UMass | Dartmouth
DEPARTMENT OF PUBLIC POLICY

Michael D. Goodman, Ph.D.
Associate Professor and Chair
mgoodman@umassd.edu
Office: 508-910-6986

May 1, 2012

Kathryn S. Matayoshi
Superintendent
Hawai'i Department of Education

Dear Superintendent Matayoshi:

My co-authors and I have had the opportunity to review the critique of our research report prepared by former Hawai'i Department of Education staffer Dr. Thomas Gans and believe that a number of the issues raised in his critique warrant a response.

Since your letter to DoDEA Director Fitzgerald of March 8, 2012 indicates that Dr. Gans' critique will be posted on the Department's website, in the interest of transparency and open discussion, we request that our response also be posted there so that interested parties have the benefit of our perspective on these issues.

While we will address several of Dr. Gans' specific critiques in more detail below, a number of these critiques are based on misleading characterizations of the purpose of our analysis, our methods, and the conclusions of our report. To provide the proper context for these comments, we briefly review our research questions, methods and conclusions here.

As we note on page four of our report, "our report was not designed to evaluate the efficacy or efficiency of Hawai'i's public schools, their programs, curriculum or educational standards or practices." Rather, our report was intended to assess "how well military impacted public schools in Hawai'i meet the state's own standards and how Hawai'i's performance compares to that of selected other school systems serving military connected students (p. 4)."

Our report presented its research findings in four major sections:

1. In the section titled "How well do the Hawai'i Public Schools meet their Own Standards?" we introduced and discussed analysis of the "Effectiveness Index" (EI), a measure that looked at schools' performance ("proficiency" is the term used by the State of Hawai'i) in the context of their student populations' eligibility for free and reduced price lunch[1]; at

[1] More information can be found in the report's methodology section in Appendix A, pp. 59-61.

schools' status under the No Child Left Behind standard for "Adequate Yearly Progress" (AYP) (a standard also set by the state of Hawai'i).

2. In the section titled "How does Hawai'i compare to the other districts of interest?" we analyzed services offered in Hawai'i and in comparison districts (such as availability of full-day kindergarten, availability of student counselors, and high school graduation credit requirements), as well as performance on tests taken in Hawai'i and in the study's comparison districts. These tests included the standardized, nationally normed TerraNova, National Assessment of Education Progress (NAEP), and the SAT.

3. In the section titled "Identifying Performance Issues" we used a "triangulation" approach that relied on several measures, including failure to "make AYP" for two or more years and below national-median TerraNova scores, to identify military-impacted schools with "evidence of performance issues."

4. In the section titled, "Moving targets: Shifting fiscal realities and changing student needs" we analyzed the state of Hawai'i's education financing system and level of funding, finding that while the state funded education at a competitive level relative to our comparison districts, the ability of Hawai'i to provide additional funding may be limited.

A major thrust of Dr. Gans' critique was that the strength of the evidence presented by each of these analyses did not warrant the conclusions drawn. However, our research team did not rely on any one of these measures in isolation, as is implied by Dr. Gans' critique.

Rather, the team used each of these as part of a more complex analytical approach, synthesizing these and other pieces of information together to provide DoDEA and Hawai'i policymakers with the information and analysis they needed to make an informed decision.

Especially useful when individual available measures are not ideally suited to answering the questions at hand, this "triangulation" approach is a standard way in which contemporary policy analysts use multiple measures and mixed methodologies to inform evidence-based policymaking. Our use of this approach is an explicit acknowledgment that measurement and specification error are inevitable parts of the research process.

Simply stated, when different ways of looking at a question point to the same result, that result can be seen as more trustworthy than when based on one indicator alone. Our use of this approach is discussed further below.

Finally, before turning our attention to some of his more specific concerns, we feel it is important to note that, despite all of his concerns about our methodological choices, Dr. Gans ultimately endorses our major research findings, suggesting that disputes over methods remain primarily of academic interest. Nevertheless, in the interest of accuracy, we feel compelled to respond to several of the issues raised by his critique.

Adequate Yearly Progress (AYP) and the Effectiveness Index (EI)

One of our major research questions was whether Hawai'i schools attended by military dependents were meeting Hawai'i's own performance standards. Both the Effectiveness Index and the analysis of "Adequate Yearly Progress" ("AYP," the standard set under No Child Left Behind) were presented in this context.

First, Dr. Gans raises two widely recognized concerns about AYP: that it sets unachievable standards in the long run, and that it does not allow for state to state comparisons. My colleagues and I readily acknowledge that the accountability policies of NCLB are both widely criticized and problematic; however, it was the policy in effect at the time we conducted our analysis and was clearly a relevant metric for answering the question of whether Hawai'i schools were meeting state standards. Additionally, contrary to what Dr. Gans' critique suggests, we never compared AYP status across states, and in fact stated, "it [is] impossible to compare NCLB outcomes across states, since each state test is different" (p. 61).

Second, as noted in our report, "[t]he goal of the EI model is to allow for the comparison of test scores for each school within a given state or district to other schools within that state or district in a manner that considers (and controls for) the poverty level of the student body" (p. 5). We presented the EI analysis alongside information about the schools' absolute performance (whether they met state proficiency goals or not). These methods were in keeping with our clearly stated goal of assessing whether Hawai'i schools met Hawai'i-defined standards of proficiency, rather than as an attempt to explain all of the factors related to student achievement in Hawai'i schools.

Our EI is calculated by predicting a school's percentage of proficient students for a given grade and test based on a linear regression of percent proficiency on percent eligible for free and reduced price lunch, and taking the difference between the predicted and actual percent proficient. Student free and reduced-price lunch is a widely accepted and used proxy for poverty.

While we take issue with several of Dr. Gans' technical characterizations,[2] his concerns are not material in light of the fact that the EI was presented as providing an important mitigating factor to a school's absolute "proficiency" status rather than a full exploration of all relevant factors. This makes Dr. Gans' critique of the explanatory power of our regressions less relevant than it might appear.[3]

We did not, as Dr. Gans suggests, use the EI to designate schools "efficient vs. inefficient" in our report. Rather, in the cited pages (pp. 8-19), we presented the EI as a highly nuanced measure: a continuous value ranging from approximately -40 (for a school performing significantly below the predicted value) to +40 (for a school performing well above the predicted value), depending on the grade and test in question.

Finally, throughout his critique Dr. Gans expresses concern about the "transparency" of our research methods and research choices. While it is not mentioned or cited in his critique, we responded to a request for information from Dr. Gans that was submitted through the University of Hawai'i's Dr. Kathleen Berg, providing him with additional technical details for the regressions that produced the EI figures. This response is included as an appendix to this letter. Had we received additional inquiries, we would have responded to them.

TerraNova and School Performance Issues

One of the major research questions we addressed (as required by law) was how the education provided by Hawai'i public schools compares to that in similar districts identified by the Department of Defense.[4]

Here again, our research team addressed this question through several lenses: a comparison of available educational and support services (such as counseling, gifted and talented programs, etc.); a comparison of National Assessment of Educational Progress (NAEP) scores, a comparison of SAT scores; and a comparison of TerraNova scores. Our conclusions were based not on the TerraNova analysis alone, but on the synthesis of all of these findings. Furthermore, the choice of the TerraNova test as one of the indicators we considered was dictated by the comparison school districts: the TerraNova was

---

[2] Dr. Gans' characterization of the "adjusted R-squared," interpretation of the standard error, and claim that an R-squared statistic of 0.5 or greater is necessary are technically and substantively problematic. Our report repeatedly warns that we were not attempting to explain all of the variation across schools, nor would it be possible to do so.

[3] In your letter you identify several methodological changes that you state should be required of future studies. In light of our concerns about the premises upon which these recommendations are made, we advise that you and DoDEA consult with an independent third party before imposing these restrictions on future researchers.

[4] As defined in 10 U.S.C. § 2164.

not only the "best," but the *only* standardized, nationally normed test taken in Hawai'i and some of the comparison districts.

The identification of schools with "evidence of performance issues" was another instance in which our analysis was based on several measures to ensure the greatest degree of accuracy. This designation was based on a combination of TerraNova performance compared to national performance, and status across two years under the AYP standard.

Furthermore, we explicitly stated that our analysis was not intended to "[e]valuate the efficacy or efficiency of Hawai'i's public schools, their programs, curriculum or educational standards or practices" (p. 4). In this context, the discussion of schools with "evidence of performance issues" was intended to be informative, not definitive.

Clearly, the accuracy of our original TerraNova analysis is of material importance to both the question of interstate comparison, and the identification of schools with evidence of performance issues. However, Dr. Gans misidentified the original TerraNova data we used in our report and therefore his specific critiques refer to an analysis that bears little resemblance to what we actually did. Dr. Gans assumed that we used student-level TerraNova data, when in fact, due to time constraints and our inability to access student-level data, we used school-level TerraNova scores as the basis for our analysis. Since we did not in fact use these data, a direct response to his concerns about our use of student-level data is not possible.

Hawai'i School funding

This section of our report concludes, "Thus, while it appears that the commitment to funding public education in Hawai'i is not lacking, over the next several years the capacity of the state to do so remains a major question" (p. 44). Dr. Gans argues that other states' funding is also insecure and we agree. As we stated in the report: "the state still faces a significant budget challenge over the next several years. Hawai'i is not alone in this regard" (p. 42). Similarly, we acknowledged Hawai'i's relatively high level of school funding compared to national averages (p. 44). While Dr. Gans points to the greater equity of Hawai'i's system of school funding, he does not dispute the basic finding that increased revenue has been hard to come by, nor does it seem an arguable question in the context of the state's school furloughs that were the impetus for our report.

Status of Barbers Point Elementary School

We made the research choice to define "on-base schools" as including schools located in military housing areas or on federal property. DoDEA advised us that Barbers Point Elementary is located in a military housing area and that is why this

school was defined in that way. It is worth noting that this choice had no material impact on our report's research findings or policy implications.

Conclusion

In sum, we remain confident that the methods used in our study were the most appropriate given the data, measurement, time and resource constraints that this project presented our research team. We remain available to discuss any additional concerns that you or your colleagues may have regarding our work.

Mahalo for your leadership and for the constant support provided by you and your colleagues during our time in Hawai'i.

Sincerely,

Michael D. Goodman
Associate Professor and Chair
Department of Public Policy
University of Massachusetts Dartmouth

Cc Ms. Marilee Fitzgerald, DoDEA
Ms. Kathy Facon, DoDEA
Mr. Michael Lynch, DoDEA

UMass | Dartmouth
DEPARTMENT OF PUBLIC POLICY

Michael D. Goodman, Ph.D.
Associate Professor and Chair
mgoodman@umassd.edu
Office: 508-910-6986

July 30, 2011

Dr. Kathleen Berg, Associate Director
Curriculum Research and Development Group
University of Hawai‘i

Dear Kathy,

I have had the opportunity to consult with my colleagues from the UMass Donahue Institute and we have prepared the following responses to the questions you asked in your email of July 15, 2011.

Question 1 - Was the basis for the "effectiveness index" a simple linear regression of percentage proficient on the percentage of free/reduced cost lunch for each school, or was there a more complicated equation?

Yes, the Effectiveness Index is based on a simple regression that uses the percentage of students eligible for free or reduced lunch services as a proxy for student poverty.

As you are well aware, the socioeconomic status (SES) of the student population has been repeatedly demonstrated to have an impact on student performance and so when assessing school and district level performance, effectively controlling for student SES permits performance comparisons that consider a major factor that is beyond the immediate control of schools and districts and allows for a more context-sensitive understanding of school and district performance.

In previous studies (both for DoDEA and others) the Effectiveness Index has been based on more social and demographic variables (including community educational-attainment, average income, poverty-rate, single-parent status, language spoken, and others)[1].

What we have discovered in recent years is that a simpler model using the percentage of students eligible for free or reduced lunch services yields very similar results (and is therefore more parsimonious) and eliminates concerns

[1] For more information see: Robert Daniel Gaudet, "Education achievement communities: A new model for "kind of community" in Massachusetts based on an analysis of community characteristics affecting educational outcomes" (January 1, 1998). Electronic Doctoral Dissertations for UMass Amherst. Paper AAI9841870.
http://scholarworks.umass.edu/dissertations/AAI9841870

about multi-collinearity. This is why we currently use this more straightforward model.

Question 2 - Can we see the analysis of variance tables for each of the regressions? Specifically, we want to see the proportion of variance accounted for by each regression (R-squared) and the statistical significance (p values).

I have attached a memo prepared by my colleagues from the UMass Donahue Institute that contains the ANOVA tables you requested. Please contact Kate Wilkinson -- kwilkinson@donahue.umassp.edu or 774.455.7378 -- if you have any questions about the attachment.

When reviewing, please keep in mind that the Effectiveness Index is used in our report to present state criterion referenced testing data in a manner that "controls for" student SES. In other words, it was not designed to "explain" why performance varies but rather to present performance data in light of student SES.

Question 3 - How is the effectiveness index obtained? Is simply the difference between the actual percentage proficient and the predicted percentage proficient, or is the a more complicated calculation?

Yes, the Index value is the residual (the difference between the score predicted by the regression and the actual score).

Question 4 - Was the use of the TerraNova to classify schools as having performance issues based on the national median individual scores or on the national medians for schools? What was the rationale for dichotomizing the TerraNova scale rather than identifying schools substantially below the median as having "performance issues?"

The identification of schools where there was "evidence of performance issues" involved more than simply Terranova test performance. See pages 39-40 in the report for a discussion of how these schools were identified.

The TerraNova data was based on medians for schools. We were not provided with student level data. Since the TerraNova was used to validate performance issues suggested by state criterion referenced test performance (defined as two consecutive years of failing to achieve the state performance target), we made a research decision to use the 50th percentile as a cutoff because it reflected performance below the national median.

In several cases schools scored just below the 50th percentile (within a few points) on some tests and in some grades and this sparked a discussion among our research team about how one might operationally define "substantially below the median" in an effort to recognize that the difference between a school that performs at the 48th percentile is not always meaningfully different than one that

scores at the 50th. Ultimately, we decided that the median was both a reasonable and intuitive cutoff and agreed that since we were presenting all the school level Terranova scores in the report (see the sidebar on pages 30-32) , state and federal policymakers and officials would have the detailed data necessary to assess whether a given school's performance profile warranted further investigation for themselves.

Please do not hesitate to contact me if you or your colleagues have any additional questions or if you would like to discuss these or related matters further.

Sincerely,

Michael D. Goodman, Ph.D.
Associate Professor and Chair
Department of Public Policy
UMass Dartmouth

To: Kathleen Berg, Ph.D., University of Hawaii
From: UMass Donahue Institute, Economic and Public Policy Research
Re: Request for Information Regarding Hawaii Effectiveness Index Calculations, SY 2008-2009
Date: July 26, 2011

In response to Question #2 of the e-mail request for information sent on July 15, 2011, the following analysis of variance tables are provided below for the Math and Reading Hawaii State Assessment tests in grades 3, 5, 8 and 10 for SY 2008-2009. Above the table, the STATA command for the regression that was used is also included. Please contact Kate Wilkinson, at kwilkinson@donahue.umassp.edu or 774.455.7378, with any additional questions on this analysis.

Hawaii "Effectiveness Index" Calculations, SY 2008-2009
Analysis of Variance Tables
UMass Donahue Institute

Grade 3 Math

```
. reg  mathps3 disadv if  mathpscnt3>10

      Source |       SS       df       MS              Number of obs =     188
-------------+------------------------------           F(  1,   186) =   64.23
       Model |  1.34991857     1  1.34991857            Prob > F      =  0.0000
    Residual |  3.90941312   186   .02101835            R-squared     =  0.2567
-------------+------------------------------           Adj R-squared =  0.2527
       Total |  5.25933168   187  .028124768            Root MSE      =  .14498

------------------------------------------------------------------------------
     mathps3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      disadv |  -.3888634   .0485224    -8.01   0.000    -.4845884   -.2931383
       _cons |   .6611768   .0257619    25.66   0.000     .6103537        .712
------------------------------------------------------------------------------
```

Grade 3 Reading

```
. reg  readps3 disadv if  readpscnt3>10

      Source |       SS       df       MS              Number of obs =     188
-------------+------------------------------           F(  1,   186) =  179.63
       Model |  2.32277537     1  2.32277537            Prob > F      =  0.0000
    Residual |  2.40516866   186  .012931014           R-squared     =  0.4913
-------------+------------------------------           Adj R-squared =  0.4886
       Total |  4.72794403   187  .025283123           Root MSE      =  .11371

------------------------------------------------------------------------------
     readps3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      disadv |  -.5100901   .0380592   -13.40   0.000    -.5851732   -.4350069
       _cons |    .863606   .0202067    42.74   0.000     .8237422    .9034698
------------------------------------------------------------------------------
```

Grade 5 Math

```
. reg  mathps5 disadv if  mathpscnt5>10

      Source |       SS       df       MS              Number of obs =     192
-------------+------------------------------           F(  1,   190) =   54.74
       Model |  1.13518677     1  1.13518677            Prob > F      =  0.0000
    Residual |  3.94001033   190  .020736896           R-squared     =  0.2237
-------------+------------------------------           Adj R-squared =  0.2196
       Total |   5.0751971   191  .026571713           Root MSE      =    .144

------------------------------------------------------------------------------
     mathps5 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      disadv |  -.3552273   .0480114    -7.40   0.000    -.4499312   -.2605234
       _cons |   .6174613   .0253693    24.34   0.000     .5674195     .667503
------------------------------------------------------------------------------
```

Grade 5 Reading

```
. reg  readps5 disadv if  readpscnt5>10

      Source |       SS       df       MS              Number of obs =     192
-------------+------------------------------           F(  1,   190) =  150.13
       Model |  1.99498962     1  1.99498962            Prob > F      =  0.0000
    Residual |  2.52479236   190  .013288381           R-squared     =  0.4414
-------------+------------------------------           Adj R-squared =  0.4385
       Total |  4.51978198   191   .02366378           Root MSE      =  .11528

------------------------------------------------------------------------------
     readps5 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      disadv |  -.4709155   .0384334   -12.25   0.000    -.5467265   -.3951046
       _cons |   .8407569   .0203083    41.40   0.000     .8006983    .8808156
------------------------------------------------------------------------------
```

Grade 8 Math

```
. reg  mathps8 disadv if  mathpscnt8>10

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(  1,    70) =   10.68
       Model |  .228949605     1  .228949605            Prob > F      =  0.0017
    Residual |  1.50105245    70  .021443606           R-squared     =  0.1323
-------------+------------------------------           Adj R-squared =  0.1199
       Total |  1.73000205    71  .024366226           Root MSE      =  .14644

------------------------------------------------------------------------------
     mathps8 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      disadv |  -.2898424   .0887035    -3.27   0.002     -.466756   -.1129288
       _cons |   .4666535   .0456227    10.23   0.000     .3756619    .5576451
------------------------------------------------------------------------------
```

Grade 8 Reading

```
. reg  readps8 disadv if  readpscnt8>10

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(  1,    70) =   22.77
       Model |  .447609641     1  .447609641           Prob > F      =  0.0000
    Residual |  1.37607133    70  .019658162           R-squared     =  0.2454
-------------+------------------------------           Adj R-squared =  0.2347
       Total |  1.82368097    71  .025685647           Root MSE      =  .14021

------------------------------------------------------------------------------
     readps8 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      disadv |  -.4052674   .0849304    -4.77   0.000    -.5746558    -.235879
       _cons |   .8484528   .0436821    19.42   0.000     .7613316     .935574
------------------------------------------------------------------------------
```

Grade 10 Math

```
. reg  mathps10 disadv if  mathpscnt10>10

      Source |       SS       df       MS              Number of obs =      56
-------------+------------------------------           F(  1,    54) =   40.81
       Model |   .54128737     1   .54128737           Prob > F      =  0.0000
    Residual |  .716267419    54  .013264211           R-squared     =  0.4304
-------------+------------------------------           Adj R-squared =  0.4199
       Total |  1.25755479    55  .022864633           Root MSE      =  .11517

------------------------------------------------------------------------------
    mathps10 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      disadv |  -.5064759    .079284    -6.39   0.000    -.6654307    -.347521
       _cons |   .5079096   .0369668    13.74   0.000     .4337956    .5820235
------------------------------------------------------------------------------
```

Grade 10 Reading

```
. reg  readps10 disadv if  readpscnt10>10

      Source |       SS       df       MS              Number of obs =      56
-------------+------------------------------           F(  1,    54) =   36.98
       Model |  .384926044     1  .384926044           Prob > F      =  0.0000
    Residual |  .562112708    54  .010409495           R-squared     =  0.4065
-------------+------------------------------           Adj R-squared =  0.3955
       Total |  .947038752    55  .017218886           Root MSE      =  .10203

------------------------------------------------------------------------------
    readps10 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      disadv |  -.4271038    .070236    -6.08   0.000    -.5679185   -.2862892
       _cons |   .9133504   .0327481    27.89   0.000     .8476944    .9790063
------------------------------------------------------------------------------
```