

**A SAS Macro for Producing Benchmarks
for Interpreting School Effect Sizes**

Brian E. Lawton
Curriculum Research & Development Group
University of Hawaii at Manoa
Honolulu, HI

December 2012

© Copyright 2012 Curriculum Research & Development Group,
University of Hawaii at Manoa

Please cite as:

Lawton, B. E. (2012). A SAS macro for producing benchmarks for interpreting school effect sizes [Computer software code]. Retrieved from <http://goo.gl/Op3q4>

```
/******DESCRIPTION******/
```

```
/*This code calculates the sizes of the differences between all pairs of schools within a year and grade. It calculates the absolute difference between two schools and divides this number by the pooled standard deviation of those two schools, performing this for every school with every other school. For each grade within a year, it calculates and reports the 1st, 2nd (median), and 3rd quartile of these effect sizes. It also calculates the average 1st, 2nd, and 3rd quartile across all years within one grade.
```

```
/***DIRECTIONS***/
```

```
/*
```

1. Format the dataset you are analyzing.
 - a. This code requires that your SAS dataset have the following three variables (named as such): Year, Grade, and School. If you do not have multiple years, or multiple grades, you should create a dummy value for all observations (e.g., "1111" for year; "11" for grade).
 - b. This code assumes you are analyzing the effects on outcome variable (math test scores, for example).
 - c. Name this outcome variable using 4 digits or less (too many digits might result in truncated names in the code).
2. Specify the following in the section where it says "DATA SPECIFICATION":
 - a. After the "%let test= ", type name of the outcome variable that is in your dataset (e.g., "Math" or "Read"). The length should be four digits or fewer in length.
 - b. After "%let OrigDataSet =", enter name of your dataset.
 - c. After "libname Effects", type the path to the directory where your dataset resides. Type within the quotes.
 - d. After "%let Folder =", type the path to the directory where you can save datasets from running this code.
 - e. After the "%let CSV_Destination =", type the path to the directory where you want the CSV output of the results saved (can be the same directory as either one of the above).
 - f. After the "%let MinNumStudents =", type the minimum number of students (or scores on the test) per school to include in your comparison.

g. After "%let InvalidScore =", type a code (e.g., "9" or "100") that is used in the dataset for indicating invalid scores on the outcome variable. If none, type a period[.]).

3. At the end of the section of the code entitled "EFFECT CALCULATION", specify all combinations of years and grades.

a. For example, if you are examining Grades 5 and 8 in the years 2002 and 2003, your code would be this:

```

    %effect (year = 2002, grade = 5)
    %effect (year = 2002, grade = 8)
    %effect (year = 2003, grade = 5)
    %effect (year = 2003, grade = 8)
*/

```

```

/*****DATA SPECIFICATION*****/

%let test = math; /*-----The length should be four or fewer digits long
(e.g., if your test is called "Reading" create a new variable called
"Read").*/

%let OrigDataSet = Your_data_set_name; /*-----Specify the name of your
dataset after = */

libname Effects "C:\. . .your data set directory"; /*-----Specify the
location of your dataset.*/

%let Folder = "C:\...location_of_saved_data"; /*-----Specify a location for
saving datasets in this code.*/

%let CSV_Destination = "C:\. . .location_of_CSV_outputs"; /*-----Specify a
location for saving CSV outputs.*/

%let MinNumStudnts = 25; /*-----Specify the minimum number of students (or
scores on the test) per school to include in your comparison.*/

%let InvalidScore = 100; /*-----Specify the code used for invalid scores. If
none is used, type a period [.]*/

libname &test._DS "&Folder"; /*There is no need to change this*/
options mprint spool;

/*If the name of your test is longer than four digits (e.g., if it is
"Reading"), create a new variable (e.g., "Read") that is four digits long.
Here's an example:

****EXAMPLE ONLY*****/

/*data Effects.&OrigDataSet.;
set Effects.YourPreviousDataSet.;
rename Read = reading;
run;*/

****SUBSET SELECTION*****/

/*This section of the code subsets the data based on the outcome variable you
specified, and removes invalid scores.*/

```

```

proc sort data = Effects.&OrigDataSet.;
    by year grade school;
run;
data temp1; set Effects.&OrigDataSet.;
    if &test. NE . and &test. NE &InvalidScore.;
run;

/*The PROC SUMMARY statement calculates the N, mean, and the standard
deviation of scores within each school. This code also eliminates schools
that do not have at least N scores per school (in our example, we are keeping
schools that have at least 25 or more students per school.)*/

proc summary data = temp1;
    by year grade school;
    var &test.;
    output out = temp2 n = n&test. mean = mean&test. std = std&test.;
run;
data &test._DS.&Test._ge_&MinNumStudnts.; set temp2;
    if n&test. ge &MinNumStudnts.;
    if grade ne 31;
run;

/*The second PROC SUMMARY statement calculates the number of schools in each
grade in a year. This will be used in the do loop in the next section.*/

proc summary data = &test._DS.&Test._ge_&MinNumStudnts.;
    by year grade;
    var school;
    output out = temp3 n = nschools;
run;

/****EFFECT CALCULATION****/

/*This section of the code is the macro. Each iteration is invoked by the
macro calls at the end of the section. For example, the call "%effect
year=2002, grade=3" tells the code to include only the data within the grade
and within the year.*/

/*The first data step creates a subset composed of only schools' scores
within the grade within the year. It outputs this to a flat file stored in
your folder location (which you specified above).*/

%macro effect (year=, grade=);
data _NULL_;
    set &test._DS.&Test._ge_&MinNumStudnts.;
    if year = &year;
    if grade = &grade;
    file "&Folder.\&Test.gr&grade._&Year.";
    put @1 school @10 n&test @20 mean&test @35 std&test;
run;

/*The PROC SQL statement uses the count of schools within the grade within
the year and creates a macro variable (nschools) which will be used in the do
loop*/

proc sql noprint;
    select (nschools)

```

```

        into: NumSchls
        from temp3 (where=(Year=&Year. and grade=&grade.)) ;
quit;
%let nschls = %trim(%left(&NumSchls));

/*The next data step uses the flat file data and do loops to calculate the
effect size of every possible pair comparison.*/

data gr&grade._&Year.;
    infile "&Folder.\&Test.gr&grade._&Year.";
    array school{&nschls};
    array size{&nschls};
    array mean{&nschls};
    array std{&nschls};
    n=&nschls;
do i=1 to n;
    input school{i} $ size{i} mean{i} std{i}; end;
do i=1 to n;
do j=i+1 to n;
    dmean =abs(mean{i}-mean{j});
    stdweight=(((std{i}*std{i})*size{i})+((std{j}*std{j})*size{j}));
    stdpool=sqrt(stdweight/(size{i}+size{j}));
    d=dmean/stdpool;
    output;
end;
end;

/*PROC UNIVARIATE calculates the descriptive statistics (n, mean, std, min,
Q1, Q2, Q3, max) of the many effect sizes within this grade and year.*/

proc univariate noprint;
var d;
output out = temp n = n mean = mean std = std max = max min = min q1 = Q1
median = median q3 = Q3;
data Avg_gr&grade._&year.;
set temp; year = &year; grade = &grade;
run;

/*PROC APPEND appends the PROC UNIVARIATE output to a dataset which will
contain all subsets' descriptive statistics. This prepares the results for
subsequent print and means procedures.*/

proc append base=Appended&test._results
data=Avg_gr&grade._&year.;
run;

/*PROC SORT with the NODUP option eliminates duplicates. This is useful
because without it, PROC APPEND would continue to append duplicating data in
the event that the user runs the program more than one time in a session. */

proc sort NODUP data=Appended&test._results;
by year grade;
run;

%mend;
options nomprint;
%effect (year = 2002, grade = 3);

```

```
%effect (year = 2002, grade = 5)
%effect (year = 2002, grade = 8)
%effect (year = 2002, grade = 10)
%effect (year = 2003, grade = 3)
%effect (year = 2003, grade = 5)
%effect (year = 2003, grade = 8)
%effect (year = 2003, grade = 10)
%effect (year = 2004, grade = 3)
%effect (year = 2004, grade = 5)
%effect (year = 2004, grade = 8)
%effect (year = 2004, grade = 10)
%effect (year = 2005, grade = 3)
%effect (year = 2005, grade = 4)
%effect (year = 2005, grade = 5)
%effect (year = 2005, grade = 6)
%effect (year = 2005, grade = 7)
%effect (year = 2005, grade = 8)
%effect (year = 2005, grade = 10)
%effect (year = 2006, grade = 3)
%effect (year = 2006, grade = 4)
%effect (year = 2006, grade = 5)
%effect (year = 2006, grade = 6)
%effect (year = 2006, grade = 7)
%effect (year = 2006, grade = 8)
%effect (year = 2006, grade = 10)
%effect (year = 2007, grade = 3)
%effect (year = 2007, grade = 4)
%effect (year = 2007, grade = 5)
%effect (year = 2007, grade = 6)
%effect (year = 2007, grade = 7)
%effect (year = 2007, grade = 8)
%effect (year = 2007, grade = 10)
%effect (year = 2008, grade = 3)
%effect (year = 2008, grade = 4)
%effect (year = 2008, grade = 5)
%effect (year = 2008, grade = 6)
%effect (year = 2008, grade = 7)
%effect (year = 2008, grade = 8)
%effect (year = 2008, grade = 10)
%effect (year = 2009, grade = 3)
%effect (year = 2009, grade = 4)
%effect (year = 2009, grade = 5)
%effect (year = 2009, grade = 6)
%effect (year = 2009, grade = 7)
%effect (year = 2009, grade = 8)
%effect (year = 2009, grade = 10)
;
```

```
/**PRINTING**/
```

```
/*This section of the code prints the descriptive statistics of the effect size results.*/
```

```
/*The datastep rearranges the order of the variables as they will print out. The ODS lines output the printout to a CSV file, which can be opened with Excel.*/
```

```
data all_&test._effect_results;
    retain grade year n mean std min Q1 median Q3 max;
set Appended&test._results;
run;
proc sort data = all_&test._effect_results;
    by grade;
run;
ODS noresults;
ODS CSV FILE = "&CSV_Destination.\all_&test._effect_results.CSV";
proc print data=all_&test._effect_results noobs;
    title "All &test. Effect Size Datasets";
run; title;
ODS CSV Close;
ODS Results;

/****MEAN EFFECTS PER GRADE****/

/*This section of the code calculates and reports the average 1st, 2nd, and
3rd quartile within one grade across all years.*/

proc means n mean std stderr maxdec=3;
    by grade;
    var min Q1 median Q3 max;
    title "&test. Effect Size Statistics";
run; title;
quit;
```