Clark County School District
Program Evaluation Subcommittee of the Superintendent's Executive Advisory Group
Framework for Appraising the Adequacy of Academic Programs: Applying Criteria to Form a Summary Judgment

EXECUTIVE SUMMARY

This outlines the accomplishment of a team of District staff and community stakeholders that convened to develop a template for evaluating the adequacy of academic programs in the District. The team developed a five-step approach.

1. Including direct and indirect costs, convert total cost into a form that permits comparison across programs and schools.
2. Compare cost per student for schools using the program to the cost per student for schools not using the program.
3. Quantify program benefit in various ways including:
   - Approaches that adjust for demography of student population and approaches that do not adjust for demography
   - Approaches that focus on schools using the program and compare derived benefit to expected or promised benefit
   - Approaches that compare school performance prior to implementation versus performance after implementation
   - Approaches that compare performance of schools using the program to the performance of schools not using it
4. Compare the benefit-to-cost ratio for schools using the program to the benefit-to-cost ratio for schools not using it.
5. Judge program adequacy after disclosing any factors that could limit the confidence of conclusions that are drawn.

CONTEXT

The District is developing a framework for evaluating the effectiveness and efficiency of academic programs. The aim is to ensure funds directly support the mission to marshal people, time, and money to achieve the District vision:

> All students progress in school and graduate prepared to succeed and contribute in a diverse global society.

The District believes establishing a process for the evaluation of academic programs helps ensure that resources are applied more-efficiently and more-effectively to the highest priorities.

PURPOSE AND EFFECT

The work of this Subcommittee arises out of a larger effort by the Superintendent's Executive Advisory Group. Understanding the intent of that larger effort gives purpose to the Subcommittee work.

The Superintendent's Executive Advisory Group was convened to provide advice to the District on improving student outcomes related to spending from various funding sources, but ultimately the taxpayer. A focus of the Executive Advisory Group is to improve return-on-investment, or ROI.

The Program Evaluation Subcommittee was convened to assist in this larger work. The Subcommittee accomplishes its aim when it: (1) develops a flexible system to evaluate any academic program; and (2) formulates suggestions that can aid District efforts to implement policies, practices and procedures that lead to program-based budgeting districtwide.

The Subcommittee has developed a generic framework for evaluating any academic program. To assist the Subcommittee in its work, a particular program that is in use in roughly half of all middle schools served as a template to determine the efficiency of the framework. The program, called SpringBoard, is focused on Mathematics and English/Language Arts. Finally, the Subcommittee plans to use this "dry run" as the basis for identifying procedures that can enable the District to implement a program-based budget.

THEORY OF ACTION

This undertaking relied on a chain of logic. The logic involves a set of linked propositions that explains how change will lead to improvement. This is sometimes called a "theory of action." (Haertel, National Academy of Science, 2009).

If we . . .
- Develop repeatable processes that link data on financial and operational performance;
- Use these to calculate the return on investment for schools, departments, and programs;
- Are forthright and transparent with our community about what is working and what is not;
- And use the insights to transform how we allocate scarce resources;

. . . then resources will be used more productively, student performance will improve, and public confidence will grow.

TERMS, ASSUMPTIONS, AND LIMITS

Terms
- "Adjusted" means that analysis takes into account (and statistically controls for) the varied demography of schools.
- "Demography" means free or reduced price lunch (FRL), English learners (ELL), or Individualized Educational Plan (IEP)
- "Educational output" means points from the Nevada School Performance Framework (hereafter referred to as "NSPF").
- "English Language Learner" means any student whose first language is not English and who is not yet fluent with respect to English (in terms of reading, writing, listening and speaking).
- "Free or Reduced Price Lunch" means the national school lunch program. This is a federally assisted lunch program that has been in existence since 1946 which has provided low cost or free lunches to economically-disadvantaged public school students whose families can show they are near or below the poverty level.
- "Individualized Educational Plan" means a plan that is jointly developed by school staff and a student's parents or guardians and is intended to guide the learning program of a student who qualifies for Special Education Services. The plans are required of public schools that receive federal aid under the federal Individuals with Disabilities Education Act.
- "Program" means a prescribed set of activities intended to achieve a specific result. By connecting means and ends, it focuses resources toward accomplishing a particular goal or objective. [i]
- "Program Evaluation" means the collection and analysis of data (related to a need, goal, or vision) against a set of pre-determined criteria for the purpose of judging the adequacy of the activities under scrutiny. [ii]
- "Research" means a systematic investigation designed to develop or contribute to generalizable knowledge. [iii]
- "Return on Investment" refers to educational output per dollar expended.

Assumptions
- The focus is on academic program.
- The unit of analysis is school; however, a school's academic success is defined in terms of the academic performance of students in the care of the school (specific metrics used for this purpose are described more fully below).
- If adequacy judgments rely on comparisons of programs then comparable measures across programs are needed.
- There is a noteworthy distinction between "program evaluation" and "research." Expectations concerning rival hypotheses or threats to the validity of findings that pertain to research differ from those that apply to program evaluation. References to "generalizability" are germane in the context of research; however, they are less so in the context of program evaluation. Unlike research, the "purpose of evaluation is to *improve*, not prove." In other words, "research strives to establish that a particular factor caused a particular effect. For example, smoking causes lung cancer. The requirements to establish causation are very high. The goal of evaluation, however, is to help improve a particular program." [iv]

Limits
In order to define what something is, it is vital to recognize what it is not. This project did not set out to produce a gold-standard framework for evaluating academic programs. Nor did it aim to produce a framework that generates unassailable conclusions about program adequacy. Instead, it sought to provide a serviceable mechanism that can help District leaders improve decision-making related to the proper mix of academic programming. This effort seeks to arm

leaders who have to choose when and whether to expand, modify, or terminate various academic programs. The desired effect of this undertaking is for the District to learn how to better apply scarce resources in a more-effective and more-efficient manner. More specifically, the intent of the project was to:

- Design a transparent, repeatable process to evaluate programs that can be implemented within existing resources.
- Provide information that is useful to staff who strive to ensure every student exits college-prepared and career-ready.
- Make defensible claims related to program adequacy that are valid for their purpose and based on reliable information.
- Create conditions that elevate performance through a focus on continuous improvement and accountability for results.

In short, the purpose of this project was to develop a way to more-systematically apply program evaluation as a tool to improve the quality of educational programming. Like most tools, when used appropriately it can create value; however, used inappropriately it can cause harm.

When undertaking any evaluation, clear answers and instant tabulations are always desirable; but the reality is that hard numbers are sometimes hard to come by. This project successfully produced a functional template for program evaluation, a starting place, if you will. And through the course of six meetings, a team of 12 individuals then applied the proposed template to a single program. SpringBoard was that "straw man." Yet, it is vital that the results of this work not be used for unintended purposes. Conclusions cited in this report come with considerations that limit their usefulness. These concerns are non-trivial. Because of these, caution should be used when interpreting results. Limitations are cited early because caveats can get in the way of sound bites, and what is cited first is usually remembered forever.

The limits include factors such as the quality and quantity of resources deployed to schools and used by staff. The limits also include factors like variability in the productivity of resources devoted to professional development. Taken together, these factors raise doubt about the fidelity of implementation across school sites. As a result, until these factors are considered and adequately addressed, inferences that are made about the adequacy of an academic program should be narrowly drawn. Reasons that explain why implementation may be inconsistent from school-to-school include the following:

- Use across schools may vary due to differences in training, materials, leadership, and staff continuity and commitment.
- Efficacy may vary due to the demographics of students enrolled in schools (i.e., poverty, English language learning, etc.)
- Consistency may differ due to the varied productivity of coaches who are deployed to train staff.

The key question for policy-makers and decision-makers is this, How wise it is to assign importance to conclusions when important variables are not yet part of the equation?

ORGANIZING CONCEPTS TO GUIDE EVALUATION OF ANY PROGRAM

Five considerations are taken into account when program evaluations are being designed and executed: [v]

- Implementation
  Was the program put in place as intended (i.e., subsumes factors like training and fidelity of implementation)?

- Effectiveness
  To what extent did the program achieve what it was intended to accomplish (i.e., this is a measure of benefit)?

- Efficiency
  Were reasonable resources devoted to implementation (i.e., were budget and time necessary and sufficient)?

- Return on investment (or cost-effectiveness)
  Did program benefit justify program cost? [vi]

- Confidence in or Limits to the Conclusions
  Did evidence show that progress made toward the goal was related to the program (and not to other factors)?

A GENERAL FIVE-STEP APPROACH TO EVALUATION OF THE RETURN ON INVESTMENT ON ANY ACADEMIC PROGRAM

The Subcommittee endorses a framework built around a five-step sequence.

1. Translate total cost (direct and indirect) into cost per student; hereafter the term "cost" will mean the per student cost.*
2. Compare cost per student for schools that used the program vs cost for schools that did not.
3. Compare benefit:
   a. Compare benefit (in terms of an unadjusted NSPF change) for users vs non-users.
   b. Compare benefit (in terms of adjusted proficiency on state-administered Criterion Referenced Tests or CRTs) for users v non-users, controlling for demography (FRL, ELL, IEP) using regression analysis or other analytical tools. **
   c. Compare benefit (in terms of unadjusted growth) for users vs non-users. ***
   d. Compare derived vs expected benefit for program users.
4. Compare benefit-to-cost ratio for program users to benefit-to-cost ratio for schools using program alternatives. ****
5. Identify factors that limit the confidence of judgment made

* Following Marguerite Roza, costs are fully loaded to include both direct and indirect costs including actual salaries (not budgeted salaries); excluded are capital expenditures, food service spending, and transportation costs. [vii] [viii]

** "Proficiency" means results from state assessment; that is, criterion referenced tests or CRT.

*** "Growth" refers to longitudinal academic growth as measured by state assessment.

**** In the specific case of SpringBoard, there were no pre-established benefits promised by the vendor.

APPLYING THIS GENERAL FRAMEWORK TO THE "STRAW MAN" (SPRINGBOARD PROGRAM)

The preceding section outlined a 5-step sequence. Here those steps are applied toward the evaluation of SpringBoard. Notes that appear at the end draw attention to factors that may affect the defensibility of conclusions that are reached.

STEPS 1 and 2: TRANSLATE TOTAL COST INTO PER STUDENT COST AND THEN COMPARE COSTS FOR USERS VS NON-USERS

Appendix 1 illustrates the following for fiscal year 2014:

- The per student cost among schools using SpringBoard is $7,033
- The per student cost among schools not using SpringBoard is $6,677
- The per student costs in SpringBoard-using schools amount to $356 more than in schools not using SpringBoard.

  Note: No adjustment made for variance in staff experience (SpringBoard-using schools vs schools not-using SpringBoard). While other factors potentially affect findings in this study and are outlined in "STEP 5" below, attention is drawn here to staff experience because nearly 90 percent of general fund costs are related to salary and benefits. Thus, staff experience is a key driver in cost calculations.

STEP 3(a): COMPARE BENEFIT FOR SCHOOLS THAT USE THE PROGRAM VS BENEFIT FOR SCHOOLS THAT DO NOT USE IT

In this context, the term "benefit" means the NSPF index. Appendix 2 illustrates the following:

- The mean NSPF index for SpringBoard-using schools in 2014 was lower than in 2012; the difference was -0.38.
  - 62.14 vs 62.52.
- The mean NSPF index for schools not using SpringBoard in 2014 was higher than in 2012; the difference was 4.47.
  - 75.12 vs 70.65
- The mean math NSPF index for SpringBoard-using schools in 2014 was lower than in 2012; the difference was -3.97.
  - 22.20 vs 26.17
- The mean math NSPF index for non-SpringBoard-using schools in 2014 was lower than in 2012; the difference was -0.61.
  - 30.18 vs 30.79
- The mean reading NSPF index for SpringBoard-using schools in 2014 was higher than in 2012; the difference was 3.58.
  - 33.74 vs 30.16
- The mean reading NSPF index for non-using schools in 2014 was higher than in 2012; the difference was 4.79.
  - 38.43 vs 33.64

Taking into account the preliminary nature of these findings, this study showed that by a variety of measures (overall NSPF scores and CRT scores both adjusted and unadjusted), schools using SpringBoard generally did not out-perform their counterparts. At the same time, we note that the mean NSPF index for SpringBoard schools in 2012 and 2014 was approximately 62; by contrast, the NSPF index for schools not using SpringBoard was approximately 71 in 2012 and 75 in 2014. In other words, prior to implementation, the overall performance level among schools using SpringBoard was lower than the overall performance level among schools not using SpringBoard. In overall terms, this gap was not closed following implementation of SpringBoard. Finally, an examination of the results shows the following.

- After adjusting for demography mean change in NSPF index (2012-2014) is 4.47 for non-using schools vs -.38 for others.
- Causality aside, lower CRT scale scores in both Math and Reading are associated with schools using SpringBoard.

STEP 3(b): COMPARE BENEFIT IN TERMS OF CRT STATUS (ADJUSTED FOR DEMOGRAPHY) FOR USERS VS NON-USERS [ix]

Appendix 3 describes the mathematical formulas used to adjust for demography of the students. Appendix 4 describes the detail results from a regression analysis. They illustrate the following:

- Regression analysis shows effects are stable across two time frames: those are 2012-2013 and 2013-2014.
  Note: Three conditions are noteworthy.
    - First, causation should not be inferred because the analysis is an observational study rather than experimental design.
    - Second, because randomized control trials were not part of this analysis, the possibility of biased samples cannot be ruled out. As a result, it is plausible that non-trivial systematic differences exist between the population of schools that use SpringBoard and the population of schools that do not; moreover these differences could account for some or all of the findings that emerge. For instance, 28 of the 32 middle schools that use SpringBoard are recipients of Title I funds. In part, these schools were selected to participate in the SpringBoard implementation because they were lower-performing.
    - Third, it is useful to underscore the rationale for adjusting academic performance based on school composition or demography. Statistical techniques used in this study (i.e., regression analysis, analysis of covariance, etc.) adjust the estimated performance so middle schools that differ demographically can be compared more-readily. More precisely, judgments about program adequacy require comparisons to be made. When undertaking a program evaluation, the performance of students in schools that use the program in question is compared to the performance of students in schools that do not use the program. Statistical adjustments enable us to take into account whether schools had many, few, or no disadvantaged students. In a case where two schools were identical in every respect except that one had far more disadvantaged students, it was assumed that the school with more disadvantaged students would require more learning per student per year (in order to arrive at the desired learning target) than a school of identical size and enrollment with fewer disadvantaged students. A technique of regression (or analysis of covariance or a test of mean difference) makes it feasible to more-readily compare schools that differ -- sometimes widely -- on a dimension like demography. Caution should be used when interpreting the results of this analysis. Using regression to control for variation in demography does not mean that academic expectations should be less for any student subgroup. The approach taken assumes all students are expected to attain the same ambitious learning targets (i.e., high school graduation). The Board-adopted vision for the Clark County School District expresses a commitment to the academic success for all who attend CCSD schools. “All students progress in school and graduate, prepared to succeed and contribute in a diverse, global society.” However, because students enter school with different levels of proficiency, the road to a high school graduation is steeper for some than for others. Disadvantaged students who start out behind must travel farther to arrive at the learning destination. That is, a greater amount of learning must take place for some student in order for them to reach the desired learning target. In this context, the term “disadvantaged” refers to students who qualify for Special Education services because they have special learning needs. Or the term refers to students who are economically disadvantaged because they qualify for free or reduced price lunch. Or the term refers to students who are English language learners. That is, these are students for whom English is not their native language.

STEP 3(c): COMPARE BENEFIT (IN TERMS OF UNADJUSTED LONGITUDINAL ACADEMIC GROWTH) FOR USERS VS NON-USERS

Appendix 5 describes the results from an analysis of longitudinal academic growth. It illustrates the following:

- Results are inconclusive.
  Note: No definitive conclusion can be reached due to the inconclusive results.

STEP 3(d): COMPARE DERIVED BENEFIT VS EXPECTED BENEFIT

While this question was not considered, the Subcommittee did review material that suggested that other school districts that had used the program as designed had seen marked improvement in the number and percentage of middle school students who registered for and successfully completed Advanced Placement courses in high school. Nonetheless, no specific performance promises (pledges of results) were apparently included in the written agreements with the vendor.

STEP 4: COMPARE BENEFIT-TO-COST RATIO FOR USER SCHOOLS TO THE BENEFIT-TO-COST RATIO FOR NON-USER SCHOOLS

The two-column display in Appendix 6 illustrates the following:
- The cost per student was greater in SpringBoard-using schools ($7,033) than in non-using schools ($6,677).
- Change in mean NSP score from 2012 to 2014 was -.38 in SpringBoard-using schools versus 4.47 in non-using schools.
  Note: Some cost variability may be attributable to differences in staff experience in user schools vs non-user schools.

STEP 5: IDENTIFY FACTORS THAT MAY AFFECT FINDINGS AND LIMIT THE CONFIDENCE OF JUDGMENTS THAT ARE MADE

Factors affecting performance in schools implementing academic programs (and in this case, SpringBoard) include these:
- Fidelity of implementation
- Professional development
- Teacher experience
- Staff turnover
- Resources quality and quantity
- Productivity of the resources
- "Implementation dip"
- Supplemental resources (e.g., multi-media, technology aids, manipulatives, parallel exams, etc.)
- Class size
- Special education, English Language Learners, Free or Reduced Price Lunch, Students with a 504 Plan
- On-going monitoring

ROLLUP APPRAISAL: JUDGING SPRINGBOARD ADEQUACY BY COMPARING USERS TO NON-USERS ON VARIOUS DIMENSIONS

	Schools Using	Schools Not Using
Step 3a: Compare benefit, unadjusted NSPF change	-	+
Step 3b: Compare benefit, adjusted CRT status	-	+
Step 3c: Compare benefit, unadjusted growth	Mixed results	Mixed results
Step 3d: Compare benefit, derived vs expected	Not calculated	Not calculated
Step 4: Compare benefit-to-cost ratio	-	+
Rollup: Overall appraisal	-	+

Note: Step 1 is not displayed because it involves transforming total cost into per student cost. Step 5 is not displayed here because it represents factors that were not addressed and might affect findings.

Caution: In this particular case involving SpringBoard, the overall appraisal involves a preliminary judgment. Because limits have not been definitively addressed, results are not yet defensible. Thus, conclusions remain inconclusive.

FLOWCHART

Attached is a two-page flowchart. It displays the workflow involved in implementing the recommended framework. It also illustrates how the evaluation framework functions. The flowchart is best viewed when printed on 11"x17" paper.

RECOMMENDATIONS AND NEXT STEPS

1. Direct future efforts toward evaluating the process used to select and/or develop academic programs. The aim would be to develop a template and a repeatable process for selection and/or development of academic programs.

2. Consider and incorporate other variables in future modeling efforts. This includes length of teacher experience, amount of class turnover, class size, and others. Lines of future inquiry should consider whether, when, and how to appraise the fidelity of program implementation. With respect to implementation, it is appropriate to consider and address and appraise the duty of a program provider (vendor) and the duty of customer (the District). Further, attention should be given to the obligations and responsibilities of site versus central staff. This aspect is more-fully described in recommendation #10 below.

3. Consideration should be given to ways that the framework can be adapted to apply to the evaluation of programs that fall outside the subjects and grades for which standardized test results are available. These programs might include early childhood programming or programs in subject areas like science, social studies, or the arts. As well, the template that was applied in this case (to SpringBoard) was modified to take into account data beyond NSPF. Included were data on longitudinal academic growth and results from Criterion Referenced Tests (CRTs). In similar fashion, it is foreseeable that in the case of academic programs with a particular focus (say college readiness) that a measure other than NSPF or CRT or longitudinal academic growth may be the proper gauge of program benefit. In the case of SpringBoard for instance, it may well be that the measure of choice for appraising program benefit focuses on data related to the rate of student participation in a school in Advanced Placement (AP) courses as well as the number and percent of students taking AP courses who earn a score on AP exams of "3" or better. Modifications of this nature are sensible and appropriate and should be the focus of future attention.

4. Invest further in information technology system (in the areas of human resources, accounting and other functions) to accumulate the data and evaluate and manage it in order to more-effectively implement suggestions included in this report.

5. After recommendations are implemented and two programs have been evaluated with this framework, conduct follow-up meetings with this Subcommittee. The purpose is to evaluate the application of the recommended framework.

6. To build understanding of and support for the recommendations and attached framework, conduct "peer review" that includes senior leadership, Teacher's Association, the Administrator's Association, and other key constituencies.

7. Developing a common framework for program evaluation provides the opportunity to insert this element early in the planning stages. Prior to acquiring any academic program, thought should be given to the manner in which program evaluation will be conducted. This will address the adage, "you can't fix with analysis what has been missed in design."

8. Consider the merits to developing a District matrix of roles and responsibilities. This would clearly specify who is responsible to whom and for what insofar as program evaluation is concerned. One offshoot of this could include the production of toolkits to assist schools in marshaling the data needed to execute program evaluation.

9. In the future, as initiatives like program evaluation take on a renewed purpose and/or a different emphasis, it is likely that questions will arise related to the proper type and amount of resources to apply to this work. Consistent with standard budgeting practice, it is assumed that the resources devoted to program evaluation will be allocated (within and across departments and divisions) commensurate with the level of priority.

10. As the District develops a repeatable process to guide selection of academic programs, it is recommended that all contracts that are reached with providers incorporate performance promises. These promises or pledges include milestones that permit progress monitoring of results. It is further recommended that these contracts specify not only the obligations of the provider (related to a schedule of support, training, maintenance, and updates as well as the timely provision of supplies, materials, help etc.) but also the expectations of District staff (insofar as training, amount and conditions of use, etc.). Further, the contracts will include evaluation components that specify: (a) the purpose of the evaluation; (b) the criteria for judging adequacy; (c) the sources of data; (d) the timing and manner of evaluation; and (e) the intended effect of evaluation results.

---

[i] Source is "Program Evaluation and Performance Management" at Sage, http://www.sagepub.com/upm-data/51113_ch_1.pdf and at http://www.cdc.gov/EVAL

[ii] According to the "Standards for Educational and Psychological Testing" (2014), "Program evaluation is the set of procedures used to make judgments about a program . . . [and] tests often provide the data that are analyzed to estimate the effect of a program on outcomes such as student achievement." [Further,] "an evaluation of a program [can] sometimes synthesize results from multiple tests [and] a variety of tests can be used for evaluating programs; examples include standardized achievement tests administered by states or districts." [In addition] "Test results are often one important source of evidence for the continuation, modification, termination, or expansion of various programs." Finally, "It is relatively rare for a test to be designed specifically for program evaluation and therefore is often necessary for those who conduct such studies to rely on measures developed for other purposes (source: "Standards for Educational and Psychological Testing", National Council on Measurement in Education, the American Psychological Association, and the American Educational Research Association, 2014, pg 203-205). Further, according to the U.S. Government Accounting Office a "program evaluation may assess the program's effects beyond its intended objectives, or estimate what would have occurred in the absence of the program, in order to assess the program's net impact." [Additionally,] "a program evaluation may systematically compare the effectiveness of alternative programs aimed at the same objective." [Further,] according to the U.S. Government Accounting Office, program evaluation, can include: "outcome evaluation as a form of program evaluation that focuses on outputs and outcomes to judge program effectiveness" [as well as] "impact evaluation as another form of program evaluation that assesses the net effect of a program by comparing program outcomes with an estimate of what would have happened in the absence of the program" [and finally] "cost-benefit and cost-effectiveness analysis compare a program's outputs or outcomes with the costs (resources expended) to produce them. When applied to existing programs, they are also considered a form of program evaluation. Cost-effectiveness analysis assesses the cost of meeting a single goal or objective and can be used to identify the least costly alternative for meeting that goal. Cost-benefit analysis aims to identify all relevant costs and benefits, usually expressed in dollar terms." (source: Nancy R. Kingsbury, Managing Director, Applied Research and Methods (May, 2011), Government Accounting Office Assets Document at http://www.gao.gov/assets/80/77277.pdf)

[iii] United States Gov't Printing Office, 45 CFR 46, 2000 Code of Federal Regulations, Title 45 regarding Public Welfare and Concerning the Basic Health and Human Services Policy for the Protection of Human Research Subjects. Further, "scientifically-based research" means the "examination of causal questions [that] requires experimental designs using random assignment or quasi-experimental or other designs that substantially reduce plausible competing explanations for the obtained results . . . This applies to research studies evaluating the impacts of . . . programs on educational outcomes." (source: American Educational Research Association (2014) at http://www.aera.net/AboutAERA/KeyPrograms/EducationResearchandResearchPolicy/AERAOffersDefinitionofScientificallyBasedRes/tabid/10877/Default.aspx)

[iv] Stufflebeam, DL, 2007, CIPP Evaluation Model Checklist, found at: http://www.wmich.edu/evalctr/archive_checklists/cippchecklist_mar07.pdf. Said differently, "The primary purpose of evaluation is to provide timely and constructive information for decision-making about particular programs, not [as research does] to advance more wide-ranging knowledge or theory." Further, "program evaluation is the systematic assessment of the processes and/or outcomes of a program with the intent of furthering its development and improvement. . . . towards the end of a program or upon completion, [evaluators] provide evaluation findings, often . . . to make decisions about program continuation or expansion." (source: Office of Educational Assessment http://www.washington.edu/oea/services/research/program_eval/faq.html). In contrast

to research, "evaluation determines the merit, worth, or value of things. The evaluation process identifies relevant values or standards that apply to what is being evaluated, performs empirical investigation using techniques from the social sciences, and then integrates conclusions." (Scriven, 1991) By comparison, "research bases its conclusions only on factual results—that is, observed, measured, or calculated data. Research does not establish standards or values and then integrate them with factual results to reach evaluative conclusions." (source: Scriven, 2003 found at http://www.hfrp.org/evaluation/the-evaluation-exchange/issue-archive/reflecting-on-the-past-and-future-of-evaluation/michael-scriven-on-the-differences-between-evaluation-and-social-science-research).

[v] In the interest of clarity it is noted that these five elements are widely-accepted principles of sound program evaluation. When applied to a particular program (as in the case of a "straw man" program later in this report), it may be the case that one or another of these considerations receives more or less attention. In part, this may be due to the availability of data, the resources available to address and rule out other factors that may limit the inferences drawn by the evaluation, the sample frame, etc. Thus, these five considerations represent a point of departure for organizing the evaluation of any academic program.

[vi] Except in the case of Board-appointed committees (audit committee, etc.), matters that reach the Board first cross the Superintendent's desk. Superintendent-appointed committees like the Superintendent's Executive Advisory Group serve at the pleasure of the Superintendent and unless specifically delegated decision-making responsibility are instead charged with formulating recommendations to the Superintendent of Schools. Under District policy, the Superintendent is ultimately and fully responsible for all operational aspects, including but not limited to decisions about whether to add, retain, or remove a program. In cases where a decision concerning the use of a particular academic program requires board approval of a contract, then decision-making authority may be shared by the Superintendent and the Board of School Trustees. But in no case will a recommendation to take action on a program reach the Board unless it first has the approval of the Superintendent of Schools.

[vii] Marguerite Roza in Hess and Osberg, ed. "Stretching the School Dollar: How Schools and Districts Can Save Money While Serving Students Best", Harvard Education Press, 2011, *Now is a Great Time to Consider the Per Unit Cost of Everything in Education*

[viii] In the interest of clear communication and to promote greater public trust in school stewardship of scarce taxpayer dollars, the Subcommittee found that accountability is best served by full disclosure of all school-related costs (both direct and in-direct). The sole exceptions are capital expenditures, spending related to food service, and transportation costs. Thus, it is recommended that the written summary of every program evaluation include an appendix listing the school-level costs "fully loaded" and fully-annotated and that this appendix include the actual compensation figures as opposed to district average salaries and benefits for various positions. In the body of the written summary of every program evaluation, it is recommended that a narrative case will be made for trimming these fully-loaded costs to focus on all costs and benefits relevant to the program under review. It is recommended that the program evaluation proceed focused on the costs and benefits that are relevant to the program under review.

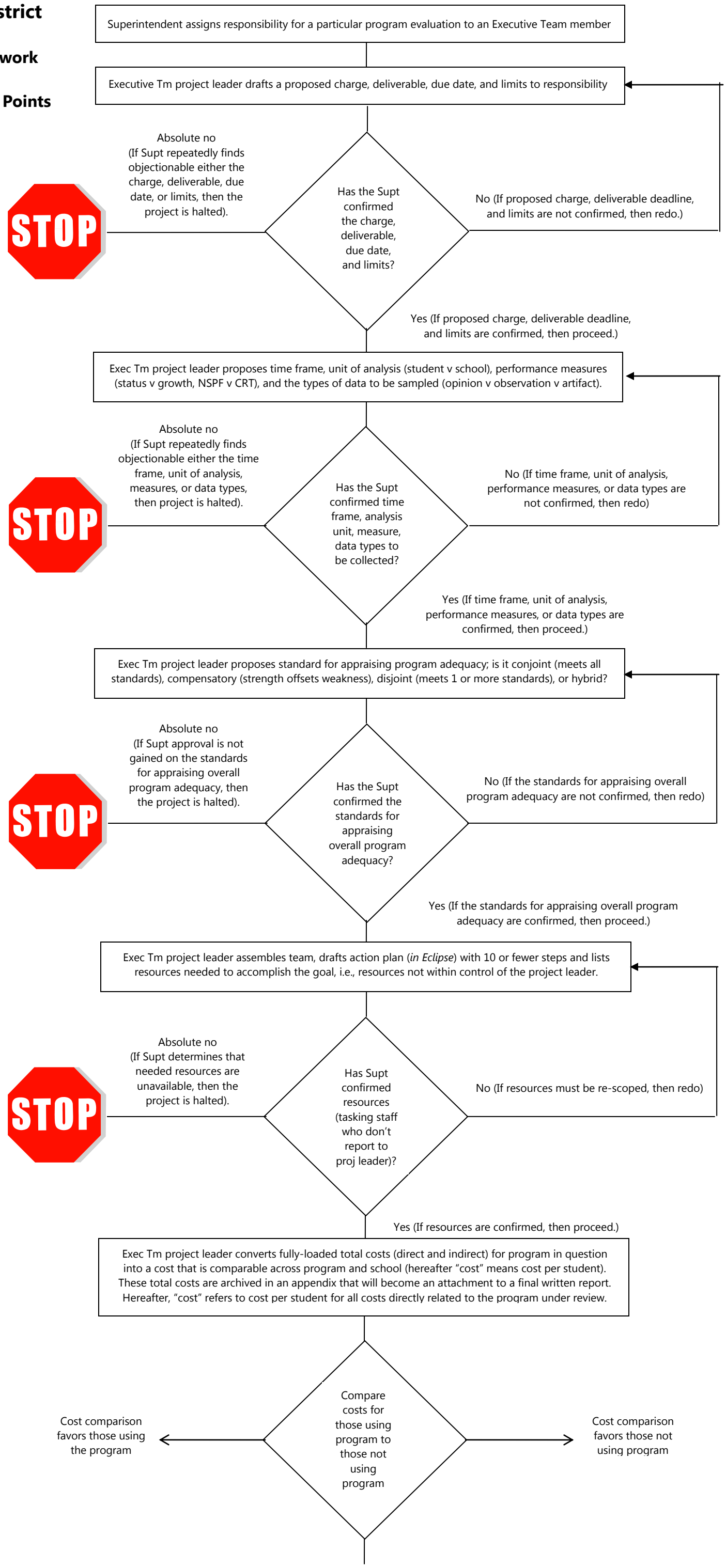
[ix] This study presents the results from analysis that statistically adjusts for demography and also results from an analysis that is unadjusted. The two perspectives provide a way to compare and contrast results. Including both perspectives makes it possible to anticipate questions or challenges that might arise.

Clark County School District

Program Evaluation Framework

Flowchart Showing Decision Points

November 20, 2014



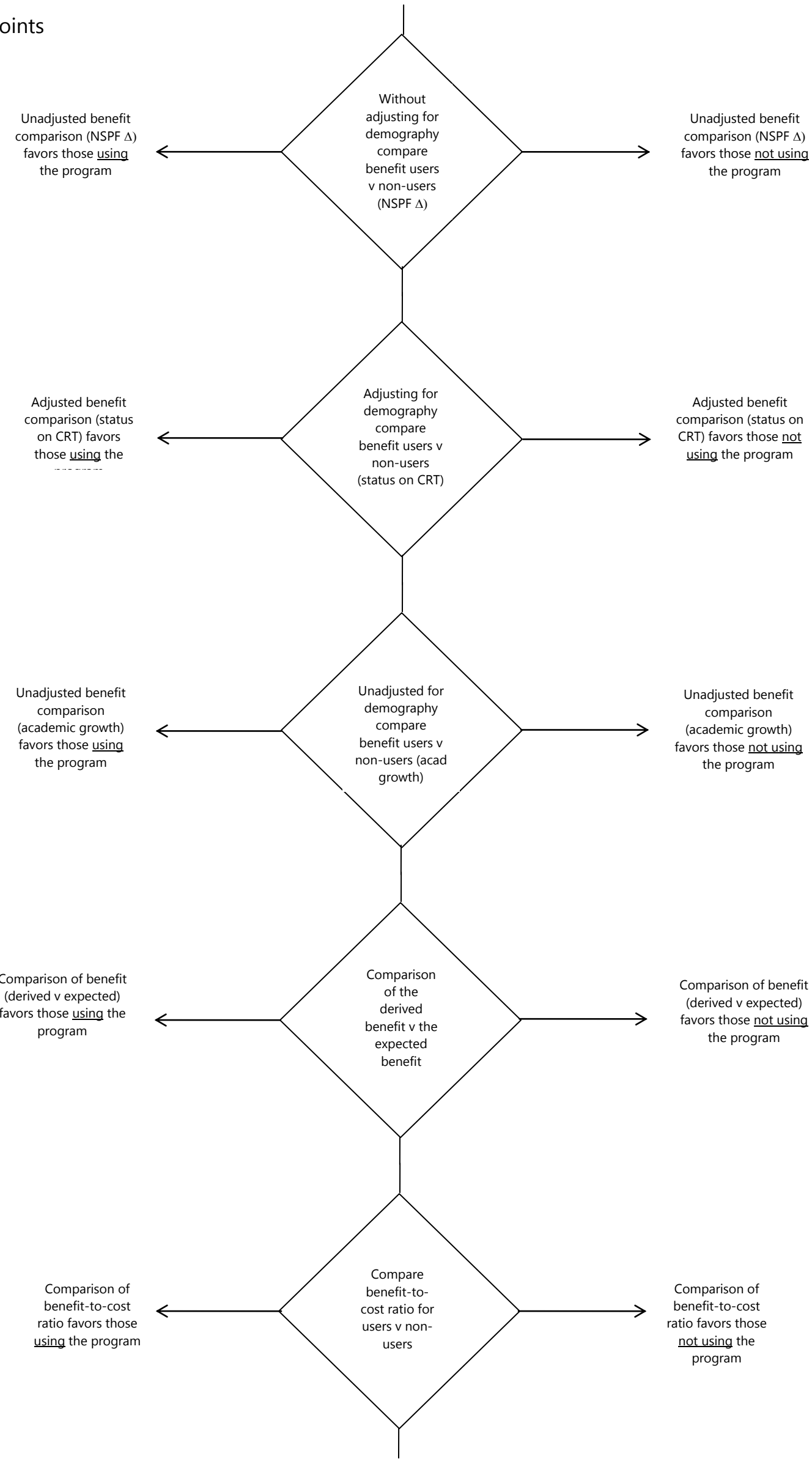
Continued on the next page

Continued . . .

Program Evaluation Framework

Flowchart Showing Decision Points

November 20, 2014



Note: For this illustration, the remainder of this flowchart assumes the metrics approved by the Superintendent (for the purpose of defining program benefit) include the Nevada School Performance Framework (NSPF), Criterion Reference Tests (CRTs) and Longitudinal Academic Growth ("growth"). As well, it is assumed that benefit is gauged by comparing derived versus expected benefit. In practice, some, all, or other metrics may be approved by the Superintendent.

	Schools Using the Program	Schools Not Using the Program
Compare cost per student	+ or -	+ or -
Compare benefit, unadjusted NSPF change)	+ or -	+ or -
Compare benefit, adjusted CRT status	+ or -	+ or -
Compare benefit, unadjusted growth	+ or -	+ or -
Compare benefit, derived vs expected	+ or -	+ or -
Compare benefit-to-cost ratio	+ or -	+ or -
Overall appraisal		