understand the world and "predict the values of dependent variables" (Gilbert & Troutzsch, 2005, p. 2). Similar to statistical modeling, researchers using simulations enter inputs about a system into a model and observe computer-generated outputs (Axelrod, 2005; Gilbert & Troutzsch, 2005). Although simulation has been a viable methodological approach in the social sciences for nearly half a century, its use has only begun to widely expand in the field over the last 15 years (Axelrod, 2005). "Virtually all disciplines of the social sciences, including anthropology, business, economics, human evolution, environmental planning, law, information, organization theory, political science, and public policy" (Axelrod, 2005, p. 2) have to some degree adopted this approach to research and used it for gaining better understanding of the social world or for predicting behavior over time. The field of education, however, is not listed among these social science fields and has rarely used simulation methods as a research tool (Axelrod, 2005) other than for theoretical studies in statistics and measurement.

Microanalytical simulation models, also referred to as microsimulation, in particular may be of great use to educational researchers who have aggregate school level data but actually need student-level data for statistical analyses. Microanalytical simulation models focus on "the individual-level, modeling individual persons with a number of attributes (e.g., sex, age, marital status, education, employment)" (Gilbert & Troitzsch, 2005, p. 58). Information is collected from a representative sample of the population on the dependent variables of interest, and individual-level data are created based on the model for the hypothetical sample. For this type of simulation, "all probabilities applied in the model have to be calculated from empirical data" (Gilbert & Troitzsch, 2005, p. 63) based on the specific sample being studied, yet actual population attribute probabilities are often difficult to obtain if they exist at all.

Dependence on individual-level information at every stage of the simulation analysis is what distinguishes microanalytical simulation models from other forms of simulation modeling (Mitton, Sutherland, & Weeks, 2000). Educational researchers in the United States are at an advantage in accessing attribute probabilities needed for microanalytical data simulation, as NCLB has instituted requirements for making longitudinal school and district level data available to the public, and agencies are working to make this possible. Therefore, the empirical probabilities needed to simulate student-level data for dependent variables, such as student achievement, by key demographic factors shown to impact student achievement such as economically disadvantaged status, race, and special needs status (U.S. Department of Education, 2004), may be obtained through state databases to conduct microanalytical data simulation.

Research Purpose and Questions

The purpose of this study was to assess the feasibility of using microanalytical simulation methods in educational research when individual-level data are unavailable, and develop specific procedural directions for others to follow to conduct similar research. Specifically, this study evaluated the validity of inferences about student-level data that were simulated from Ohio Achievement Test (OAT) school-level data by examining comparability of the results between *simulated* student-level achievement data (based on the parameters of the state-reported school level data) and *actual* student-level achievement data. Specifically, data comparability was assessed with regard to percent proficient statistics and student test scores by selected demographic characteristics.

Data used in this study were from a longitudinal CSR evaluation study of an urban junior high school in Ohio. Practical need is what drove our desire to find an alternative method to obtaining student-level data since our evaluation team ran into great difficulty acquiring the longitudinal student-level data we desired from the school district being evaluated. Although in the larger study our goal was to simulate multiple years of data for varying grade levels to statistically assess intervention and comparison group achievement outcomes, the focus of this study was on validating our microanalytical simulation approach. Thus, we only evaluated one year of one 7th grade cohort in the intervention group. The following research questions were addressed:

1. How does the percentage of students who are determined to be percent proficient compare between compare between the state reported aggregate data obtained from the Ohio Department of Education's (ODE) website and the simulated student-level data modeled after the state level data?
2. When actual student test scores are matched to simulated student test scores based on key demographic variables (i.e., economically disadvantaged status, race, and special needs status), is there a statistically significant relationship between test scores?
3. If statistically significant differences in actual student test scores exist by socioeconomic status, special needs status, or race, are the same differences reflected in the simulated student test scores?

Methods

Data Collection and Instrumentation

Data used in this study were from a longitudinal CSR evaluation study of an urban junior high school in Ohio. Ohio Achievement Tests (OATs) are given to middle grades students in Ohio each spring over the content areas designated for their specific grade levels (e.g., reading, mathematics, social studies, science, and/or writing). Content validity evidence for these assessments is high because they were created based on blueprints of the state content standards that Ohio public school teachers are required to teach in their classrooms. All OAT scores are classified into five levels of achievement. From highest to lowest they are: Advanced, Accelerated, Proficient, Basic, and Limited. Regardless of the grade level or content, a scaled score of 400 is the lowest OAT Proficient score needed for passing. Scores from individual subject area tests, however, vary in their range, mean, standard deviation, standard error of measure, and reliability indices. Table 2 provides detailed information on the 2007-08 test statistics resultant from seventh grade math and reading OATs. Table 3 illustrates the achievement level scaled cut score points for each test as they slightly vary between tests. Information for Tables 2 and 3 was obtained from the Ohio Department of Education website (www.ode.state.oh.us/) (ODE, 2008) and was used to inform the parameters in simulating student-level data.

Table 2

Yearly Ohio Achievement Test Statistics by Subject Area for Seventh Grade 2008

	Subject Area	
Statistic	Math	Science
N-count	133,556	133,907
Max Raw Score	50	48
Max Scaled Score	569	540
Min Scaled Score	275	267
Raw Score *M*	22.84	27.03
Raw Score *SD*	10.21	9.35
Raw Score *SEM*	3.35	3.24
Scaled Score *M*	416.95	418.45
Scaled Score *SD*	32.22	29.22
Scaled Score *SEM*	10.55	10.12
Reliability	0.89	0.88

Table 3

Achievement Level Scaled Cut Score Points by Subject Area for Seventh Grade

Subject Area	Limited	Basic	Proficient	Accelerated	Advanced
Reading	< 379	379	400	432	452
Math	< 378	378	400	436	458

All 2007-08 school level OAT math and reading school level results for seventh grade students in the urban junior high school of interest were retrieved from the ODE (www.ode.state.oh.us/). These school-level results were used as parameters for simulating student-level data. On the ODE website "reports are made available for more advanced users of the data who are interested in comparing multiple years of information for several schools or districts" (ODE, 2008, ¶1). Actual student-level achievement data for the 2007-08 school year were obtained from the school district of interest's data center for the purpose of comparing them to the simulated student-level data.

Sampling Method and Sample

A sample of 200 students was selected to be proportionally representative of the 2007-08 school demographics (economically disadvantaged by race and special needs). The stratified sampling was done to elicit a balanced sample matched on the key demographic factors shown to impact student achievement (U.S. Department of Education, 2004). This sample comes from a larger longitudinal cohort study where students from nine consecutive seventh grade cohorts in the school of interest were matched based on the above listed key demographic factors so academic and non-academic outcomes could be compared over time to examine the impact of a CSR program. Over the term of the larger study, the seventh grade cohorts ranged in size from 351 to 438 students (M = 391.56, SD = 31.49). Thus, a sample size of 200 students each year was selected to allow for accurate matching of cohort students from year to year allowing for slight enrollment size and demographic variation between yearly cohorts. As a result, we use a sample of 200 seventh grade students from the 2007-08 school year in this study. Table 4 details the frequencies and percentages of students that were selected based on the student demographic data obtained from ODE's online database.

Table 4

Students Selected for Study based on Demographic Factors Retrieved from ODE's Online Database (N=200)

Economically Disadvantaged	Race	Special Needs	*f*	%
N	Minority	N	20	10
Y	Minority	N	40	20
N	White, Non-Hispanic	N	43	21.5
Y	White, Non-Hispanic	N	63	31.5
Y	White, Non-Hispanic	Y	18	9
Y	Minority	Y	10	5
N	White, Non-Hispanic	Y	5	2.5
N	Minority	Y	1	0.5

Methodological Framework

With simulation methods earning their place among social science methodologies, researchers have created guidelines for conducting sound research of this type. According to Law and Kelton

(2000), there are 10 steps to follow in order to conduct a reliable simulation study. The following sequence of steps is modified slightly from Law and Kelton's (2000) text (pp. 84 – 86):

1. Formulate problem and plan the study.
2. Collect data and define a model.
    a. Collect data to specify model parameters and input probability distributions.
    b. Define the data with an "assumptions document," which is the conceptual model.
3. Validate the conceptual model.
    a. Perform a structured walk-through of the conceptual model using the assumptions document before an audience of managers, analysts and subject matter experts (SMEs).
4. Construct a computer program and verify.
5. Make pilot runs for validation purposes.
6. Identify program validity.
    a. Compare model measures to the existing system measures.
    b. SMEs and analysts should review the model results for correctness.
7. Design experiments.
    a. Specify length of run.
    b. Specify number of independent simulation runs using different random numbers.
8. Make production runs.
9. Analyze output data.
10. Document, present, and use results.

Simulation process specific to this study. To conduct this study, only Steps 1 – 6 of the above simulation process were used because the focus was on assessing the validity of the microanalytical simulation method. Thus, reporting on steps 7 – 10 with regard to experimentation, production runs, analysis of simulated data, and results presentation of the experimentation is unnecessary. This simulation study was conducted by implementing two main phases: conceptual planning and data validation. In the conceptual planning phase, Steps 1 - 3 of Law and Kelton's simulation study procedures were addressed. In Step 1, the problem was defined as the lack of availability of student-level data to the researchers. A plan was then developed to simulate student-level data based on the school level demographics and OAT data which were accessible through the ODE website.

In Step 2, disaggregated school level data were downloaded from the ODE website's public access portal and used to determine the parameters to guide how the individual student-level data would be simulated. The most challenging part of creating the model used in this or any simulation study was deciding what information to include and what to omit (Gilbert & Troitzsch, 2005). We decided to include only student-level demographic data that have been shown to impact student achievement including economically disadvantaged status, race, and special needs status (U.S. Department of Education, 2004). Models for the dependent variables (seventh grade math and reading student achievement) were generated to reflect student-level data that were proportionally allocated based on school level results for the key demographic factors. See Appendix A for a sample conceptual model based on demographically disaggregated achievement data obtained from the ODE website.

Once conceptual models for the simulation were created, they were validated in Step 3 through a structured walkthrough with subject matter experts (SMEs) to ensure all relevant factors were included in the conceptual models for each planned simulation model. Four SMEs were consulted: two were the external evaluators of the CSR program being studied and had been working on the project for seven years, and two SMEs were experts in high-stakes educational testing. In addition to providing information on the level of detail the model needed, SMEs also gave valuable insight on sample size and population parameters. One specific example SMEs were particularly helpful with related to population parameters. From the ODE website we knew the 2007-08 seventh grade math OAT scores in the state ranged from 275 – 569 (M = 416.95, SD = 32.22). Because our simulated data were based on the ranges of scores for each OAT category all middle OAT category ranges (Basic, Proficient, and Accelerated) were defined with distinct beginning and ending point values. However, the tails of the distribution were less defined as we knew where each value began but did not know the specific end values for our sample (i.e., Limited is and score < 379 and Advanced is any score ≥ 452). To determine appropriate end values for this population (urban middle school students) the SEMs assisted in conceptualizing the end-point parameters. It was decided that this population may likely score as low as three standard deviations below the state mean (placing the Limited end point at 320) and possibly two standard deviations above the mean (placing the Advanced end point at 481).

Data validation was the second phase of this study's simulation research process, and encompassed Law and Kelton's (2000) Steps 4 – 6. Microsoft Excel was used to construct a computer program for generating the student-level achievement data in Step 4. Although other simulation programs were available that could outperform Excel in terms of speed and simplicity in programming (e.g., Arena, AutoMod, Extend, ProModel, WITNESS), the decision to use Excel for simulating achievement data over alternative simulation programs was primarily based on the fact that Excel would be easily accessible and familiar to most researchers (Harnett & Horrell, 1998). Excel is commonly used in simulation research, and specifically has demonstrated its usefulness when conducting microanalytic simulations (e.g., Gohler & Geisler, 2012; Public Health Agency of Canada, 2006). While Excel add-ins for conducting microanalytic simulation have been developed (e.g., *Ersatz* created by Epi Gear International, 2012), microanalytic simulation can be conducted without using additional add-ins as shown in this study. Appendix B presents the step-by-step process of simulating data and the computer programming commands needed for simulating student-level achievement data from school level achievement data.

Using the simulation program that was created in Excel for this simulation process, pilot runs were performed in Step 5 to validate that the student-level data being simulated were in fact representative of the conceptual model (Appendix A), and the program was working as desired. This is also known as debugging. If the simulated data were not representative of the conceptual model, programming modifications would be needed and additional pilot runs would be conducted until the program worked correctly.

Finally, Step 6 assessed the program's validity. Program validation ensures that the behavior of the modeled data corresponds to the target system (real world phenomenon) (Gilbert & Troitzsch, 2005). However, "there is no completely definitive approach for validating the model of the proposed system" (Law & Kelton, 2000, p. 279). Simulation researchers instead examine the "closeness" of the simulated model data to the actual system data that exist in order to consider modeled or simulated data valid. For this study, two versions of the inspection approach were implemented: basic inspection and correlated inspection (Law & Kelton, 2000). The basic inspection approach uses one or more computed statistics from the observed system and descriptively compares this to the corresponding statistics from the simulated model output. For student achievement data in this study, this meant checking to see if the overall simulated percent proficient for each grade level on each OAT was similar to the corresponding state reported school level percent proficient. To do this, student-level data for the 200 students in seventh grade were simulated based on the state reported demographic and achievement parameters. Then the percent scoring 400 (minimum proficient score on all OATs) or greater were calculated from the individual-level simulated data.

Basic inspection, however, is vulnerable to innate randomness of observations in the simulated model and actual system because "each statistic is essentially a sample of size 1 from some underlying population" (Law & Kelton, 2000, p. 283). Therefore, the correlated inspection approach was also used for comparing model to system data. With the school district providing student-level achievement data from the 2007-08 school year, it was possible to implement the correlated inspection approach for this school year's achievement data. Correlated inspection of the spring 2008 reading and math OAT simulated and actual scores was performed for the sample of 200 seventh graders matched by key demographic factors. One-hundred simulated data runs were completed for each test to assess the program's ability to reproduce similar data regardless of the run. We share the results from the first five simulated data runs as they are representative of all simulated data runs conducted, and a goal of this study was to develop a program that would allow researchers to run one set of simulated data that should be representative of the system data rather than needing to average multiple runs. If a valid program were developed, running any one set of simulated data would be sufficient. Student-level system data (actual student OAT scores) were compared to the simulated student-level data through bivariate correlations for seventh grade reading and math achievement data, where the strength of the relationship between the simulated student-level data and the actual student-level data were assessed.

Results validation was a third method of assessing the validity of the simulated data. This validation included running multiple independent *t*-tests to look for potential differences in OAT scores based on special needs status (Regular Education vs. Special Education), race (White vs. Minority), and economically disadvantaged status (Non-Economically Disadvantaged vs. Economically Disadvantaged). If any statistical differences were found based on these factors in the actual (system) student-level data, it was expected that the same differences would be identified in the simulated student-level data for each simulated data run. Again, five simulated data runs are shared for independent *t*-tests to assess replicability. Results validation was used along with basic and correlated inspection approaches to review the models for correctness and validate the assumptions of the simulation models.

Data Validation Results

Results from validating the simulation program in Step 6 are presented in this section by research question. While all data were simulated using Microsoft Excel, simulated data were imported into SPSS 15.0 and this software was utilized for the actual Step 6 analysis.

How does the percentage of students who are determined to be percent proficient compare between the state reported aggregate data obtained from the ODE website and the simulated student-level data modeled after the state level data?

Through basic inspection, Table 5 shows that all simulated percent proficient students from the averaged five sample runs were very closely matched to the actual state reported percent proficient students for each academic subject. Simulated data in both cases were 0.1% below the state reported percent proficient illustrating the extreme closeness of the simulated data to that of the actual state reported data.

Table 5

Basic Inspection Validation for Seventh Grade Ohio Achievement Test Data

Subject Area	State Reported % Proficient	Simulated % Proficient (Five Runs Averaged)
Math	33.6%	33.5%
Reading	51.6%	51.5%

When actual student test scores are matched to simulated student test scores based on key demographic variables including economically disadvantaged status, race, and special needs status, is there a statistically significant relationship between test scores?

Correlated inspection results are shown in Tables 6 and 7 and depict the inter-correlations, means, and standard deviations for the actual student OAT scores and the five simulated runs of seventh grade math and reading OAT results from spring of 2008. Overall, regardless of the grade level or test, inter-correlations between the actual and simulated student-level data were all statistically significant ($p < .001$) and highly positive, where nearly perfect correlations averaged $r = .986$ and ranged from $r = .980$ to $r = .994$. The resultant average $r^2 = .972$ indicated that on average 97.2% of the variance in the actual scores was accounted for by the simulated scores. Additionally, within demographic groups, the intercorrelations between simulated data sets were all statistically significant ($p < .001$) and highly positive again where nearly perfect correlations averaged $r = .997$ and ranged from $r = .993$ to $r = .999$. The average $r^2 = .994$ indicated that on average one simulated run accounted for 99.4% of the variance in another simulated run.

If statistical differences in actual student test scores exist by socioeconomic status (non-economically disadvantaged vs. economically disadvantaged), special needs status (regular education vs. special education), or race (White vs. minority), are the same statistical differences reflected in the simulated student test scores?

Table 6

Intercorrelations, Means, and Standard Deviations for Actual and Simulated Seventh Grade Student-Level Reading OAT Scores from Spring 2008 (N=200)

Measure	Actual	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	*M*	*SD*
Actual	—						399.64	28.13
Sim 1	.993	—					399.39	30.62
Sim 2	.992	.999	—				398.89	30.18
Sim 3	.992	.998	.998	—			399.80	30.11
Sim 4	.990	.998	.997	.998	—		399.90	31.25
Sim 5	.994	.996	.995	.995	.993	—	400.89	29.07

Note. All correlation coefficients are statistically significant ($p < .001$).

Table 7

Intercorrelations, Means, and Standard Deviations for Actual and Simulated Seventh Grade Student-Level Math OAT Scores from Spring 2008 (N=200)

Measure	Actual	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	*M*	*SD*
Actual	—						392.13	22.62
Sim 1	.977	—					391.47	29.22
Sim 2	.981	.997	—				390.23	28.79
Sim 3	.980	.998	.996	—			390.59	27.74
Sim 4	.980	.998	.998	.996	—		390.53	28.56
Sim 5	.982	.998	.997	.998	.998	—	390.59	28.49

Note. All correlation coefficients are statistically significant ($p < .001$).

Tables 8-10 show that in the actual student-level data there were no statistically significant differences in seventh grade reading or math OAT scores by economically disadvantaged status, or for reading OATs by race. There were, however, statistically significant differences found in seventh grade reading and math OAT scores by special needs status ($p < .001$) with regular education students having statistically significantly higher scores than special needs students; and for math OATs by race ($p < .05$) with White students having statistically significantly higher scores than minority students. These same results were found in all five simulated student-level data sets. The only variation in results is that all simulated math data sets indicated a statistically significant difference between races at the $p < .01$ level rather than $p < .05$ level that the actual data revealed.

When comparing actual means and standard deviations for each subgroup in reading and math to simulated means and standard deviations a similar pattern was seen across tests. Simulated means were similar to actual means with differences ranging from .01 to 4.28 points different and averaged 1.04 points different with most simulated means being lower ($n = 44$; 73%) than actual means across tests. Simulated standard deviations were not as similar as they ranged from .86 to 11.43 points different and averaged 4.06 points different across tests. All simulated standard deviations were higher than actual standard deviations across subgroups and tests. When comparing means, standard deviations, and *t*-statistics of simulated data runs to the actual data, a single simulated data run did not turn out to be "the best" as the most similar values were spread across multiple simulated test runs. Further, because all simulated data runs resulted in similar outcomes as the actual data and no one simulated run appeared to be better than another, support is thus established for using any of the individual simulated data runs as representative of the system (actual data).

Discussion and Implications

Results from this study demonstrated that overall the simulated student-level data were highly representative of the state reported school level data and actual data. One reason we believe this to be the case is because implementing a well-researched model for completing a sound simulation study (Law & Kelton, 2000) increased the likelihood of success. Secondly, the ease of accessibility to school level data disaggregated by important demographic factors (economically disadvantaged status, race, and special needs status) for each of the outcome variables was invaluable in the building of a valid conceptual model for simulation. These very specific school level data allowed for the use of a microanalytical simulation model where individual-level data, or student-level data, were simulated based on the actual attribute probabilities found in the population (Gilbert & Troitzsvh, 2005). Without the specific probabilistic break-down of the outcome variables by key demographic factors, the conceptual models used for simulation would certainly have been less comparable to the actual system data, thus producing weaker models and less justification for simulated data validation.

Although overall results of this study were positive, the main weakness noted when comparing the simulated student-level data to actual student-level data was in terms of the standard deviation comparison. Across all validation tests run, the standard deviations from the simulated data were always higher than those from the actual data, going as much as 11.43 points greater. Although this discrepancy was found, it did not appear to impact the correlated inspection or validation results as all simulated and actual test results arrived at the same conclusions, and "the most definitive test of a simulation model's validity is to establish that its output data closely resemble the output data that would be expected from the actual (proposed) system"

Table 8

Seventh Grade Reading and Math OAT Score Differences Between Non-Economically Disadvantaged and Economically Disadvantaged Students (N=200)

	Non-Economically Disadvantaged (*n* = 69)		Economically Disadvantaged (*n* = 131)							
Data Source	*M*	*SD*	*M*	*SD*	Mean Difference	95% CI Lower	95% CI Upper	*df*	*t*	η^2
Actual Data – 7R	399.62	26.54	399.64	29.02	0.02	-8.29	8.25	198	-.01	< .01
Sim Data 1 – 7R	398.78	29.46	399.70	31.33	-0.92	-9.93	8.09	198	-.20	.02
Sim Data 2 – 7R	398.20	28.97	399.25	30.90	-1.05	-9.92	7.82	198	-.23	.02
Sim Data 3 – 7R	399.29	29.22	400.07	30.68	-0.78	-9.63	8.07	198	-.17	.01
Sim Data 4 – 7R	399.46	30.47	400.12	31.76	-0.66	-9.85	8.53	198	-.14	.01
Sim Data 5 – 7R	400.59	27.91	401.05	29.77	-0.46	-9.01	8.09	198	-.10	< .01
Actual Data – 7M	393.58	16.44	391.37	25.30	2.21	-4.43	8.85	189.35	.75	.05
Sim Data 1 – 7M	393.49	22.19	390.40	32.34	3.09	-5.49	11.67	184.49	.80	.06
Sim Data 2 – 7M	392.10	21.26	389.24	32.09	2.86	-5.60	11.32	187.72	.76	.05
Sim Data 3 – 7M	392.12	21.21	389.79	30.68	2.33	-5.82	10.48	183.73	.63	.04
Sim Data 4 – 7M	392.07	21.54	389.72	31.69	2.35	-6.04	10.74	185.40	.62	.04
Sim Data 5 – 7M	392.29	21.24	389.69	31.68	2.60	-5.77	10.97	186.68	.69	.05

Note. 7R=Seventh Grade Reading OATs, 7M=Seventh Grade Math OATs. For all Math data sets, equal variances were not assumed and Levene's correction was applied.

Table 9

Seventh Grade Reading and Math OAT Score Differences Between Regular Education and Special Education Students (N=200)

	Regular Education (n = 166)		Special Education (n = 34)							
Data Source	*M*	*SD*	*M*	*SD*	Mean Difference	95% CI Lower	95% CI Upper	*df*	*t*	η^2
Actual Data – 7R	405.05	26.83	373.18	17.40	31.87	22.40	41.34	69.64	8.76***	.58
Sim Data 1 – 7R	405.33	28.89	370.35	20.85	34.98	24.69	45.27	198	6.70***	.57
Sim Data 2 – 7R	404.70	28.49	370.53	20.90	34.17	24.01	44.33	198	6.63***	.56
Sim Data 3 – 7R	405.60	28.37	371.47	21.08	34.13	24.00	44.26	198	6.64***	.56
Sim Data 4 – 7R	405.93	29.39	370.44	22.19	35.49	24.98	46.00	198	6.65***	.56
Sim Data 5 – 7R	406.41	27.79	373.94	18.23	32.47	22.66	42.28	68.77	8.55***	.57
Actual Data – 7M	395.41	22.20	376.12	17.39	19.19	11.22	27.16	198	4.77***	.44
Sim Data 1 – 7M	395.42	28.60	372.18	24.43	23.24	12.87	33.61	198	4.42***	.40
Sim Data 2 – 7M	393.99	28.32	371.85	23.89	22.14	11.88	32.40	198	4.26***	.39
Sim Data 3 – 7M	394.32	27.19	372.38	23.10	21.94	12.08	31.80	198	4.39***	.40
Sim Data 4 – 7M	394.25	28.19	372.35	23.17	21.90	11.72	32.08	198	4.24***	.39
Sim Data 5 – 7M	394.32	28.06	372.38	23.35	21.94	11.79	32.09	198	4.26***	.39

Note. 7R=Seventh Grade Reading OATs, 7M=Seventh Grade Math OATs. For the actual and one simulated Reading OAT data sets, equal variances were not assumed and Levene’s correction was applied. ***(p < .001).

Table 10

Seventh Grade Reading OAT Score Differences Between White and Minority Students (N=200)

	White (n = 129)		Minority (n = 71)							
Data Source	M	SD	M	SD	Mean Difference	95% CI Lower	95% CI Upper	df	t	η^2
Actual Data – 7R	402.28	28.84	394.83	26.29	7.45	-0.70	15.60	198	1.80	.13
Sim Data 1 – 7R	402.22	31.88	394.24	27.68	7.98	-0.90	16.86	198	1.77	.13
Sim Data 2 – 7R	401.61	31.50	393.94	27.15	7.67	-1.08	16.42	198	1.73	.13
Sim Data 3 – 7R	402.73	31.32	394.48	27.19	8.25	-0.47	16.97	198	1.87	.14
Sim Data 4 – 7R	402.81	32.41	394.59	28.46	8.22	-0.83	17.27	198	1.79	.13
Sim Data 5 – 7R	403.85	40.27	395.51	26.11	8.34	-2.12	18.80	198	1.96	.12
Actual Data – 7M	394.84	23.84	387.20	19.40	7.64	1.12	14.16	198	2.31*	.17
Sim Data 1 – 7M	395.42	29.85	384.28	26.77	11.14	2.75	19.53	198	2.61**	.19
Sim Data 2 – 7M	394.25	29.45	382.92	26.20	11.33	3.07	19.59	198	2.71**	.20
Sim Data 3 – 7M	394.38	28.28	383.70	25.51	10.68	2.71	18.65	198	2.64**	.19
Sim Data 4 – 7M	394.60	29.16	383.13	26.05	11.47	3.28	19.66	198	2.76**	.20
Sim Data 5 – 7M	394.59	29.16	383.32	25.86	11.27	3.10	19.44	198	2.72**	.20

Note. 7R=Seventh Grade Reading OATs, 7M=Seventh Grade Math OATs. *($p < .05$), **($p < .01$).

(Law & Kelton, 2000, p. 279). We do, however, speculate that the higher standard deviations for simulated data are a result of simulating normally distributed data for each achievement category. Using a normal or uniform distribution for simulating data may have allowed for slightly greater variability in the simulated data sets comparison to that of the actual data, as the simulated data likely had more scores in the tails of each achievement level range than would occur naturally in actual student scores.

Although simulation as a methodology has become more widely accepted in the social sciences in general (Axelrod, 2005; Gilbert & Troitzsch, 2005), educational researchers have yet to adopt simulation as a common methodological approach (Axelrod, 2005). This study introduced an application of microanalytical simulation methods to the field of education when needed student-level data were not available to the researchers, but school level data existed. Specifically, this study demonstrated the process and provided evidence that simulating student-level data in the field of education might be an effective research method *if* adequate school level modeling information were available. With technology advancements and NCLB requirements for school, district, and state level data availability for the public (Hood, 2007), the challenge of finding sufficient modeling information is immensely lessened.

Future simulation research in education should focus on employing a greater range of simulation methods to bring the field more in line with other disciplines in the social sciences that are already implementing such research methods (e.g., anthropology, business, economics, environmental planning, law, information, organization theory, political sciences, and public policy). Simulation methods could assist in combating the criticism education program evaluations have recently received (Applebaum & Schwartzbeck, 2002; Borman, Hewes, & Overman, 2002; Crowley & Hauser, 2007; Slavin, 2002) as educational researchers and evaluators would have greater opportunities to execute more rigorous methodological designs (experimental and quasi-experimental) because they would have accesses to necessary data sources to do so by generating their own student-level data from the population parameters available. For example, microanalytical simulation methods could be helpful in determining pre-intervention baseline levels if researchers did not collect such student-level data in advance of program implementation and could not obtain it retrospectively. Unfortunately, educational evaluators/researchers commonly become involved in a program's evaluation after its inception. In these cases, collecting baseline data on students' achievement, attendance, behavior, or other outcome measures of importance prior to the program's implementation may be difficult or impossible. With appropriate school level data sources and microanalytical simulation, baseline data could be simulated based on the known population parameters. Researchers could also simulate comparison group data from demographically similar schools that are not involved in the reform/intervention program and have no vested interest in supplying actual student-level data to the researchers. As a result of using microanalytical simulation methods and more rigorous research designs allowing for comparison – intervention vs. control group, longitudinal study or a combination of both – a more sound understanding of program efficacy could be determined allowing for better informed data driven decisions to be made regarding program strengths and areas of needed improvement, continued funding, and whether or not scale-up should occur.

Although this study provided a supportive foundation for using microanalytical simulation as a research method in the field of education, this recommendation does not come without due caution. With the limited, yet encouraging, validity evidence for using this research method, microanalytical simulation should only be used for exploratory purposes at this time and not for making high stakes decisions. Common practice has shown that most microanalytical simulation models are considered works in progress, “as far as their builders are concerned – major enterprises requiring many person-years of expertise, attention to detail and stamina” (Mitton, Sutherland, & Weeks, 2000, p. 10). Therefore, model revision for the sake of improvement is often done in an effort to obtain a valid working model requiring multiple iterations. Further, a model that works well in one situation may not function satisfactorily in other circumstances, requiring appropriate adjustments to be made. Specifically, this microanlaytical simulation model, although it demonstrated valid and replicable outcomes based on Ohio aggregate achievement and student demographic data, would most certainly need adjustment and should be tested further if used for data from a different state or with outcome measures other than state student achievement tests (e.g., graduation rate, attendance rate, and behavior incidents).

Additional considerations with regard to the process of simulating student-level data from school level data are essential. Educational researchers must first adopt and follow a well-defined procedure for performing such studies in a methodical approach to avoid haphazard methods that are nonreplicable. Law and Kelton (2000) and Gilbert and Troitzsvh (2005) provide in-depth background on simulation methods and history in the social sciences. They also offer similar useful guidelines for performing valid simulation studies. Finally, researchers interested in simulating individual-level data from population data need to realize that creating conceptual models from which to simulate student-level data is an iterative process, and the first model established *may not* be the best model. Building an accurate conceptual model of the system requires a considerable amount time and research, multiple sources of information, and the use of SMEs (subject matter experts) to fully inform the development of a valid conceptual model to simulate data from (Law & Kelton, 2000). Failing to follow well-established simulation guidelines and taking short-cuts during the conceptual model development stage would inevitably lead to invalid results that provide inappropriate information for decision making regardless of how remarkable the simulation results appear.

References

American Educational Research Association (AERA). (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher, 35*, 33-40.

American Psychological Association (APA). (2002). *Publication manual* (5th ed.). Washington, DC: APA.

Applebaum, D., & Schwartzbeck, T. (2002). Defining, measuring and supporting success: Meeting the challenges of comprehensive school reform research. In *CSR Connection*. Washington, D.C.: National Clearinghouse for Comprehensive School Reform. Retrieved from ERIC database. (ED467446).

Axelrod, R. (2005). *Advancing the art of simulation in the social sciences.* Retrieved from http://www-personal.umich.edu/~axe/research/AdvancingArtofSim.pdf

Beghetto, R. (2003). *Scientifically based research.* Washington, D.C.: Institute of Education Sciences. (ERIC Document Reproduction Service No. ED475107).

Crowley, J., & Hauser, A. (2007). Evaluating whole school improvement models: Creating meaningful and reasonable standards for review. *Journal of Education For Students Placed At Risk, 12*(1), 37-58.

Data Quality Campaign. (2010). *Inaugural overview of states' actions to leverage data to improve student success.* Austin, TX: National Center for Educational Achievement. Retrieved from http://www.dataqualitycampaign.org/resources/details/846

EpiGear International. (2012). *Product list.* Retrieved from http://www.epigear.com/index_files/products.html

Gilbert, N., & Troitzsch, K. (2005). *Simulation for the social scientist* (2nd ed.). London: Open University Press.

Gohler, A., & Geisler, B. (2012). Automated deterministic sensitivity analysis for first-order Monte Carlo simulations with Excel and TreeAge. *Institute for Technology Assessment.* Retrieved from http://www.mgh-ita.org/index.php?option=com_content&task=view&id=163&Itemid=84

Harnett, D., & Horrell, J. (1998). *Data, statistics, and decision models with Excel.* New York: John Wiley & Sons, Inc.

Hood, L. (2007). The impact of data on education. *Carnegie Reporter, 4*(4), 1-5.

Law, A., & Kelton, W. (2000). *Simulation modeling and analysis* (3rd ed.). New York: McGraw-Hill.

MathWave Technologies. (2011). *How to generate random numbers in Excel worksheets.* Retrieved from http://www.mathwave.com/articles/random-numbers-excel-worksheets.html

Milton, L., Sutherland, H., & Weeks, M. (Eds.). (2000). *Microsimulation modeling for policy analysis: Challenges and innovations.* Cambridge, United Kingdom: Cambridge University Press.

Public Health Agency of Canada. (2006). Population health impact of disease in Canada: Workbooks and microsimulation. Retrieved from http://www.phac-aspc.gc.ca/phi-isp/workbooks-eng.php

School Communities That Work. (2008). *A user's guide to getting school-level budget and enrollment data.* Providence, RI: Annenberg Institute for School Reform at Brown University.

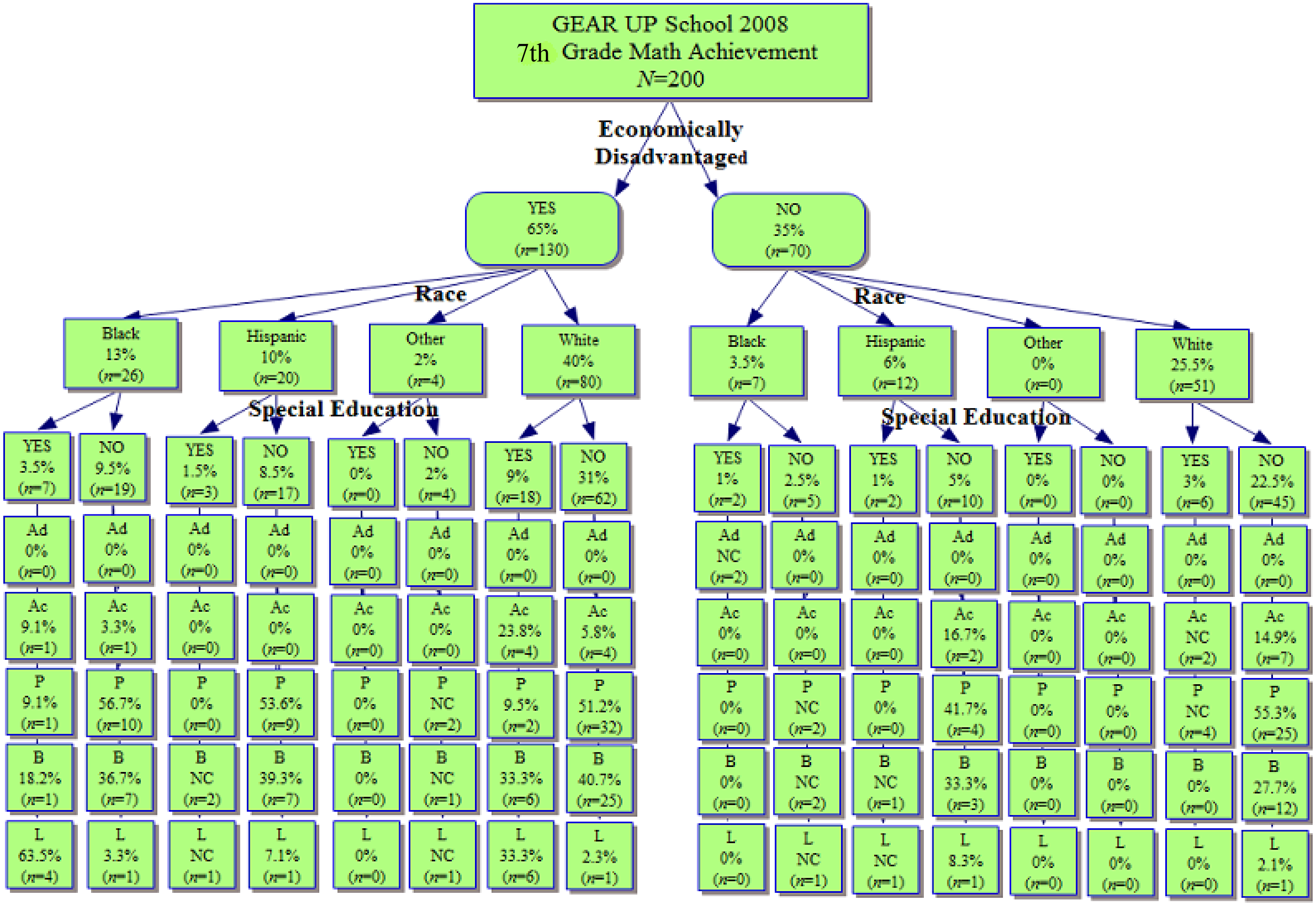
Slavin, R. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*, 15-21.

Slavin, R. (2008). Evidence-based reform in education: Which evidence matters? *Educational Researcher, 37*, 47-50.

U.S. Department of Education. (2004). *Executive Summary of the No Child Left Behind Act of 2001.* Retrieved from http://www.ed.gov/nclb/overview/intro/execsumm.html

Appendix A

Sample Conceptual Model for Simulating Student-level Data



Appendix B

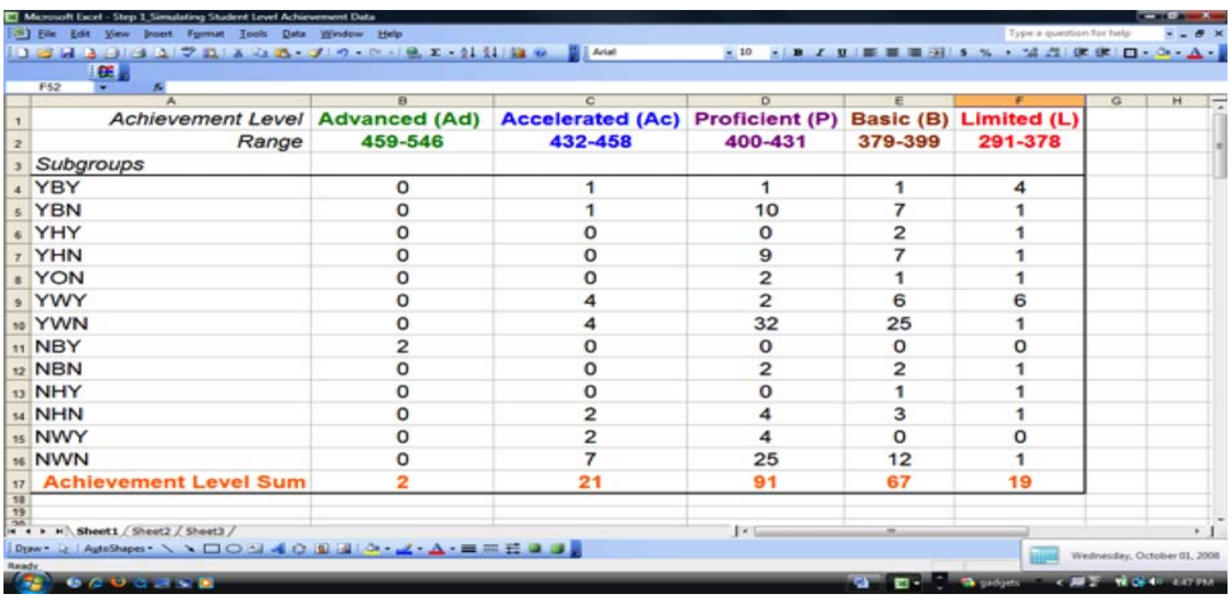
Directions for Generating Student-level Achievement Data from School Level Data in Excel

Excel 2003 with or without the Data Analysis Package could be employed to generate the simulated data. Step-by-step directions for generating student-level achievement data from the reported school level data are provided. Seventh grade math (2008) OAT results for the model sample are applied in the demonstration below.

Creating the Simulation File

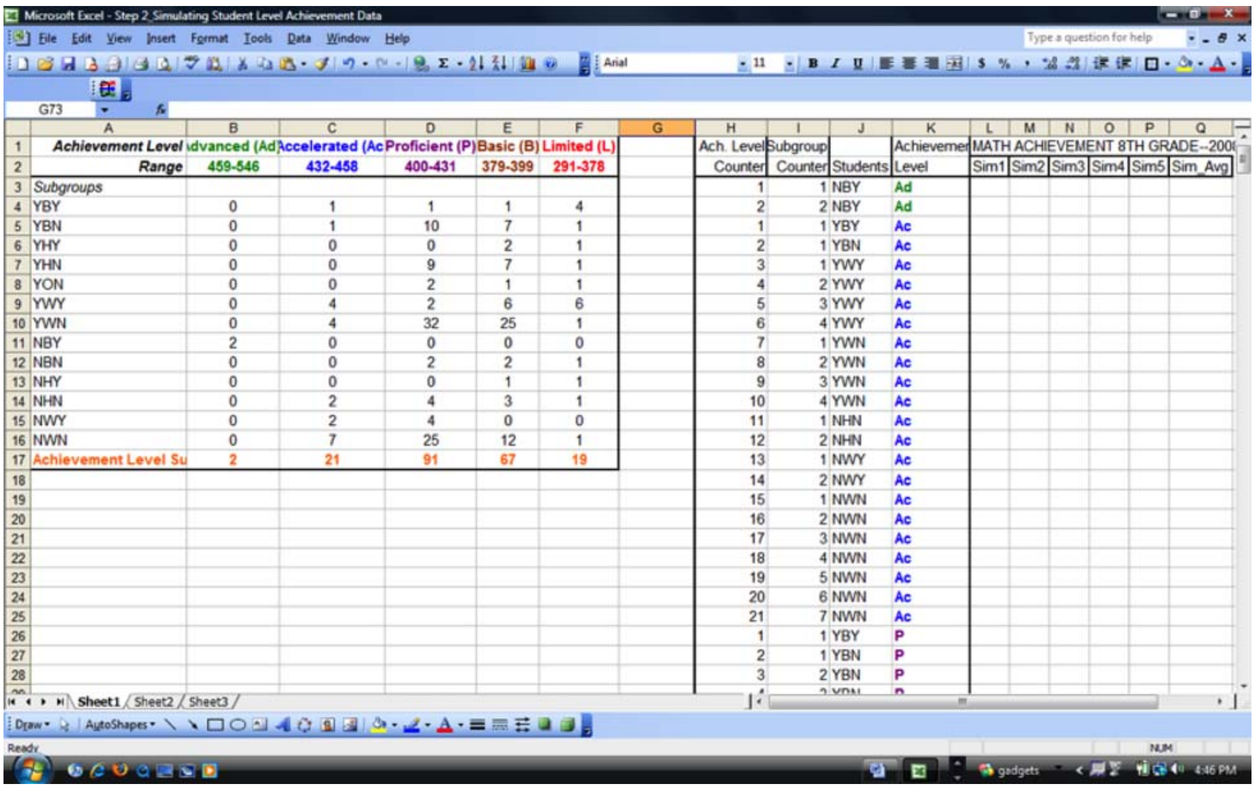
Step 1. Input achievement model specifications into Excel. Column A in Figure 1 has all of the student subgroups for the sample to be simulated coded by economically disadvantaged status, race, and special education status. Economically disadvantaged is either Y (yes) or N (no); race codes are B (Black), H (Hispanic), O (Other), and W (White); special education status is either Y (yes) or N (no). For example, the first subgroup code is YBY, this represents the student subgroup of economically disadvantaged (Y), Black (B), special education (Y) students. Another code explained is NHN which represents the student subgroup of non-economically disadvantaged (N), Hispanic (H), non-special education (N) students. All possible combinations of economically disadvantaged status, race, and special education status for the representative sample of students at the Gear Up school are provided. Total frequencies for each subgroup were derived during the sampling plan and are shown in Figure 1.

Columns B through F represent the total number of subgroup students in each of the OAT achievement levels (Advanced, Accelerated, Proficient, Basic, and Limited). Ranges for each achievement level are a necessary factor when generating student-level data and may be obtained on the Ohio Department of Education website (http://www.ode.state.oh.us/). For example, cell D7 indicates 9 economically disadvantaged, Hispanic, non-special education (YHN) students from the sample scored in the Proficient range (400-431) on the 2008 Math OAT.



Achievement Level	Advanced (Ad)	Accelerated (Ac)	Proficient (P)	Basic (B)	Limited (L)
Range	459-546	432-458	400-431	379-399	291-378
Subgroups					
YBY	0	1	1	1	4
YBN	0	1	10	7	1
YHY	0	0	0	2	1
YHN	0	0	9	7	1
YON	0	0	2	1	1
YWY	0	4	2	6	6
YWN	0	4	32	25	1
NBY	2	0	0	0	0
NBN	0	0	2	2	1
NHY	0	0	0	1	1
NHN	0	2	4	3	1
NWY	0	2	4	0	0
NWN	0	7	25	12	1
Achievement Level Sum	2	21	91	67	19

Step 2. Create individual students based on model specifications. Columns H and I in Figure 2 are not necessary, but highly recommended as they are very helpful for keeping track of the number of students in each achievement level by subgroup. Column H, Achievement Level Counter, keeps track of the number of students (regardless of subgroup) for each achievement level. Column I tracks the number of students per subgroup in each achievement level. To explain the Achievement Level Counter further, look at Cells H3 and H4. The numbers in these cells are 1 and 2 respectively and are counting the total number of students (regardless of subgroup) in the Advanced level (Cell B17). At Cell H5, the count begins over from 1 and continues to 21 in Cell H25 to track the total number of students (regardless of subgroup) in the Accelerated level (Cell C17). This pattern of tracking total number of students per achievement level continues through Limited with each category beginning at 1 and ending with the corresponding Achievement Level Sum in Row 17.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Achievement Level	dvanced (Ad	ccelerated (Ac	Proficient (P)	Basic (B)	Limited (L)		Ach. Level	Subgroup		Achievemen	MATH ACHIEVEMENT 8TH GRADE--200					
2	Range	459-546	432-458	400-431	379-399	291-378		Counter	Counter	Students	Level	Sim1	Sim2	Sim3	Sim4	Sim5	Sim_Avg
3	Subgroups							1	1	NBY	Ad						
4	YBY	0	1	1	1	4		2	2	NBY	Ad						
5	YBN	0	1	10	7	1		1	1	YBY	Ac						
6	YHY	0	0	0	2	1		2	1	YBN	Ac						
7	YHN	0	0	9	7	1		3	1	YWY	Ac						
8	YON	0	0	2	1	1		4	2	YWY	Ac						
9	YWY	0	4	2	6	6		5	3	YWY	Ac						
10	YWN	0	4	32	25	1		6	4	YWY	Ac						
11	NBY	2	0	0	0	0		7	1	YWN	Ac						
12	NBN	0	0	2	2	1		8	2	YWN	Ac						
13	NHY	0	0	0	1	1		9	3	YWN	Ac						
14	NHN	0	2	4	3	1		10	4	YWN	Ac						
15	NWY	0	2	4	0	0		11	1	NHN	Ac						
16	NWN	0	7	25	12	1		12	2	NHN	Ac						
17	Achievement Level Su	2	21	91	67	19		13	1	NWY	Ac						
18								14	2	NWY	Ac						
19								15	1	NWN	Ac						
20								16	2	NWN	Ac						
21								17	3	NWN	Ac						
22								18	4	NWN	Ac						
23								19	5	NWN	Ac						
24								20	6	NWN	Ac						
25								21	7	NWN	Ac						
26								1	1	YBY	P						
27								2	1	YBN	P						
28								3	2	YBN	P						

The Subgroup Counter in Column I tracks the number of students from each subgroup at each achievement level. Looking at Cells I3 and I4, the numbers in these cells are 1 and 2 respectively. These numbers represent the number of students from the first subgroup that have Advanced level student scores (NBY—Cell B11). There are no other students with Advanced level scores so the counter is reset and begins again for the Accelerated level. Cells I5, I6 both have a 1 in them because YBY and YBN subgroups each only have one student scoring in the Accelerated level. Cell I7 begins with 1 and continues through 4 as the YXY subgroup has four students in the Accelerated level. These counters are practical checks of the data counts and will aid in creating the simulated data file. Corresponding Student subgroup codes and Achievement Level labels are inserted into Columns J and K.

Creating the number counter in Excel is very simple and done the same way for both number counters in Figure 2. To explain the programming, the Achievement Level Counter in Column H will be used. First, type "1" into Cell H3 to begin the count. In Cell H4, type the command =H3+1 which will place a "2" in this cell. Copy this command and paste it in the same amount of cells as the sample is (200 in this case). Each cells H3 through H202 will have the numbers 1 through 200 respectively. To reset the counter to

“1” in Cell H5, simply type the number “1” in and the counter will reset. Continue entering “1” in each cell where the count restarts for the number of students in an achievement level (regardless of subgroup).

Step 3. Generate simulated student-level data. To produce simulated student-level data from the model, the range from which a desired random score is to be generated needs to be known. The distribution type selected for use in generating the simulated data was a normal or uniform distribution because we had no expectation that student scores in any of the achievement level categories would be skewed. The Excel formula used is the following:

=TRUNC(RAND()*(High Score – Low Score)+Low Score)

The TRUNC command means truncate, indicating the decimals will be dropped after the calculation has been completed. This insures there will be only whole numbers and the High number will not stretch beyond its range into the next range of achievement level scores. Generating a random number between 0 and .999 is done by the RAND() command. If using Excel 2002 or earlier versions this function should not be used with large simulation models (hundreds of variables run thousands of times) as it can run out of random numbers. This limitation has however been fixed in Excel 2003 and later versions (MathWave Technologies, 2001). The random number is then multiplied by the range of achievement scores for the level (High Score – Low Score). The Low Score is then added to the product of the random number times the range to provide a random student-level score within the appropriate range. This formula is pasted down as far as needed for the specified achievement level and uses a new random number (0-.999) to generate each student achievement score for every cell.

For example, creating random scores for 21 students in the Accelerated range (432 – 458) begins by inserting the formula =TRUNC(RAND()*(458-432)+432) into Cell L5. This cell is then copied and pasted into the 20 cells below to generate the remaining simulated Accelerated math scores. See Figure 3. This same procedure is completed for each achievement level using the proper range of scores for each level.

Repeat the steps to generating student-level data from school level data as many times as deemed necessary for retaining reliable results. There is no set number of times this should be done, however, enough runs need to be completed that the researcher feels confident the program is producing data that are a valid representation of the system.