

Setting Performance Standards for the VAL-ED Assessment of Principal Leadership

Andrew Porter – University of Pennsylvania
Ellen Goldring – Vanderbilt University
Steve Elliott– Vanderbilt University
Joseph Murphy – Vanderbilt University
Morgan Polikoff – University of Pennsylvania
Xiu Cravens – Vanderbilt University

August 29, 2008

The Vanderbilt Assessment of Leadership in Education (VAL-ED) is a 360 degree assessment of school principal leadership developed during the period 2006 through 2008 with Wallace Foundation support. The instrument is based on a conceptual framework of six core components crossed with six key processes. The conceptual framework was extracted from the literature on school leadership that improves student academic achievement (Murphy, et al, 2006; Goldring, et al, 2007; Porter, et al, 2006). The core components are characteristics of effective schools: high standards for student learning, rigorous curriculum (content), quality instruction (pedagogy), culture of learning and professional behavior, connections to external communities, and performance accountability. The key processes are behaviors that principals can employ to realize the six core components for their school: planning, implementing, supporting, advocating, communicating, and monitoring. As shown in figure 1, the conceptual framework identifies 36 combinations of core components and key processes or cells that identify unique types of behaviors an effective school principal would exhibit.

The VAL-ED assessment is a paper and pencil or online assessment consisting of two forms, A and C, each containing 72 items, two items per cell. The forms are randomly equivalent sets of items stratified by cell and designed to be parallel. The 360 degree assessment surveys the principal, the principal's supervisor, and all of the teachers in the principal's school. In each case, the respondent rates the principal on a five-point scale from ineffective to outstandingly effective for each of the 72 behaviors identified.

The VAL-ED is designed, developed, and tested to be both reliable and valid. A series of pilot studies, fairness reviews, cognitive interviews, and other psychometric evaluations have been completed in the process of creating the two parallel forms of the instrument. In spring 2008, a national field trial was conducted using the paper and pencil version of both forms A and C.

The results of the principal assessment are reported in terms of profiles on the mean item response scale (ratings 1 through 5) by respondent group, by individual core components, and by individual key processes. In addition, results are reported in percentile ranks according to the results of the national field trial. Finally, the results are to be reported in terms of performance levels: distinguished, proficient, basic, and below basic.

In August of 2008, a standard setting panel of 22 individuals was convened to set the performance standards for the VAL-ED.

The Performance Level Descriptors

The project research team wrote performance level descriptors (PLD's) for each proficiency level. After multiple iterations, the proficiency level descriptors were critiqued by seven educators, including two supervisors of principals, three principals (one each from elementary, middle, and high schools), and two teachers (one each from elementary and high school). The educators were from five different states. The critiques were done individually by each of the seven.

Packets to the seven educators included unlabeled proficiency level descriptors and a copy of the VAL-ED instrument with the framework and definitions. All respondents were able to correctly sort the unlabelled PLDs into the correct proficiency levels. The seven educators agreed the proficiency level descriptions distinguished and differentiated among the performance levels. Based on feedback, the original proficiency level descriptors were modified slightly. The final proficiency level descriptors are as follows:

- Below Basic – A leader at the below basic level of proficiency exhibits leadership behaviors of core components and key processes at levels of effectiveness that over time are unlikely to influence teachers to bring the school to a point that results in acceptable value-added to student achievement and social learning for students.
- Basic – A leader at the basic level of proficiency exhibits leadership behaviors of core components and key processes at levels of effectiveness that over time are likely to influence teachers to bring the school to a point that results in acceptable value-added to student achievement and social learning for some sub-groups of students, but not all.
- Proficient – A proficient leader exhibits leadership behaviors of core components and key processes at levels of effectiveness that over time are likely to influence teachers to bring the school to a point that results in acceptable value-added to student achievement and social learning for all students.
- Distinguished – A distinguished leader exhibits leadership behaviors of core components and key processes at levels of effectiveness that over time are virtually certain to influence teachers to bring the school to a point that results in strong value-added to student achievement and social learning for all students.

The Task

Setting proficiency levels is a relatively new activity within educational assessment and is typically done for student achievement testing. Over time, different procedures have been identified, including the Bookmark procedure, the Angoff procedure, the Jaeger-Mills method, and the Contrasting Groups method (Green, Trimble, and Lewis, 2003). The Bookmark method has emerged as the most popular method for setting proficiency levels, at least on student achievement testing. Perhaps the popularity of the method stems from its being straight forward and data based.

Setting standards for the VAL-ED is different from standard setting on student achievement tests, however, in three ways. First, the VAL-ED is an assessment of school principals at the elementary, middle, and high school level. Second, the items on the VAL-ED do not have dichotomous answers, as is typical for multiple choice items on a student achievement test; rather, the responses are on a five-point effectiveness scale with five being the highest rating and one the lowest. Further, there are three response groups assessing each principal.

After conferring with the project's panel of practitioner experts, we decided that one set of performance standards should be established that applies to all principals, regardless of their level of schooling (elementary, middle, or high school) and the location of their school (urban, suburban, or rural). Further, we decided to set performance standards based on all of the information across all 72 items and all three response groups. We wanted the simplicity of one set of standards based on the best principal performance data available. Thus, the task for the standard setting panel convened in August was to determine cut scores to differentiate distinguished from proficient, proficient from basic, and basic from below basic on the effectiveness item response scale, ranging from 1.0 to 5.0.

An approach to this task was identified using the Bookmark procedure. The approach was put out for critique by three experts in setting proficiency standards, Gary Phillips at the American Institutes for Research, Steve Ferrara at CTB, and Greg Cizek at the University of North Carolina. Further, Steve Ferrara, who has used the Bookmark procedure in setting standards on student achievement successfully on many occasions, was recruited to run the standard setting event. While these experts were drawn upon in the process of determining the procedure to be followed and implementing it, any errors of a conceptual or implementation nature are not the fault of the experts upon whose advice we drew, but rather the research team of the project.

Standard Setting Panel Composition

We targeted 22 panelists: 10 principals, 4 teachers, 4 supervisors of principals, 2 researchers of school leadership, and 2 education policymakers. In all cases, we sought experts. For teachers and principals, we sought a distribution across levels of schooling. Finally, we wanted a national panel.

The obtained and convened panel matched in composition the target: 10 principals, 4 teachers, 4 supervisors, 2 leadership researchers, and 2 policymakers. Panelists came from nineteen different states and the District of Columbia. Six panelists were from the South, five from the West, eight from the East, and three from the Midwest. Ten panelists were female. The panel included 1 Hispanic and 6 African Americans, as well as fifteen Whites. The distribution across levels of schooling for principals was three high school, three middle school, and four elementary. For teachers, the distribution was one high school, one middle school, and two elementary.

Expertise is an illusive concept. Of our ten principals, four were recent recipients of Principal of the Year designation for their state. An additional two were designated Distinguished Principal by the American Association of School Administrators (AASA). One was a principal of the year finalist (for their state) and two were principals in National Association of Secondary School Principals Breakthrough Schools. Among the four teachers, one was recipient of the Middle School National Teacher of the Year award, one was a recent State Teacher of the Year, another was a Distinguished Distributive Leadership Teacher, and yet another was recognized as an Outstanding Educator in their district. The four principal supervisors were all district superintendents with an AASA designation of excellence. Of the two policymakers, one was at the state

level and the other with a large school district. The two researchers are both well-known contributors to the research literature on school leadership.

Standard Setting Process

We elected to use a Bookmark method for setting the performance levels of distinguished, proficient, basic, and below basic. Defining the four performance levels required setting three cuts. For a Bookmark method, panelists work their way through a booklet of items, one item per page, where the items are ordered according to difficulty. The task is for each panelist to independently place a bookmark in the booklet on the item that is appropriate in content and difficulty to be the easiest item that a just proficient (or just distinguished or just basic) principal would perform well on. Each item represents a behavior from one of the 36 cells defined by the 6 by 6 conceptual framework. There are 72 items on a form of the assessment, 2 per cell. The item-ordered booklet consisted of the 72 items for Form A. The decision to use Form A was based on the need to make the task for panelists manageable, the fact that Forms A and C were constructed to be paralleled by randomly assigning two items to each form for each of the 36 cells in the conceptual framework and the findings from the national field trial that the two forms have nearly identical means and standard deviations.

To order the items in difficulty, an aggregate variable was created. The aggregate variable was defined as the arithmetic mean of an item's mean item response across the principal, the supervisor, and the mean for the teachers in the principal's school. The principal, the supervisor, and the mean of the teachers were thus equally weighted in creating the aggregate variable. The variable could range from 1 to 5, representing the levels on the effectiveness rating scale, where a 1 indicates "ineffective;" 2, "minimally effective;" 3, "satisfactorily effective;" 4, "highly effective;" and 5, "outstandingly effective."

Item means on the aggregate variable ranged from a low of 3.18 for the most difficult item to 4.04 for the easiest item. The distribution of schools on the aggregate variable ranged from 2.57 for the lowest-rated principal to 4.51 for the highest-rated principal. The range for schools was larger than the range for the item means, as might be expected. Means are measures of central tendency.

The data for the item-ordered booklet comes from a national field trial completed in the spring of 2008 through early summer of 2008. The goal was to recruit 300 schools: 100 high schools, 100 middle schools, and 100 elementary schools. The schools were to be distributed geographically. We targeted 150 urban schools, 100 suburban schools, and 50 rural schools. For the 150 urban schools, 50 were to be drawn from Wallace Foundation grantee districts, 50 drawn from Wallace grantee states but not the districts, and 50 urban schools drawn from non-Wallace grantee districts and states. The process involved randomly selecting districts with a probability in proportion to number of students. Within districts, elementary, middle, and high schools were randomly selected. When a school or district declined participation, a new district or school was randomly selected. Districts and schools were contacted directly by a member of the research team.

Ninety-nine districts were randomly selected and contacted and 60 elected to participate. Within the 60 participating districts, schools were randomly selected and contacted, 309 said they would participate, and 276 returned at least some data. The obtained sample, including 218 complete response sets (teacher, principal, and supervisor responses), is described in Table 1. Of the 218 complete response sets, 103 were Form A, the form that was used for the Ordered-Item Booklet, and 115 were Form C. On Form A, there were 36 urban schools responding, 43 suburban schools, and 24 rural schools. Elementary schools comprised 38 of the Form A responses; middle schools, 33; and high schools, 32. Of the schools in Wallace districts, 33 responded to Form A. Finally, 20 of the complete Form A responses came from the Midwest, 26 from the West, 26 from the Northeast, and 31 from the South.

The Standard Setting Meeting

The standard setting panel was convened at Vanderbilt University in Nashville, Tennessee, on August 12 and 13, 2008. The agenda for the meeting is found in Appendix A. Day 1 began at 8:30 with welcome and introductions. Andy Porter described the VAL-ED instrument, the national field trial, the purpose of the meeting, and introduced the proficiency level descriptors.

Ellen Goldring described a study of the standard setting process that she wished to conduct together with Xiu Cravens to learn how panelists understood the task and process and their thoughts on the appropriateness and quality of each. The study was voluntary for panelists, who, if they wished to participate, signed an informed consent agreement. All 22 panelists agreed to participate, and as a result were interviewed during and at the end of the process in ways designed to ensure no interruption to the primary task of setting performance standards.

Steve Ferrara, with the assistance of Ellen Goldring, worked with the panel for orientation and training to use the Bookmark standard setting procedure with the VAL-ED. The remainder of the morning of the first day was spent on orientation and training. See Appendix B for the PowerPoint slides used.

The panelists were told that the process consists of three rounds. At the end of each round, participants place bookmarks individually and independently one of another, but after discussion as a group. Panelists were organized into five tables where the composition of panelists at each table was mixed across principals, teachers, supervisors, researchers, and policymakers. Panelists were instructed that the task is to set cut scores on the rating scale that separate the performance levels of distinguished, proficient, basic, and below basic. The panelists then again reviewed and discussed each of the four performance level descriptors.

The first task was for panelists to take the instrument and complete it with a specific principal they know in mind. The next task was to place bookmarks to determine proficient. Panelists were instructed to answer each of two questions for each item:

- What behaviors must a principal exhibit in order to achieve the rating of at least highly effective on this item?
- What makes it more difficult to achieve a rating of at least highly effective on this item than on all previous items in this book?

More specifically, panelists were instructed to “place your bookmark on the page where a just barely proficient principal would be rated at least highly effective on the effectiveness rating scale” (at least highly effective is at least a 4.0 on the 5-point scale). The ordered-item booklet did not present item means because panelists were to focus on the nature of the behaviors. Next, panelists placed their bookmarks for where a just barely distinguished principal would be rated at least highly effective and finally, bookmarks where a just barely basic principal would be rated at least highly effective. Panelists practiced the task and had their questions answered before they undertook round 1. Round 1 was completed in the afternoon of the first day. Rounds 2 and 3 occurred on day 2.

Prior to each round, each panelist signed a form indicating that they were ready to proceed. All panelists signed the form at each point in the process. In each round, panelists worked collaboratively at their table to answer the two questions. They then individually and independently recorded their responses on the item-ordered booklet pages. Rounds 2 and 3 were conducted as was round 1, except rounds 2 and 3 began with interpretation and discussion of feedback from the prior round. Each table received information on where the median of panelists at their table placed the cut for each of the three cuts. They were also given the median cut for the collection of five tables (i.e. the room).

Impact data was given at the beginning of round 2 based on the room’s medians for each of the three cuts in round 1. Impact data is more typically given at the beginning of round 3 than at the beginning of round 2. The decision to give impact data earlier than is typical was because of the unique nature of the Bookmark procedure when used with the VAL-ED and its 5-point rating scale. Typically, the Bookmark procedure is used with student achievement items that have a right or wrong answer or can yield multiple score points. When an item can yield multiple score points, the item is represented once for each score point in the item-ordered booklet. As has been described, the items in the item-ordered booklet for the VAL-ED standard setting process were ordered according to each item’s mean on the aggregate variable on the mean item response continuum from 1 to 5. Had a panelist set his or her bookmark on page 1 for distinguishing below basic from basic, it would have resulted in impact data of 10.6% of the principals in the national field trial being designated as below basic. Further, if a panelist set his or her bookmark on the last page of the booklet, it would have resulted in impact data of 11.9% of the principals being distinguished. Understandably, panelists are reluctant to set their bookmarks on either the first or the last page. As will become clear, while panelists converged within table as to where they set their bookmarks across the three rounds, the median bookmarks did not appear to be affected by the “impact data.” In short, the cuts between basic and below basic, proficient and basic, and proficient and distinguished

didn't change much for the room median from round 1 to round 3 and therefore, neither did the impact data.

Perhaps the fact that impact data appeared to not change panelists' decisions about where to place their cuts is simply reflecting the fact that impact data rarely have much impact on where standards are set by panelists in bookmarking (author's personal experience from sitting on four state technical advisory committees). An additional possibility is that when the national field trial data were described to the panelists, it was made clear that the VAL-ED was being used in a field trial setting, not an operational setting. Perhaps the results from the national field trial are not reflective of what would be the results if the VAL-ED were actually being used to assess principals in ways that might be used for high stakes decisions (reports from the national field trial were sent to principals only, and it was up to the principals receiving the information as to whether they distributed it further).

At the end of the process, panelists discussed how comfortable they were with their recommended bookmark placements as well as implications of the impact data. Results of the evaluation are summarized below, but are quite positive with the exception that 24% of the panelists felt uncomfortable with the possibility that the cut between proficient and distinguished was not sufficiently demanding and/or the cut between below basic and basic was too demanding

Results of the Standard Setting

Results from the three rounds of the standard setting process are captured in Tables 2, 3, and 4. Table 2 is for proficient, table 3 for distinguished, and table 4 for basic. There were five tables operating independently with four or five individuals at each table. Bookmarks were placed on pages in the item-ordered booklet so the medians are the medians by table and for the room of the pages on which bookmarks were placed. Next to the median page is the mean item response on the aggregate variable for the item on that page. The final cuts recommended by the panel are found in the lower right entries of each of the three tables. The cut to distinguish proficient from basic is 3.60, the cut between distinguished and proficient is 3.77, and the cut between basic and below basic is 3.42. The impact data is as follows: 30% of the principals in the national field trial fall below basic; 50% fall below proficient; and 70% fall below distinguished.

To determine the item difficulty associated with the median bookmark page, proficiency was straightforward. The item identified by the median page number had a mean item response on the aggregate variable and that was the cut for proficient. For distinguished, the difference between the proficiency cut and the item identified for the distinguished cut was added to the proficient cut score to obtain the distinguished cut. For example, the proficiency cut was 3.48 and the item difficulty associated with the median bookmark page for the distinguished cut was 3.22. The distinguished cut score was set at $3.48 + (3.48 - 3.22) = 3.74$. The transformation was necessary due to the ordering of items where the items became more difficult as the item means went down. To set the basic cut, the same technique was used.

Inspection of tables 2, 3, and 4 identifies differences among tables for each round for each cut. As can be seen, the differences in median page numbers are greater than the differences in the item means. While for a particular cut, there was some change across rounds within a table, there was very little change across rounds for the room. For most tables, there was very little change between round 2 and round 3.

Table 5 reports the standard error of the cut score for basic, proficient, and distinguished based on round 3 data. The standard error of a cut score is a function of the extent to which the cuts are replicated across the tables. The estimation procedure was taken from Huynh (2003) and is stated in terms of page numbers. Thus, the standard error of the basic cut is 2.58 pages in the item-ordered booklet, the standard error of the proficient cut is 2.74 pages, and the standard error of the distinguished cut is .97 pages. Again, each page represents a single item of the 72 items on the principal assessment. These standard errors are relatively small indicating that the groups of panelists were relatively consistent in their recommendations.

Table 6 presents the standard deviations of bookmark placement by round. The standard deviation is calculated on the page numbers of the cuts set by each of the individual 22 panelists. As can be seen, the standard deviation decreases markedly across rounds at each of the cuts. The standard deviation among individuals for round 3 for basic was 5.14, for proficient 5.42, and for distinguished a small 2.12. The decreasing size of standard deviations across rounds makes clear that the panelists were converging on a recommendation for each cut.

It is possible to show graphically this convergence. Figure 2 shows the cuts across rounds for each of the 22 panelists for proficient, figure 3 for distinguished, and figure 4 for basic. These figures are a graphic representation of the same finding reported in terms of decreasing standard deviations.

Post Standard Setting Reconsideration

Based on this feedback from panelists, a week and a half after the standard setting panel met panelists received an email thanking them once again for participating in the process. The email recognized that everyone was pleased with the cut for proficient or not. The proficient cut was set on the mean item response scale at 3.60 and with impact data of 50% of the principals in the national field trial proficient. The email went on to recognize some concerns expressed about the other two cuts. The email stated:

Setting performance standards on an assessment is part science and part art. In all cases, when a panel of experts sets proficiency cuts, they are recommendations. It's not at all uncommon that those recommendations are modified to some extent, for example, by a state's school board when setting proficiency cuts on student achievement tests. We have come up with two options for slightly modifying the cuts you recommended for the distinction between basic and below basic and between proficient and distinguished. The cuts you recommended were 3.42 between basic and below basic, yielding 30% of the principals in the National Field Trial

below basic, and 3.77 for the cut between proficient and distinguished, yielding 30% of the principals in the National Field Trial distinguished. We consider this Option 1. In no order of preference, we have two additional options for you to consider:

- Option 2. Set the cut between basic and below basic at 3.29 and the cut between proficient and distinguished at 3.87. This would yield impact data of 17% of the principals below basic and 22% of the principals distinguished. The rationale for the numbers 3.29 and 3.87 is that of the five tables in the process, the table that set the lowest cut for the distinction between basic and below basic set it at 3.29 and the table that set the highest cut for the distinction between proficient and distinguished set the cut at 3.87. Thus, Option 2 goes with your recommendations, but just for one table, rather than the average across all five tables.
- Option 3. This option again sets the cut at 3.29 for the distinction between basic and below basic, but sets the cut for the distinction between proficient and distinguished at 4.0. You already know the rationale for 3.29 and the impact. The rationale for 4.0 is that this requires a principal to achieve on average across the 72 items at least an average of highly effective. In short, if a principal got a 4 on the 5-point scale on every one of the 72 items, they would be right at the cut between proficient and distinguished. They could of course achieve a score of 4.0 by being a 5 on some of the items and less than a 4 on some of the other items. In any event, this would yield 14.2% of the principals in the National Field Trial distinguished.

We would like your help in deciding among these new options, or of course, the cuts you recommended at the meeting in Nashville (but which you said you weren't completely comfortable with).

Please respond to this email by saying which of the three options you could endorse. You could say that you endorse only one of the three options or you could endorse two of the three, or all three options. Again, Option 1 is to set the cuts where they were at the end of the meeting, which was 3.42 and 3.77. Option 2 is to set the cuts at 3.29 and 3.87. Option 3 is to set the cuts at 3.29 and 4.0, and. In all three options, the cut for proficient is 3.60, which is where it was set at the end of the meeting and which we all agree is a good place.

All 22 panelists responded. Of those, eighteen support Option 3. Of those eighteen, three indicated that they support either Option 2 or Option 3. Of the other four panelists, one favored Option 1, one favored either Options 1 or 2, and two favored Option 2. Thus, our decision is to keep the cut between basic and proficient where it was set during the bookmarking process, but set the cut between basic and below basic to 3.29, resulting in 17% of principals below basic on the national field trial, and set the cut between

proficient and distinguished to 4.0, resulting in 14.2% of the principals in the national field trial being designated as distinguished.

Panelists' Evaluations of the Process and Results

At the end of the process, panelists were given an evaluation survey. Nineteen panelists completed and handed in their survey before leaving the meeting. Three panelists had to leave the meeting slightly early. Of those three, evaluation surveys were obtained from two. Thus, the following results are based on 21 of the 22 panelists.

Twenty of the 21 respondents said they were generally satisfied with the placement of the proficient cut score and sixteen said they were generally satisfied with the basic and distinguished cut scores. Unfortunately, the item was double-barreled. Had we asked separately about basic and distinguished, we would have learned that there was slightly more concern about distinguished than basic. In both cases, however, of those who were concerned, they thought the basic cut was set a bit too high and the distinguished a bit too low. Four individuals indicated they would move distinguished to be more demanding and two indicated that they would move basic to be less demanding.

As to the process of the bookmark standard setting event, panelists were strongly positive. 100% of the panelists indicated that a) workshop leaders clearly explained the purpose of the meeting, b) workshop leaders clearly explained the task, c) the large and small group discussions helped understand the process, d) they were able to follow the instructions and complete the recording form accurately, e) the discussions after the first round of bookmark placements were helpful, f) the information showing the distribution of bookmark placements was helpful, and g) the facilities and food service helped to create a good working environment. All but one panelist said the examples and exercises helped them understand how to perform their task and the discussions after the second round of bookmark placements were helpful to them. (See Table 7)

100% of the panelists agreed or strongly agreed with each of the following: a) I understood the purpose of the standard setting workshop; b) the overview and training on the Bookmark method gave me the information I needed to complete the assignment; c) the introduction to the VAL-ED gave me the information I needed to complete my assignment; d) the training on the performance level descriptors gave me the information I needed to complete my assignment; and e) the agreement data gave me the information I needed to complete my assignment. (See Table 8)

As for the clarity of the performance level descriptors (PLDs), virtually all of the panelists found them either very clear or somewhat clear. In the case of the basic PLD, three found it somewhat clear. For the proficient PLD, one found it somewhat unclear, and for the distinguished PLD, two found it somewhat unclear. None of the panelists found the performance level descriptors "very unclear." (See Table 9)

Panelists agreed that table and whole group discussions were important in their bookmark placement decisions. Panelists also agreed that the performance level descriptors, agreement data, and impact data were also important.

Finally, panelists thought there was enough time to complete each of the tasks that they were requested to do, with a few saying they had more than enough time and with only one panelist indicating that there was too little time for the overview and training on the bookmark method. Nineteen of the twenty who responded to the question, “The standard setting process was fair and unbiased” agreed that it was. The one who disagreed said, “Not biased in a bad way, but certainly biased by the experiences of the folks involved.”

In sum, panelists strongly supported where they had placed the proficient cut and three fourths supported where they had placed the basic and distinguished cuts. Nevertheless, as described above, the panelists were asked a week and a half after the standard setting workshop whether they were still comfortable with where they set the basic and distinguished cuts and given a couple options as to how they might be changed. In that stage of the process, all but one of the respondents indicated that they wished to make the basic versus below basic cut less demanding and the proficient versus distinguished cut more demanding.

Summary and Conclusions

A new assessment of K-12 school principal leadership was developed with Wallace Foundation support, the VAL-ED. In the spring and summer of 2008, a national field trial was completed, consisting of 218 schools from across the nation, with roughly one third elementary, one third middle, and one third high schools. Data from the national field trial will be used to create national norms. In addition to presenting results of a principal assessment in terms of percentile ranks based on the national norms, results will also be presented in terms of performance levels: distinguished, proficient, basic, and below basic.

A Bookmark method was used to set the cut scores between each performance level. A panel of 22 experts was recruited from across the nation, consisting of ten principals, four teachers, four supervisors of principals, two researchers of school leadership, and two education policymakers. The panel was convened in August of 2008. At the end of the standard setting event, panelists had placed three cuts on the effectiveness rating scale continuum from 1.0 to 5.0. The cut to distinguish proficient from basic was set at 3.60, the cut between distinguished and proficient at 3.77, and the cut between basic and below basic at 3.42. These cuts resulted in 30% of principals in the national field trial below basic, 50% below proficient, and 70% below distinguished.

Panelists were positive about the process and their satisfaction with the cuts set. Nevertheless, 24% expressed some concern about where the cuts were set between basic and below basic and distinguished and proficient. In response to panelists’ concerns, a post-standard setting communication with panelists asked them whether they wished to a) keep the cuts where they were set at the end of the standard setting event, b) move the cuts to the median cut for the least demanding table (there were five tables that operated independently in the standard setting task) for the distinction between basic and below basic and the most demanding table for the distinction between proficient and distinguished, or c) move the cut for basic and below basic to the least demanding table

and the cut for the distinction between proficient and distinguished to 4.0 on the impact scale. All 22 panelists responded. Twenty-one favored moving the basic to below basic cut to be less demanding and the proficient to distinguished cut to be more demanding. Of those twenty-one, all favored moving the basic to below basic cut to 3.29, yielding 17% of principals below basic, and eighteen of the 21 favored moving the cut between distinguished and proficient to 4.00, yielding 14.2% of the principals distinguished. The final decisions, then, is to set the cuts consistent with the panelists' preferences. The cut between basic and below basic is 3.29, between basic and proficient is 3.60, and between proficient and distinguished is 4.00.

Table 1
Field Test Responses

		Form A		Form C		Overall	
		No.	% of total	No.	% of total	No.	% of total
Level of Schooling	Elementary	38	37%	48	42%	86	39%
	Middle	33	32%	37	32%	70	32%
	High	32	31%	30	26%	62	28%
Wallace Districts	Wallace	33	32%	30	26%	63	29%
	Not Wallace	70	68%	85	74%	155	71%
Geographic Location	West	26	25%	25	22%	51	23%
	South	31	30%	35	30%	66	30%
	Midwest	20	19%	27	23%	47	22%
	Northeast	26	25%	28	24%	54	25%
Urbanicity	Urban	36	35%	49	43%	85	39%
	Suburban	43	42%	41	36%	84	39%
	Rural	24	23%	25	22%	49	22%
Total	Total	103	-	115	-	218	-

Note: percentages may not add to 100% due to rounding

Table 2
Proficient Bookmark Page Number Median Cuts and Item Medians

Table	Round 1		Round 2		Round 3	
	Page	Item Mean	Page	Item Mean	Page	Item Mean
1	31.00	3.63	30.00	3.64	30.00	3.64
2	37.00	3.60	44.50	3.53	44.50	3.53
3	32.00	3.62	30.00	3.64	34.00	3.61
4	37.00	3.60	37.00	3.60	37.00	3.60
5	35.50	3.60	39.00	3.58	39.00	3.58
Room	35.50	3.60	37.00	3.60	37.00	3.60

Table 3
Distinguished Bookmark Page Number Median Cuts and Item Medians

Table	Round 1		Round 2		Round 3	
	Page	Item Median	Page	Item Median	Page	Item Median
1	55.50	3.79	55.50	3.80	62.50	3.87
2	58.50	3.76	63.00	3.66	63.00	3.66
3	50.00	3.74	59.00	3.84	59.00	3.39
4	59.00	3.76	59.00	3.76	59.00	3.76
5	61.00	3.80	59.00	3.73	59.00	3.73
Room	58.50	3.18	59.00	3.76	59.50	3.77

Table 4
Basic Bookmark Page Number Median Cuts and Item Medians

Table	Round 1		Round 2		Round 3	
	Page	Item Median	Page	Item Median	Page	Item Median
1	12.00	3.48	8.00	3.43	9.00	3.43
2	14.50	3.42	15.00	3.29	15.00	3.29
3	12.00	3.45	16.00	3.53	16.00	3.48
4	29.00	3.55	13.00	3.41	7.00	3.33
5	19.00	3.48	24.00	3.48	21.50	3.46
Room	14.00	3.43	15.00	3.42	15.00	3.42

Table 5
Standard Errors of Cut Scores at Each Performance Level

	Basic	Proficient	Distinguished
OIB Pages	2.58	2.74	0.97

Table 6
Standard Deviations of Bookmark Placements by Round

Round	Basic	Proficient	Distinguished
1	8.05	9.56	8.75
2	5.70	5.78	3.08
3	5.14	5.42	2.12

Table 7
Helpfulness of the Process

	Agree	Disagree
The workshop leaders clearly explained the purpose of the meeting.	21	0
The workshop leaders clearly explained my task.	21	0
The examples and exercises helped me understand how to perform my task.	20	1
The large and small group discussions helped me understand the process.	21	0
I was able to follow the instructions and complete the recording form accurately.	21	0
The discussions after the first round of bookmark placements were helpful to me.	21	0
The discussions after the second round of bookmark placements were helpful to me.	20	1
The information showing the distribution of bookmark placements was helpful to me.	20	0
The facilities and food service helped to create a good working environment.	21	0

Table 8
Access to Information Needed

	Strongly agree	Agree	Disagree	Strongly disagree
I understood the purpose of this standard setting workshop	14	7	0	0
The overview and training on the bookmark method gave me the information I needed to complete the assignment	15	6	0	0
The introduction to the VAL-ED gave me the information I needed to complete my assignment	13	8	0	0
The training on the Performance Level Descriptors gave me the information I needed to complete my assignment.	13	8	0	0
The agreement data gave me the information I needed to complete my assignment	14	7	0	0

Table 9
Clarity

Rate the clarity of ...	Very clear	Somewhat clear	Somewhat unclear	Very unclear
The instructions provided by the trainers.	17	4	0	0
Description of the Basic performance level	7	11	3	0
Description of the Proficient performance level	9	11	1	0
Description of the Distinguished performance level	9	10	2	0

Key Processes						
Core Components	Planning	Implementing	Supporting	Advocating	Communicating	Monitoring
High Standards for Student Learning						
Rigorous Curriculum (content)						
Quality Instruction (pedagogy)						
Culture of Learning & Professional Behavior						
Connections to External Communities						
Performance Accountability						

Figure 1: VAL-ED Conceptual Framework

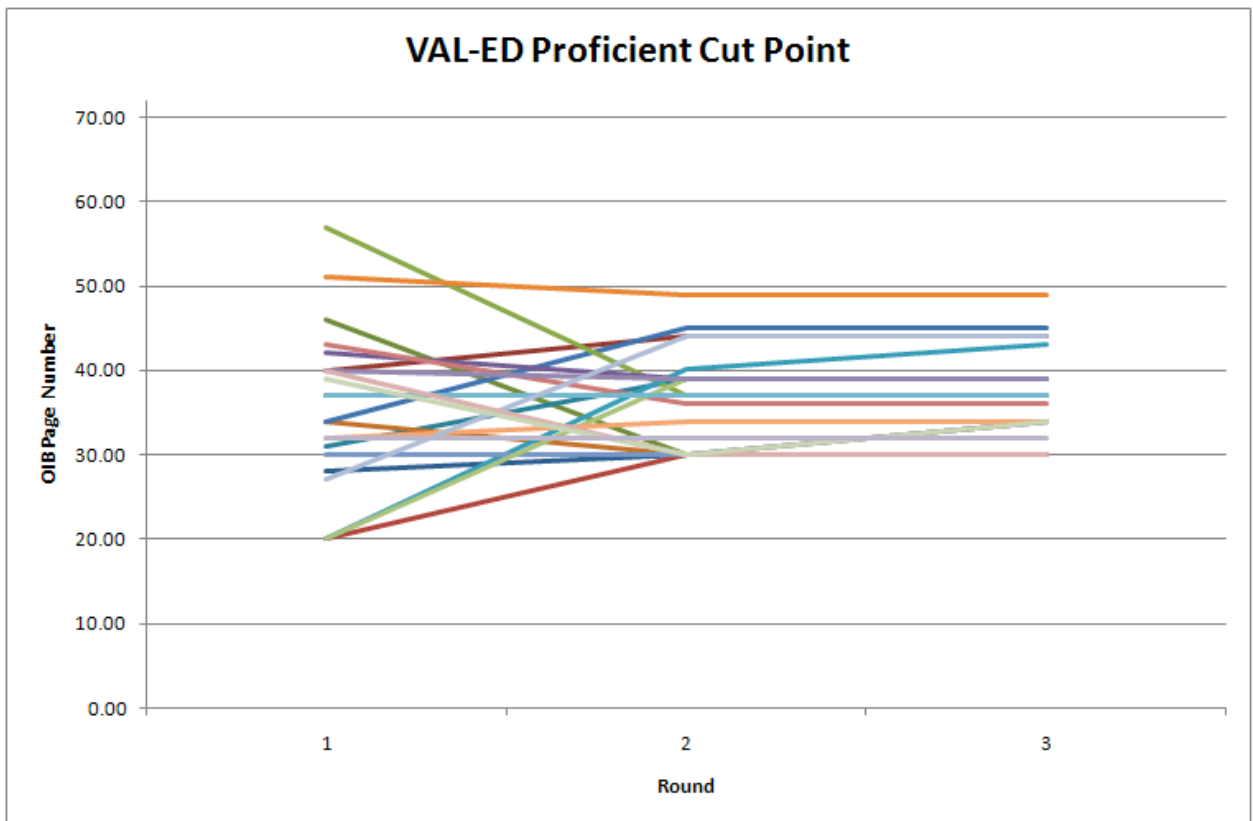


Figure 2: Proficient Cuts in Page Numbers for Each Panelist across Rounds

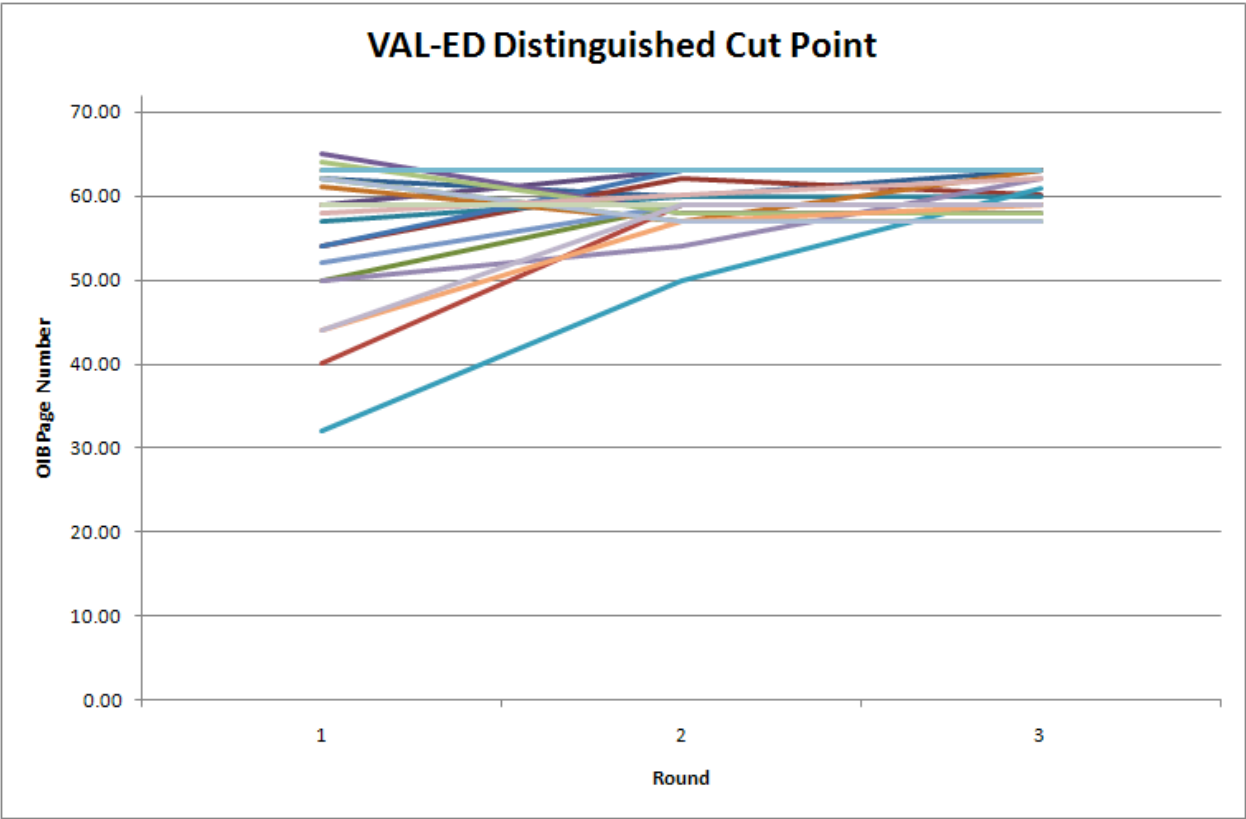


Figure 3: Distinguished Cuts in Page Numbers for Each Panelist across Rounds

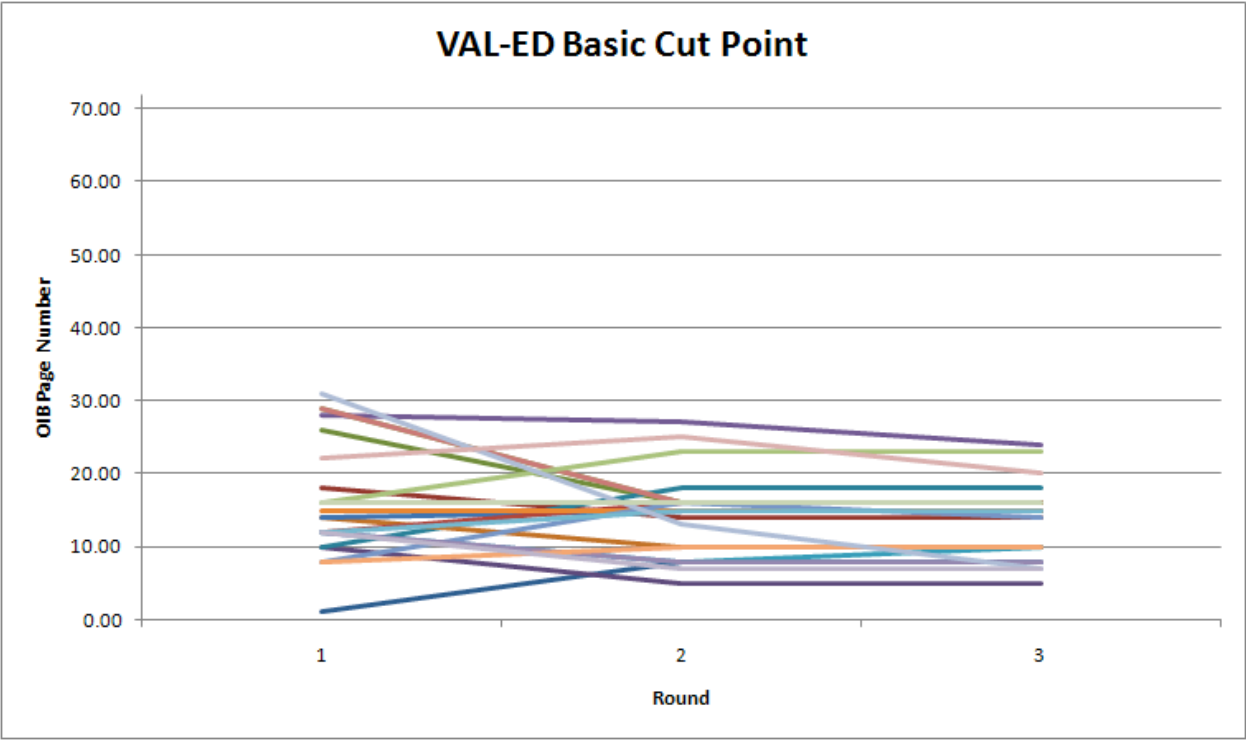


Figure 4: Basic Cuts in Page Numbers for Each Panelist across Rounds

References

- Goldring, E., Porter, A.C., Murphy, J., Elliott, S., Cravens, X. (2007). Assessing learning-centered leadership: Connections to research, professional standards, and current practices. *Journal of Leadership and Policy in Schools*.
- Green, D.R., Trimble, C.S., Lewis, D.M. (2003). Interpreting the results of three different standard-setting procedures. *Educational Measurement: Issues and Practice*, 22(1), 22-32.
- Huynh, H. (2003, August). Technical Memorandum for Computing Standard Error in Bookmark Standard Setting. (The South Carolina PACT 2003 Standard Setting Support Project). Columbia: University of South Carolina.
- Murphy, J., Elliott, S.N., Goldring, E.B., & Porter, A.C. (2006). *Learning-centered leadership: A conceptual foundation*. New York, NY: Wallace Foundation.
www.wallacefoundation.org/ELAN/TR/KnowledgeCategories/DevelopingLeaders/PerfMeasurement/PerfMeasurementResource.ConceptualFoundation.htm
- Porter, A.C., Goldring, E.B., Murphy, J., Elliott, S.N., & Cravens, X. (2006). *A framework for the assessment of learning-centered leadership*. New York, NY: Wallace Foundation.
www.wallacefoundation.org/ELAN/TR/KnowledgeCategories/DevelopingLeaders/PerfMeasurement/PerfMeasurementResource.ConceptualFoundation.htm

Appendix A

VAL-ED Standard Setting
Agenda

Day 1	Activity	Lead Responsibility
8:30-9:00	Welcome and Introductions	Andy Porter
	Studying the Process	Ellen Goldring
9:00-12:00 (15 minute break at 10:30)	Training and practice 1. Review the framework, proficiency level descriptors, and assessment design (Andy) 2. Respond to a small sample of items on the assessment (Ellen) 3. Write brief definitions of target principals (Steve) 4. Formal training on concepts and procedures (Steve) 5. Practice on the bookmark placement task (Steve)	Steve Ferrara
12:00-12:45	Lunch	--
12:45-4:45 (15 minute break at 2:30)	Round 1 1. Review the OIB and answer the two questions 2. Sign readiness forms 3. Place bookmarks	Steve Ferrara
4:45-5:00	Collect materials, debrief, plan for day 2	Steve
5:00-5:30	Interviews	Process study team
Day 2	Activity	Lead Responsibility
8:30-10:30	Round 2 1. Interpret and discuss feedback from round 1	Steve Ferrara

	2. Sign readiness form	
	3. Place bookmarks	
10:30-11:00	Interviews and Break	Process study team
11:00-12:00	Round 3	Steve Ferrara
	1. Interpret and discuss feedback from round 2	
	2. Sign readiness form	
	3. Place bookmarks	
12:00-12:45	Lunch and Interviews	
12:45-1:30	Round 3 Continues	
1:30-2:00	Final debriefing (Steve)	Steve Ferrara
	Refinement comments on the PLDs (Ellen)	
	Workshop evaluation (Andy)	
2:00-2:45	Interviews	Process study team

Setting Performance Standards for the VAL-ED

Orientation and Training to Use the Bookmark Standard Setting Procedure

August 12 -13, 2008
Vanderbilt University
Nashville, TN



Workshop goal

- Recommend cut scores for the Vanderbilt Assessment of Leadership in Education (VAL-ED) that correspond to the performance level descriptors for Below Basic, Basic, Proficient, and Distinguished levels of leadership effectiveness



Introductions

- ⌘ Welcome and opening comments
- ⌘ Panelist introductions
- ⌘ VAL-ED project staff
- ⌘ Workshop leader

Bookmark Standard Setting

VAL-ED

3



Today's activities

- ⌘ Orient you to the workshop (overview)
- ⌘ Training and practice in conducting the Bookmark procedure
- ⌘ Round 1
 - ⌘ First opportunity to recommend three cut scores

Bookmark Standard Setting

VAL-ED

4



Security and confidentiality



- All VAL -ED items are secure and copyrighted by Vanderbilt University
- Please DO NOT:
 - Remove any secure material from the meeting room
 - Discuss the placement of your cut scores or items among yourselves outside the sessions
 - Discuss secure materials with non -participants
 - Discuss the cut scores after this workshop
- It is OK to discuss the Bookmark process

Bookmark Standard Setting

VAL-ED

5

Paperwork and logistics



Bookmark Standard Setting

VAL-ED

6

The VAL-ED

- ⌘ Conceptual framework
- ⌘ PLDs
- ⌘ Assessment design
- ⌘ Development and field test process
- ⌘ Respond to some items

Bookmark Standard Setting

VAL-ED

7

Orientation to the workshop

Bookmark Standard Setting

VAL-ED

8



Workshop materials



- ⌘ Agenda
- ⌘ Performance level descriptors
- ⌘ The VAL-ED instrument
- ⌘ Ordered item book
- ⌘ Readiness forms
- ⌘ Recording forms

Bookmark Standard Setting

VAL-ED

9

Overview of standard setting



- ⌘ Systematic process
 - ⌘ Trained experts use their knowledge of effective leadership and principals' performance
 - ⌘ Recommend cut scores required to attain each performance level
- ⌘ Conducted in three rounds
 - ⌘ You will have three rounds to review items, answer the two questions, and discuss your insights with your colleagues
 - ⌘ At the end of each round, participants will place bookmarks—individually

Bookmark Standard Setting

VAL-ED

10

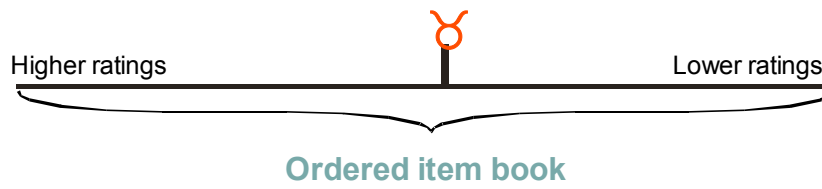
Introduction to a performance standard



Principals' overall leadership effectiveness ratings are based on average ratings (a) on all items, (b) by teachers, supervisors, and the principals themselves.

Both principals' performance ratings and average item ratings are located on the same scale.

You will place a bookmark in an ordered item book on a page to represent a performance standard.



Bookmark Standard Setting

VAL-ED

11

What do we mean by “standards”?



✧ VAL-ED conceptual framework

- A conceptual framework that describes leadership skills that principals should possess or develop

✧ Performance standards

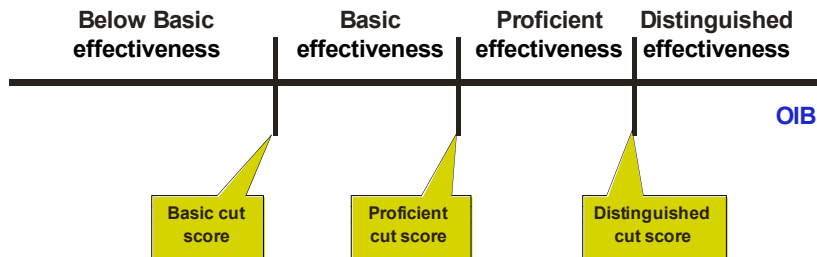
- PLDs
 - Definitions of Distinguished, Proficient, Basic, and Below Basic leadership effectiveness
- Cut scores on the rating scale that separate the performance levels
 - Defined as pages in the ordered item book on which you place bookmarks

Bookmark Standard Setting

VAL-ED

12

VAL-ED performance levels



Bookmark Standard Setting

VAL-ED

13

Performance level descriptors (PLDs)



- ⌘ Define leadership at four levels of leadership performance and effectiveness
- ⌘ Distinguished, Proficient, Basic, and Below Basic

Bookmark Standard Setting

VAL-ED

14

Distinguished leadership



- ⌘ A distinguished leader exhibits learning - centered leadership behaviors at levels of effectiveness that are **virtually certain** to influence teachers positively and result in **strong** value-added to student achievement and social learning for **all students** .

Bookmark Standard Setting

VAL-ED

15

Proficient leadership



- ⌘ A proficient leader exhibits learning - centered leadership behaviors at levels of effectiveness that are **likely** to influence teachers positively and result in **acceptable** value-added to student achievement and social learning for **all students** .

Bookmark Standard Setting

VAL-ED

1

Basic leadership



- ⌘ A leader at the basic level of proficiency exhibits learning -centered leadership behaviors at levels of effectiveness that are **likely** to influence teachers positively and that result in **acceptable** value-added to student achievement and social learning for **some** sub-groups of students, but **not all**.

Bookmark Standard Setting

VAL-ED

17

Below Basic leadership



- ⌘ A leader at the below basic level of proficiency exhibits learning -centered leadership behaviors at levels of effectiveness that are **unlikely** to influence teachers positively **nor result in acceptable** value-added to student achievement and social learning for students.

Bookmark Standard Setting

VAL-ED

17



The Bookmark method

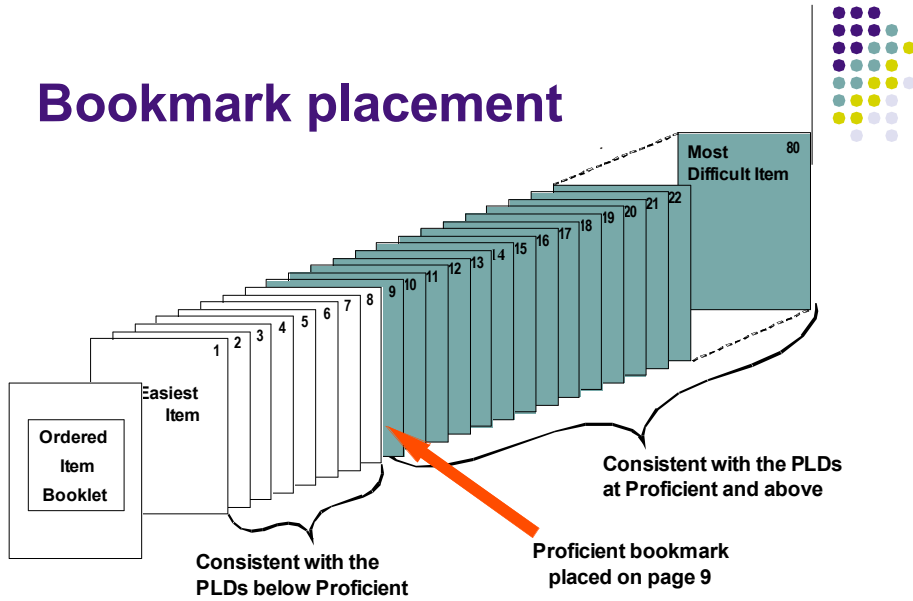
- Empirical support
- Used in ~25 state and district student achievement testing programs since 2000
- Has withstood legal challenges and is approved in NCLB peer reviews



The Bookmark method (cont'd)

- Using an ordered item book, experts:
 - Find the location in the ordered item book that separates performance levels, and then
 - Literally, place a bookmark at that location in the ordered item book
- You will place bookmarks in three rounds after discussion and deliberation

Bookmark placement



Bookmark Standard Setting

VAL-ED

21

A way to think about the bookmark's location

- ✧ Read a book partway through, place your bookmark
- ✧ You can be expected to be able to tell the story on the pages before the bookmark
- ✧ You cannot be expected to tell what happens from the bookmark and on

Bookmark Standard Setting

VAL-ED

22



Ordered item book

- Items are taken from the spring 2008 administration of Form A of VAL -ED
- Teachers, supervisors, and principals themselves rated a principal 's leadership effectiveness
- Ordering of the items in the booklet is based on the average ratings of principals



Ordered item book (cont 'd)

- One item per page
- Item with highest average rating is on page 1
 - It's relatively easy to achieve a high rating on this item
- Item with lowest average rating is on the last page
 - It's relatively difficult to achieve a high rating on this item

Studying the ordered item book



- In preparation for round 1, you will answer the two questions for each item:
 - ⌘ What behaviors must a principal exhibit in order to achieve a rating of at least highly effective on this item?
 - ⌘ What makes it more difficult to achieve a rating of at least highly effective on this item than on all previous items in this book?

What do you need to consider to decide where to place a bookmark?



- ⌘ Answers to the two questions
- ⌘ Just barely Proficient (etc.)
 - ⌘ That's coming up

Just barely: illustration using Proficient



- ⌘ Principals who are just barely Proficient exhibit learning-centered leadership behaviors at levels of effectiveness that are likely to influence teachers positively and result in acceptable value -added to student achievement and social learning for all students, but just barely
- ⌘ I would expect them to:
 - ⌘ Exhibit many, but not all learning -centered leadership behaviors
 - ⌘ Add value for most, but not all student subgroups
 - ⌘ Do these things much, but not all of the time

Bookmark Standard Setting

VAL-ED

27

Your turn



- ⌘ Just barely Proficient
-

- ⌘ Implications for just barely Distinguished and just barely Basic

Bookmark Standard Setting

VAL-ED

28

Bookmark placement task



- ⌘ Where do I place my bookmark?
- ⌘ How do I decide where to place my bookmark?

Bookmark Standard Setting

VAL-ED

29

Bookmark placement task



- ⌘ Place your bookmark on the page where a just barely Proficient principal would be rated at least highly effective on the Effectiveness Rating Scale.

Bookmark Standard Setting

VAL-ED

30

Bookmark placement task (cont.)



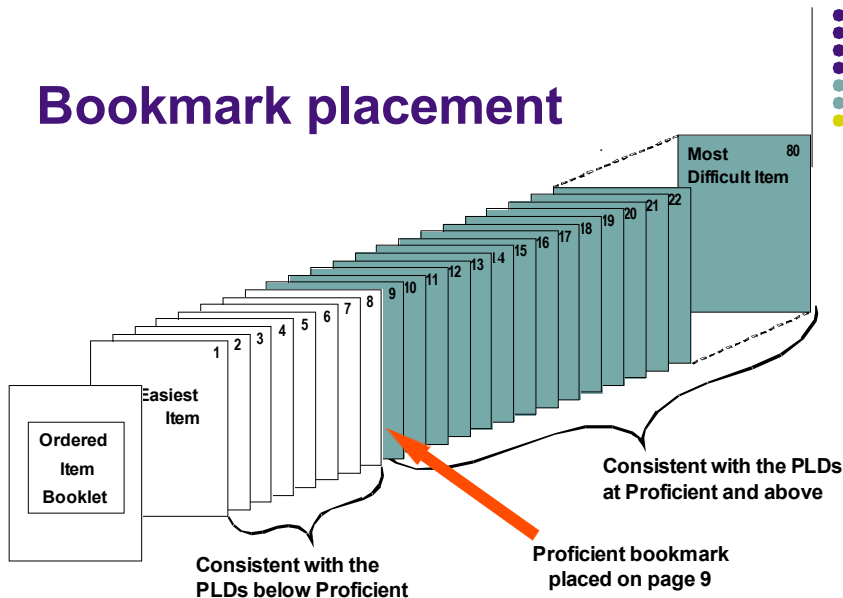
- ⌘ You would expect clearly Proficient principals to achieve a highly effective rating on items beyond the current page
- ⌘ You would expect principals who are clearly in the Basic level to achieve a lower rating on this item

Bookmark Standard Setting

VAL-ED

31

Bookmark placement



Bookmark Standard Setting

VAL-ED

32

Bookmark placement task (cont.)



- ⌘ Place your bookmark on the page where a just barely Distinguished principal would be rated at least highly effective on the Effectiveness Rating Scale.
- ⌘ Place your bookmark on the page where a just barely Basic principal would be rated at least highly effective.

Bookmark Standard Setting

VAL-ED

33

Notes on bookmark placement



- The bookmark (conceptually) divides two item sets, not two items
- Don't get hung up on the component and process represented by any single item
- It's OK to disregard an item that seems out of order

Bookmark Standard Setting

VAL-ED

34

Does an item seem out of order?



- Item orderings are based on average principal ratings
- Item order is influenced by a number of factors
- It's OK to look past an item that seems out of order
- Other puzzlements ...

Bookmark Standard Setting

VAL-ED

35

Training and practice



⌘ Modeling

- ⌘ Answering the two questions
- ⌘ Determining whether to place a bookmark on a page in the OIB

⌘ Practice

- ⌘ Answering the two questions
- ⌘ Determining whether to place a bookmark on a page in the OIB

⌘ Using the practice OIB

Bookmark Standard Setting

VAL-ED

36



The two questions

- ⌘ What behaviors must a principal exhibit in order to achieve a rating of at least highly effective on this item?
- ⌘ What makes it more difficult to achieve a rating of at least highly effective on this item than on all previous items in this book?



Your turn to practice

- ⌘ Answering the two questions
- ⌘ Determining whether to place a bookmark on a page in the OIB
- ⌘ Use the practice OIB

⌘ Are you ready to undertake round 1?

A note about table discussions



- ⌘ You are colleagues —take turns speaking
- ⌘ You all are experts —with different perspectives, different interpretations—who will make different, equally appropriate and reasonable judgments
- ⌘ Independent judgments
- ⌘ There is no right answer nor wrong answer —you’re using your judgment to make recommendations
- ⌘ Consensus is not the goal —this is a convergence process
- ⌘ Sharing insights and rationales is necessary — persuasion and argumentation are undesirable

Workshop overview



- ⌘ Today
 - ⌘ Round 1: bookmarks for Proficient, Basic, and Advanced
 - ⌘ Interviews
- ⌘ Tomorrow
 - ⌘ Round 2: feedback, bookmarks
 - ⌘ Round 3: feedback, bookmarks
 - ⌘ PLD refinement comments
 - ⌘ Final debriefing
 - ⌘ Workshop evaluation
 - ⌘ Interviews

Round 1



Bookmark Standard Setting

VAL-ED

41

Materials for round 1

- ⌘ PLD
- ⌘ Notes on JBP, JBD, JBB
- ⌘ OIB
- ⌘ Readiness form
- ⌘ Bookmark recording form



Bookmark Standard Setting

VAL-ED

42



Preparation for round 1

- ⌘ Answer the two questions for all items
- ⌘ Review the PLDs and the bookmark placement task
- ⌘ Sign the readiness form
- ⌘ Place your bookmarks—independently



Answer the two questions

- ⌘ Work collaboratively at your table
 - ⌘ Record your responses on the OIB pages
-
- ⌘ What behaviors must a principal exhibit in order to achieve a rating of at least highly effective on this item?
 - ⌘ What makes it more difficult to achieve a rating of at least highly effective on this item than on all previous items in this book?

Review the PLDs and “JBs”



- ⌘ JBP
- ⌘ JBD
- ⌘ JBB

Your judgmental task



- ⌘ Place your bookmark on the page where a just barely Proficient principal would be rated at least highly effective on the Effectiveness Rating Scale.
- ⌘ Place your bookmark on the page where a just barely Distinguished principal would be rated at least outstandingly effective on the Effectiveness Rating Scale.
- ⌘ Place your bookmark on the page where a just barely Basic principal would be rated satisfactorily effective or lower.

When you're ready to work independently



- ⌘ Sign the readiness form
- ⌘ Place your bookmarks
 - ⌘ Proficient first
 - ⌘ Then Distinguished
 - ⌘ Then Basic
- ⌘ Complete the bookmark recording form

Bookmark Standard Setting

VAL-ED

47

Round 1 recording form



- ⌘ Take a look

Bookmark Standard Setting

VAL-ED

48

Round 2



Bookmark Standard Setting

VAL-ED

49

Materials for round 2



- ⌘ PLD
- ⌘ Notes on JBP, JBD, JBB
- ⌘ OIB
- ⌘ Readiness form
- ⌘ Bookmark recording form
- ⌘ **Feedback from round 1**

Bookmark Standard Setting

VAL-ED

50

Preparation for round 2



- ⌘ Agreement information
- ⌘ Discussion questions: tables, then room

Bookmark Standard Setting

VAL-ED

51

Discussion questions



- ⌘ Goals
 - ⌘ Add important information to your thinking
 - ⌘ Develop common understandings
 - ⌘ Inform discussion and possible re-evaluation of decisions about locating bookmarks
- ⌘ Expectation is converging judgments
 - ⌘ Consensus and convincing each other is not a goal

Bookmark Standard Setting

VAL-ED

52

Let's take a look



Bookmark Standard Setting

VAL-ED

53

After Discussing the Agreement Information



- ⌘ How comfortable are you with your recommended bookmark placements?
 - ⌘ Compared to your table and room medians
- ⌘ What are the knowledge and skill demands (and the difficulty) of the items between the:
 - ⌘ Lowest and highest bookmark placements at your table?
- ⌘ Remember:
 - ⌘ Item-based rationales for adjusting or not adjusting bookmark placements

Bookmark Standard Setting

VAL-ED

54

When you're ready to work independently



- ⌘ Sign the readiness form
- ⌘ Place your bookmarks
 - ⌘ Proficient first
 - ⌘ Then Distinguished
 - ⌘ Then Basic
- ⌘ Complete the bookmark recording form

Bookmark Standard Setting

VAL-ED

55

Your judgmental task



- ⌘ Place your bookmark on the page where a just barely Proficient principal would be rated at least highly effective on the Effectiveness Rating Scale.
- ⌘ Place your bookmark on the page where a just barely Distinguished principal would be rated at least highly effective.
- ⌘ Place your bookmark on the page where a just barely Basic principal would be rated at least highly effective.

Bookmark Standard Setting

VAL-ED

56

Round 3



Bookmark Standard Setting

VAL-ED

57

Materials for round 3



- ⌘ PLD
- ⌘ Notes on JBP, JBD, JBB
- ⌘ OIB
- ⌘ Readiness form
- ⌘ Bookmark recording form
- ⌘ **Feedback from round 1**
- ⌘ **Impact data**

Bookmark Standard Setting

VAL-ED

58

Preparation for round 3



- ⌘ Agreement and impact information
- ⌘ Discussion questions: tables, then room

Bookmark Standard Setting

VAL-ED

59

Discussion questions



- ⌘ Goals
 - ⌘ Add important information to your thinking
 - ⌘ Develop common understandings
 - ⌘ Inform discussion and possible re-evaluation of decisions about locating bookmarks
- ⌘ Expectation is converging judgments
 - ⌘ Consensus and convincing each other is not a goal

Bookmark Standard Setting

VAL-ED

60

Let's take a look



Bookmark Standard Setting

VAL-ED

61

After discussing the agreement information



- ⌘ How comfortable are you with your recommended bookmark placements?
 - ⌘ Compared to your table and room medians
- ⌘ What are the knowledge and skill demands (and the difficulty) of the items between the:
 - ⌘ Lowest and highest bookmark placements at your table?
- ⌘ Implications of the impact data

Bookmark Standard Setting

VAL-ED

62

When you're ready to work independently



- ⌘ Sign the readiness form
- ⌘ Place your bookmarks
 - ⌘ Proficient first
 - ⌘ Then Distinguished
 - ⌘ Then Basic
- ⌘ Complete the bookmark recording form

Your judgmental task



- ⌘ Place your bookmark on the page where a just barely Proficient principal would be rated at least highly effective on the Effectiveness Rating Scale.
- ⌘ Place your bookmark on the page where a just barely Distinguished principal would be rated at least highly effective.
- ⌘ Place your bookmark on the page where a just barely Basic principal would be rated at least highly effective.

After round 3

- ⌘ Final results
- ⌘ Refinement comments on the PLDs
- ⌘ Workshop evaluations
- ⌘ Interviews

Bookmark Standard Setting

VAL-ED

65

⌘ End

Bookmark Standard Setting

VAL-ED

66

