Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 16, Number 13, October 2011 ISSN 1531-7714

Test Score Reporting Referenced to Doubly-Moderated Cut Scores Using Splines

William D. Schafer and Xiaodong Hou
University of Maryland

This study discusses and presents an example of a use of spline functions to establish and report test scores using a moderated system of any number of cut scores. Our main goals include studying the need for and establishing moderated standards and creating a reporting scale that is referenced to all the standards. Our secondary goals are to make possible straightforward interpretations about growth, and to report to users, scores that capitalize on their existing beliefs. Data from one state are used as an example to demonstrate how a complete system of cut scores might be developed and implemented.

In a typical state (or other application, such as National Assessment of Educational Progress) cut scores for proficiency (e.g., achievement) levels are developed through panel recommendations and may or may not be modified before finalized. Yet it is well-known that process using panel recommendations can and often do result in impacts that may not be very satisfying. For example, it has been shown that within states, achievement levels for different grade levels within the same content show marked and even inconsistent patterns of differences in rates of students achieving what has been called "proficiency." These patterns are said to show poor vertical moderation (Lissitz & Huynh, 2003). It has also been shown that rates are different for different content areas, such as reading and math, showing poor horizontal moderation (Schafer, 2005). Schafer, Liu & Wang (2007) have documented both these effects across states. Yet it is difficult to argue that there is something about reading vs. math that should alter proficiency rates, and the same is true across grade levels. It is not an unreasonable position that these differences are the result of different panels rather than inherent content differences. These inconsistencies can lead to poor policy since decisions are often made on the basis of test scores results. For example, if one grade level (say, sixth) were to have extraordinarily low percents of students proficient in one content area (say, reading), then a state might decide to provide disproportionate resources to sixth grade reading and the detriment of other grade levels and contents. But decisions like that can result from inconsistent standards (cut scores) rather than inherent content differences. Some way, either to justify disparate impacts or to moderate impacts, should be found before determining policy based on comparing percentages of students in proficiency levels for grade levels and content areas.

We outline a method that was used to establish a system of scores that references cut scores that are doubly moderated, horizontally (across content areas; Schafer, 2005) and vertically (across grade levels; Lissitz & Huynh, 2003). Our main goals include establishing moderated standards that have a basis in panel-recommended cut scores and creating a reporting scale that is inherently referenced to all the standards. Our secondary goals are to suggest ways to use the reporting scale to interpret differential growth across persons, contents, and educational units, and to better convey valid inferences to test score users.

Moderation, or consistency of impact is a relatively new way to evaluate state cut score systems

and most states show poor moderation (Schafer, Liu & Wang, 2007), with considerable normative variation in impact rates across grades and across content areas. One result of variation is to overemphasize the effects of some grades and contents over others in school evaluations with the potential for poorly justified corrective actions. We will propose a normative process for adjusting cut scores based on student performance so that they are doubly moderated, yet are based on the degree of idealism (sometimes called rigor) vs. realism represented in original panel recommendations as well as outside considerations. Policy-suggested interpretations may also be built into the moderation process and our example includes some of these, as well.

We also outline a reporting scale that references all the standards (whether or not they have been doubly moderated). Our particular implementation translates easily into letter grades or achievement levels that everyone is familiar with, though that is not a necessary characteristic of such scales. When the standards are doubly moderated, simplified interpretations of student growth can be a desirable by-product of referencing the reporting scale to cut scores. We describe below a process that could be used to evaluate the need for a moderated scale, a way to use existing data to develop adjusted cut scores that are doubly-moderated, and the use of a type of spline functions to generate a redefined scale that references the revised cut scores using a look-up table. We conclude with some considerations that may influence educators in deciding whether to use the approach. We use a volunteer state (Maryland) as an example to demonstrate a completely moderated system of cut scores for the statewide assessments from grade3-8 and high school.

METHODS

Source(s) of the information

Existing large-scale test results are necessary to implement the approach we suggest. Our results were calculated from actual statewide distributions from the state's 2008 main assessments. We also were guided by recent National Assessment of Educational Progress (NAEP) percentile ranks taken from the official NAEP website. We recommend that scales be developed after the assessments have been in place for two or more years so that the results are relatively stable. This is important for two reasons. When new assessments are introduced, standard-setting panels typically make their cut-score recommendations in the absence of reliable impact data since (a) the assessments are usually based on new (or newly revised) content standards that are only beginning to be introduced by educators and (b) the assessments are often given with weakened consequences (perhaps even using no-fault administrations) so that student (and teacher) effort may be atypical until the assessments and the system of accountability they drive are fully implemented. Since the look-up tables to be developed will be used without change each year, it is important that they be based on stable data from typically behaving students and educators.

Evaluation of the Current Cut-Score System

In order to study and document need for moderating cut scores in our example state, tables of the percentile ranks of the various cut scores used in the state at different grade levels and in different contents are presented along with the most recent national and statewide percentile ranks of cut scores on NAEP on all available corresponding tests. We suggest including the NAEP results as an external opportunity for guidance since statewide cut scores have been compared with NAEP's in the literature, often with disappointing results. Any state (or other jurisdiction) using this process will have cut scores already implemented, so our example generalizes easily to other contexts.

In Maryland, there were assessments given in grades three through eight in reading and mathematics, in grades five and eight in science, and in high school as end-of-course tests in biology, English, government, and mathematics. The high school tests were required for graduation, and a student needed to "pass" each test with a minimum score (this requirement has been deleted by the state; we retain it here because it provides a useful reference for the scale as well as for illustrative purposes) as well as obtain a certain average across the four tests. For accountability purposes, there were three achievement levels, basic, proficient, and advanced, for all reading, mathematics, and science tests, including biology.

Tables 1 and 2 display the percentile ranks of the various cut scores used in Maryland along with the percentile ranks of cut scores on National Assessment of Educational Progress (NAEP) on corresponding tests. The percentile ranks associated with the Maryland cuts were calculated from actual statewide

distributions from the 2008 main assessment; we are grateful to the Maryland State Department of Education (MSDE) for sharing these results with us.

Table 1: Percentile Ranks of 2008 Maryland Cut Scores

	Reading		Math		Science		Govern-ment	
Grade	B/P	P/A	B/P	P/A	B/P	P/A		
3	19	88	18	73				
4	14	72	12	59				
5	13	49	20	76	37	92		
6	18	63	24	70				
7	19	57	32	79				
8	27	66	38	72	39	96		
	Min	Avg	Min	Avg	Min	Avg	Min	Avg
HS	28	40	28	39	25	34	23	29

B/P: Cut Score between Maryland Basic and Maryland Proficient

P/A: Cut Score between Maryland Proficient and Maryland Advanced

Min: Cut Score for the minimum any test may be for Maryland high school graduation

Avg: Cut Score for the average of the Maryland high school test scores for graduation

HS: High school

The NAEP percentile ranks were taken from the official NAEP website results for the Nation and for Maryland; the then-most-recent data were from 2007 for reading and math and 2005 for science. The NAEP results were included as an external reference and since statewide cut scores have been compared with NAEP's in the literature. Some observations and suggestions follow the results. While these are unique to the state, in other implementations, similar inferences will be suggested by the data. While the idiosyncratic cut scores in this particular state are presented, in each implementation there will be an existing system of cut scores that should be substituted for these; if there are not, then implementation at that time is premature since the data will be atypical.

We note that there is considerable variation in impacts across grades and content areas for the Maryland cut scores. The percentage proficient and above varies from a high of 88 (grade 4 math) to a low of 61 (grade 8 science); the percentage advanced varies from a high of 51 (grade 5 reading) to a low of 4 (grade 8 science). We concluded that they are not very well moderated, which is not surprising since moderation is a relatively new way to evaluate state cut score systems. Indeed, the great majority of states show poor moderation; see Schafer, Liu, & Wang (2007).

Table 2: Maryland Percentile Ranks of NAEP Cut Scores

	Reading			Math			Science		
	2007 National Percentile Ranks								
Grade	Bas	Pro	Adv	Bas	Pro	Adv	Bas	Pro	Adv
4	33	67	92	18	61	94	32	71	97
8	26	69	97	29	68	93	41	71	97
	2007 Maryland Percentile Ranks								
4	31	64	90	20	60	92	36	74	98
8	24	67	97	26	64	90	46	74	96

Science scores are from 2005

Bas: Cut Score between NAEP Below Basic and NAEP Basic

Pro: Cut Score between NAEP Basic and NAEP Proficient

Adv: Cut Score between NAEP Proficient and NAEP Advanced

One effect of the variation is to overemphasize the effects of some grades and contents over others in school evaluations. For example, it is somewhat more difficult to achieve Proficient in math than in reading, grades 5-8, so more schools are identified for math than for reading, which overemphasizes math in school accountability at those grades; the reverse holds for grades 3-4, however. At the high school level, math and English are normatively more difficult than biology and government for both cut scores used for graduation decisions. Although this statewide system was developed using the best information at the time the original cut scores were established, we suggest that the system is difficult to justify in the light of the results in Table 1. Similar issues exist in virtually every state (Schafer, Liu, & Wang, 2007) and likely most other programs that are considering our suggestions.

Development of the proposed cut scores

In order to balance the impacts of grades and contents, a revised statewide system was developed based on moderated cut scores using equivalent percentiles. The averages of the percentile ranks (based on 2008 data) for the same cut for grades 3-8 in all content areas were used to provide guidance in the development of the cut scores to be proposed. We discuss in this section, cuts that capitalize on

consistencies among these average percentile ranks, the relations of those cuts to NAEP data, and some policy concerns unique for but important to the state (such as an interest in interpreting early grade results in relation to eventual high school performance). Policy considerations will be idiosyncratic to various states and other jurisdictions, but considerations are likely to parallel those in our example.

Policy considerations and development of the reporting scale

Since Maryland based graduation decisions in part on test performance, we considered interpretations of score reports at all levels in relation to passing cuts for high school. We also found it convenient to express the various score ranges in the familiar terms of letter grades, A through F. For high school graduation, we capitalized on two significant cuts in the state: (1) each content must be at or above a minimum score and (2) the average must be at or above a higher minimum. For lower grades, the same percentile rank for the same cut would be used for grades 3-8 in all three contents. Table 3 shows the average percentile ranks that correspond to the various cut scores in Tables 1 and 2, ordered from low to high.

Table 3: Ordered Average Percentile Rank across All Available Grades (3-8; High School) and Contents; 2008 data

Cut Score	PR
Maryland High School Min	26
Maryland 3-8 Basic/Proficient	27
National NAEP Below Basic/Basic	30
Maryland NAEP Below Basic/Basic	31
Maryland High School Avg	36
Maryland NAEP Basic/Proficient	67
National NAEP Basic/Proficient	68
Maryland 3-8 Proficient/Advanced	77
Maryland NAEP Proficient/Advanced	94
National NAEP Proficient/Advanced	95

Since we are considering interpretations of score reports at all grade levels in relation to existing passing cuts for high school in the state, we found it convenient to express score ranges in the familiar terms of letter grades (A through F) which have interpretations consistent with the bullets below. While arbitrary, this decision has some intriguing advantages that we will discuss later. Considerations we used in suggesting which letter grades correspond to which cut scores for this state appear below. Unique considerations in any other state or application may lead to other criteria.

- the cut for D at the high school level should be at the minimum score for graduation for any one test (i.e., F is failure)
- the cut for C at the high school level should be at the average cut required for graduation across all tests
- There should be parallel cuts for the above two letter grades at each of the grade/content combinations in grades 3-8.
- the cut for A should correspond to the cut for Advanced for accountability reporting purposes
- the cut for B should be established to be close to the NAEP cut for Basic/Proficient, but also so that it allows a reasonable range on either side for differentiation.

It seems desirable also that the letter grades be associated with a familiar score scale used nationally. One such scale that at once is very popular and seems reasonable for our purpose is to associate 59 and below with F (we suggest that the minimum possible score, called the lowest obtainable scale score, or LOSS on the current scales be set at 50), 60-69 with D, 70-79 with C, 80-89 with B, and 90-100 with A (100 would correspond to the maximum possible on the current scale, called the highest obtainable scale score, or HOSS). These values would need to be related to the current reporting scale, but once the conversions are fixed, they would be very easy to accomplish in the future through the re-use of the look-up table at each grade/content combination. Note that the cut percentile ranks are drawn from Table 2 and follow from the bulleted criteria, except for the Intermediate cut PR of 56, which is the average of 36 and 77, the Cut PR's surrounding it.

These considerations led us to the criteria to be implemented outlined in Table 4.

Table 4: Criteria to be implemented

Cut	Grade Cut	Cut PR	Score Range	2008%	Possible Category Label
Below Min.	F	N/A	50 - 59	26	Below Basic
Passing Content	D	26	60 – 69	10	Basic
Passing Average	C	36	70 – 79	20	Proficient
Intermediate	B	56	80 – 89	21	Highly Proficient
Advanced	A	77	90 – 100	23	Advanced

Smoothing method

In order to achieve all these goals, it is necessary for each content-grade combination to generate a smoothed, monotonic function that passes through the various scale scores x (horizontal axis) associated with the percentiles in the table as they pass through the lower values y in each score range (vertical axis) in order to map the existing scale (240 to 650) to the new scale (50-100). We used monotonic cubic Hermite spline functions to develop the conversions. These splines produce smoothed results that pass through all the points (x_i, y_i) entered into the procedure and do not change directionality, where $i = 1, \dots, n$; n is the number of points which are entered into the procedure (n = 6 in the current study).

Suppose $k = 1, \dots, n-1$, for the interval $(x_k,\ x_{k+1})$, the cubic Hermite spline can be defined as:

$$f_k(x) = h_{00}(t)y_k + h_{10}(t)(x_{k+1} - x_k)m_k + h_{01}(t)y_{k+1} + h_{11}(t)(x_{k+1} - x_k)m_{k+1}$$

where $t = \frac{x - x_k}{x_{k+1} - x_k}$, and h is the basis functions, which can expressed as

$$h_{00}(t) = (1 + 2t)(1 - t)^2,$$
$$h_{10}(t) = t(1 - t)^2,$$
$$h_{01}(t) = t^2(3 - 2t),$$
$$h_{11}(t) = t^2(t - 1).$$

There are several methods of selecting the tangent m to maintain the monotonicity of the Hermite spline function. Our study followed Fritsch-Carlson (Fritsch & Carlson, 1980) method described in the following steps.

1. The slopes of the lines linking two consecutive points are calculated as:
   $\Delta_k = \frac{y_{k+1} - y_k}{x_{k+1} - x_k}$ for $k = 1, \dots, n-1$.
2. The initial tangents m at every data point are:
   $\mathrm{m}_k = \frac{\Delta_{k-1} + \Delta_k}{2}$ for $k = 2, \dots, n-1$; $\mathrm{m}_1 = \Delta_1$ and $\mathrm{m}_n = \Delta_{n-1}$
   These initial tangents may be updated in following steps.
3. When $y_k = y_{k+1}$ for $k = 1, \dots, n-1$, then $m_k = m_{k+1}$ is set to zero to make the spline flat in order to preserve monotonicity of the function. Steps 4 and 5 for those k will be skipped.
4. Let $\alpha_\mathrm{k} = \frac{\mathrm{m_k}}{\Delta_\mathrm{k}}$ and $\beta_k = \frac{m_{k+1}}{\Delta_k}$. If α or β is zero (i.e. the input data points are not monotone), then m_k and m_{k+1} are set to be zero to ensure that the piecewise monotone curves can still be generated. This step is for the non-monotone data points, which is impossible in the application we suggest.
5. In order to achieve monotonicity of the Spline function, if $\alpha_{\mathrm{k}_k}^2 + \beta_\mathrm{k}^2 > 9$, then set $m_k = \tau_k \alpha_k \Delta_k$ and $m_{k+1} = \tau_k \beta_k \Delta_k$ where $\tau_k = \frac{3}{\sqrt{\alpha_{\mathrm{k}_k}^2 + \beta_\mathrm{k}^2}}$.

Although only one example is presented below, we developed an independent spline for each grade-content combination. Graphs of the spline functions and the associated look-up tables in the example state have all been developed and appear in Schafer, Hou, and Lissitz (2009), which is available electronically. The graphs reveal that the conversions are close to linear in regions that do not involve either LOSS or HOSS and flatten as they approach either extreme, which we view as an advantage. The look-up tables (also available electronically through a link in Schafer, et al., 2009) color-code the current cuts as a basis for comparison.

The following example is for grade 3 math. The proposed cut percentile ranks for this grade-content from the above table were used to obtain their

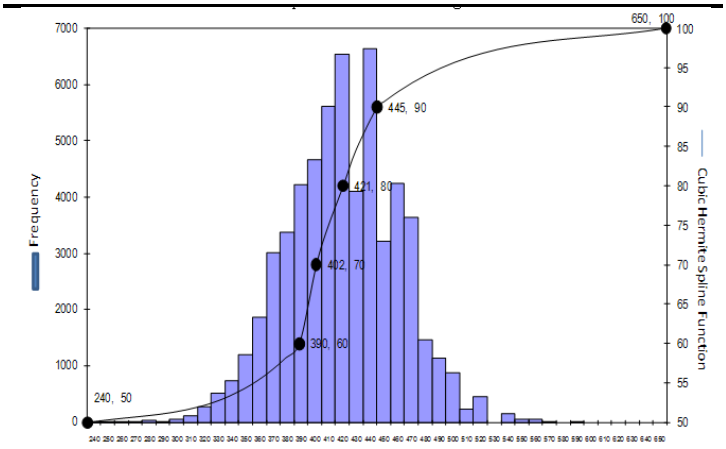
corresponding cut percentiles; the last column are the arbitrary cut scores used to reference the reporting scale to the moderated achievement levels. The values LOSS (240) and HOSS (650) are the lowest and highest obtainable scale scores. These were chosen arbitrarily for all Maryland tests and are thus boundary percentiles; they are included here because they were entered into the spline function.

Table 5: Percentiles and cut scores for six cut points.

Cut Point	Cut Percentile x_i	Cut Score y_i
LOSS	240	50
D/F	390	60
C/D	402	70
B/C	421	80
A/B	445	90
HOSS	650	100

A graph of the spline function is also shown and demonstrates fit to the six points used as input to the spline process and bolded in the graph. For context, the histogram also shows the 2008 statewide distribution for this grade-content.

Figure 1: Cubic Hermite Spline and Histogram



The following section of the corresponding look-up table suggests what the process of conversion to the splines looks like and how it would be implemented in practice (the values in red are targets that were used as input to the splines). The first column, Grade 3, corresponds to the graph above.

Table 6: Part of the look-up table

50-100 scale / Current scale	Math					
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
650	100	100	100	100	100	100
649	99.9703	99.9677	99.9674	99.9659	99.9662	99.9664
648	99.9411	99.936	99.9354	99.9325	99.9331	99.9334
…	…	…	…	…	…	…
456	91.4775	90.2955	90.1494	90	91.1021	90.7124
455	91.3511	90.1486	90	89.8304	90.9702	90.5734
454	91.2231	90	89.8292	89.6245	90.8366	90.4327
…	…	…	…	…	…	…

The moderated system of cut scores appears to have distinct advantages over what is currently done elsewhere:

- the cuts do not over- or under-emphasize any content area or grade, defined normatively, a clear drawback in the large majority of states (including Maryland).
- the transformation is nearly linear for approximately half the students in the central regions of each of the distributions (easily noted in the graphs for all content-grade combinations).
- differences among scores at the extremes are minimized (note how the curve in the graph flattens out at either end), precisely where the standard errors of measurement are greatest (as the technical manuals of virtually all states show) and thus where interpretations of score differences should be made most cautiously (The widths of conditional standard-error bands associated with extreme scores would thus be reduced somewhat, which might make their use more appealing, perhaps even for sub-scores.).
- they facilitate a notion of expected growth in that, normatively speaking; one would expect one year's growth in any content area should place a student in the same place, relative to the cut scores, as he or she was the prior year (similar, but not identical to the recommendation in Schafer, 2006, in which a

zero-to-100 scale was prefaced with the grade level; i.e., 350 and 450 would be the same relative to the cuts in the two grades, although subtracting 50 from the proposed system and multiplying by two would accomplish his suggestion, in which a change of 100 would equal one-year's growth).

- they are easily translated into letter grades (achievement levels) that have meaning in terms of graduation decisions or comparable levels of performance.
- they are expressed in terms that facilitate reasonable interpretations on the part of anyone at all familiar with American education (e.g., they could be described as "curved" results, expressed on a "percentage" scale, which resembles how a naïve user interprets curved percent-correct scores developed by teachers, and these are actually defensible interpretations for the suggested scale).
- they allow the traditional computation of grade-point-averages across students and across schools, which have face validity for educational decision making in many other contexts.
- there is no change in the technical properties of the assessments since they would continue to be developed using established scale scores; the conversions to the reporting scale would always be done as a last step and quite easily since the existing tables would just be re-used.
- they are informed by actual student results on tests taken under the conditions for which the tests are to be used in the future, unlike impact data presented to panels in standard-setting studies for the original cuts (e.g., "no-fault" administrations).
- their achievement levels could be interpreted using released items, associated with their RP67 scale positions on the 50-100 scale [locations where two-thirds of examinees would be expected to respond correctly; see Huynh (1998)]. That would allow interpretation of achievement levels (e.g., achievement level descriptions) directly in terms of actual items to which students responded. This is not new, and could be used to elaborate any existing system of achievement level descriptions.].
- the look-up tables facilitate historical calculation of NCLB criteria for school-level, district-level, and state-level decision-making, so that past trends can be expressed on the 50-100 metric; although we would not recommend that schools be identified for historically based sanctions, it seems reasonable to remove sanctions from schools if they no longer would have been identified.

Drawbacks to implementing the system include aversion to change, allegiance to the original standard-setting process, and expense, including resources for both programming the extra steps and reports, and public education.

CONCLUSIONS

We have described a complete system that could be implemented immediately, as long as an existing scale exists and users feel its percentiles are well estimated. In order to develop a rich example, however, we made several decisions that might not be best from the point of view of any other state or jurisdiction, or perhaps even Maryland. Modifications are certainly possible and fortunately most can be accomplished easily, such as setting different percentiles than we used for the fixed scores on the 50-100 scale. If the latter were done, it would be straightforward to develop revised splines and look-up tables.

The method we are suggesting is actually a criterion-referenced one, but with a normative basis. Being "advanced" may be approximately in the top quartile in year one, but in subsequent years, more and more students may place in the "advanced" range since the spline functions are not re-estimated, but simply re-used. The normative basis comes from their initial estimation, but once established, they become fixed criteria. Achievement level descriptions may need some modifications, or they may be described through exemplary released items, which their locations on the 50-100 scale similarly released.

We feel the underlying goals we have had in our work, a moderated cut-score system and a transparent method of score reporting that is referenced to the moderated cuts, are desirable and attainable outcomes. In any given state or other jurisdiction, the achievement levels currently in place, existing data, and different policy goals would lead to other sets of points to enter into the spline process, but our suggestions seem to generalize easily to a wide array of possibilities. In

addition, consideration by both technical experts and user constituencies is certainly appropriate in any application.

Any state or other jurisdiction considering this approach might try working up reports in both their current and the new formats and using them in conducting focus groups to generate suggestions about whether and how to proceed. In our view, consistency available in a doubly-moderated system and transparency in reporting using a familiar scale are advantages that are difficult to forego.

REFERENCES

Fritsch, F. N. & Carlson, R. E. (1980). Monotone Piecewise Cubic Interpolation *Journal on Numerical Analysis (SIAM) 17* (2), 238–246.

Huynh, H. (1998). On score locations of binary and partial credit items and their applications to item mapping and criterion-referenced interpretation. *Journal of Educational and Behavioral Statistics, 23* (19), 35-56.

Lissitz, R. W. & Huynh, H. (2003). Vertical equating for state assessments: Issues and solutions in determination of adequate yearly progress and school accountability. *Practical Assessment, Research, and Education, 8,* 1-10. Retrieved May 30, 2010, from http://pareonline.net/getvn.asp?v=8&n=10.

Schafer, W. D. (2005). Criteria for standard setting from the sponsor's perspective. *Applied Measurement in Education*, *18* (1), 61-81.

Schafer, W. D. (2006). Growth s*cales as an alternative to vertical scales. Practical Assessment, Research and Evaluation, 11,* 1-6. Retrieved May 30, 2010 from http://pareonline.net/pdf/v11n4.pdf.

Schafer, W. D., Hou, X., & Lissitz, R. W. (2009). *Consideration of test score reporting based on cut scores.* Retrieved February 17, 2011 from http://www.marces.org/completed/2009Project%20Report.doc

Schafer, W. D., Liu, M. & Wang, H (2007). Content and Grade Trends in State Assessments and NAEP. *Practical Assessment Research & Evaluation*, *12* (9). Retrieved May 30, 2010 from http://pareonline.net/pdf/v12n9.pdf.

Citation:

Schafer, William D & Xiaodong Hou (2011). Test Score Reporting Referenced to Doubly-Moderated Cut Scores Using Splines. *Practical Assessment, Research & Evaluation*, 16(13). Available online: http://pareonline.net/getvn.asp?v=16&n=13.

Acknowledgement

This paper was partially funded by the Maryland State Department of Education (MSDE) through the Maryland Assessment Research Center for Education Success (MARCES) at the University of Maryland. The opinions expressed are those of the authors and not necessarily those of MARCES or MSDE. The authors are indebted to MSDE for sharing the statewide assessment results necessary to complete the work described here.

Authors:

William D. Schafer is Affiliated Professor (Emeritus), Maryland Assessment Research Center for Education Success, Department of Measurement, Statistics, and Evaluation, University of Maryland, College Park, MD. He specializes in assessment and accountability.

Xiaodong Hou is a Ph.D. Candidate in the Department of Measurement, Statistics, and Evaluation, University of Maryland, College Park, MD. She specializes in quantitative research methods.