

**Benchmarking Capacity Building:  
A Developmental View** | November  
2011

---



Paper presented at the 2011 annual meeting  
of the American Evaluation Association,  
Anaheim, CA.

Kristine L. Chadwick, Ph.D.

Kimberly S. Cowley, Ed.D.

Kimberly D. Good, Ph.D.

Sharon D. Harsh, Ed.D.

Appalachia Regional Comprehensive Center at Edvantia,

P.O. Box 1348

Charleston, WV 25325-1348

800.624.9120

[info@edvantia.org](mailto:info@edvantia.org)

The Appalachia Regional Comprehensive Center (ARCC) at Edvantia is one of 21 technical assistance centers funded by the U.S. Department of Education (16 regional centers and five content centers). The ARCC provides the state education agencies in North Carolina, Kentucky, Tennessee, Virginia, and West Virginia with intensive technical assistance to address the No Child Left Behind (NCLB) requirements and meet student achievement goals. The ARCC at Edvantia is a dynamic, collaborative network consisting of the Center for Equity and Excellence in Education at George Washington University, the Eastern Stream Center on Resources and Training (ESCORT), the National Center for Family Literacy, the SERVE Center for Continuous Improvement at the University of North Carolina—Greensboro, and the Southern Regional Education Board.

Since 1966, Edvantia has been in the business of developing knowledge, resources, and professional services to improve education. A not-for-profit corporation with primary offices in Tennessee and West Virginia, Edvantia offers comprehensive services grounded in research and best practice that help educators meet the needs of all students. Edvantia's national expertise is focused in the following specialty areas: program evaluation, applied research, data analytics, school coaching, district performance improvement, and state education agency capacity building. Edvantia's clients and funders include federal agencies, state and local education agencies, professional associations, foundations, education product vendors, and others committed to improving education.

For information about Edvantia research, products, or services, contact



P.O. Box 1348, Charleston, WV 25325 • 304.347.0400 • 800.624.9120 • fax 304.347.0487  
One Vantage Way, Suite C-200, Nashville, TN 37228 • 615.565.0101 • fax 615.565.0112  
[info@edvantia.org](mailto:info@edvantia.org) • [www.edvantia.org](http://www.edvantia.org)

© 2011 by Edvantia, Inc.

Reproduction of this document is permitted with the following citation:  
Cowley, K. S., Good, K.D., & Finch, N.L. (2011, November). *The best laid plans . . . often go astray: Conducting a mixed-methods evaluation of a changing project*. Paper presented at the annual meeting of the American Evaluation Association, Anaheim, CA.

The contents of this paper were developed under grant number S283B050004 from the U.S. Department of Education. However, those contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the federal government.

Edvantia is an equal employment opportunity/affirmative action employer.

# **Benchmarking Capacity Building: A Developmental View**

## **Introduction to Benchmarking**

Benchmarking originated in the early 1980s by the Chief Executive Officer of Xerox as a business and industry management tool for a quality control initiative (Camp, 1993). Known as corporate benchmarking, the tool provides business and industry with a systematic process to compare the practices of two or more companies or two or more divisions within a company. The comparisons identify differences or gaps between the performance of the two organizations, particularly when the benchmarking organization compares its practices to a known leader in the industry. The process allows the benchmarking organization to gain an external perspective on the processes and skills that create superior performance (Maintenance Resources, n.d.). The processes and skills that allow an organization to achieve or surpass an industry standard are enablers and include, but are not limited to, culture, leadership, organization vision, focus, staff attitude, and motivation (Fullan, 2008).

Unlike industry benchmarking, which compares performance and enablers across departments or organizations, the Appalachia Regional Comprehensive Center (ARCC) at Edvantia benchmarking process compares the performance of an individual with a research-based theory that describes the stages or levels of capacity building. The Concerns Based Adoption Model (CBAM) Levels of Use (LoU) (Hall & Hord, 2011) is used in this benchmarking process to determine the changes in capacity from the beginning to the end of an initiative. The technical assistance initiative, infused with customized ARCC support, is viewed as the enabler and catalyst for capacity building.

Within this paper the authors first describe the types of benchmarking and the benchmarking process. The next section describes the application of benchmarking in three areas: education, human behavior, and capacity building. After the foundation on benchmarking has been laid, the authors detail the ARCC's approach to capacity building and the model used to explore how to benchmark individual capacity building. The paper concludes with limitations, lessons learned, and recommendations.

## **Types of Benchmarking**

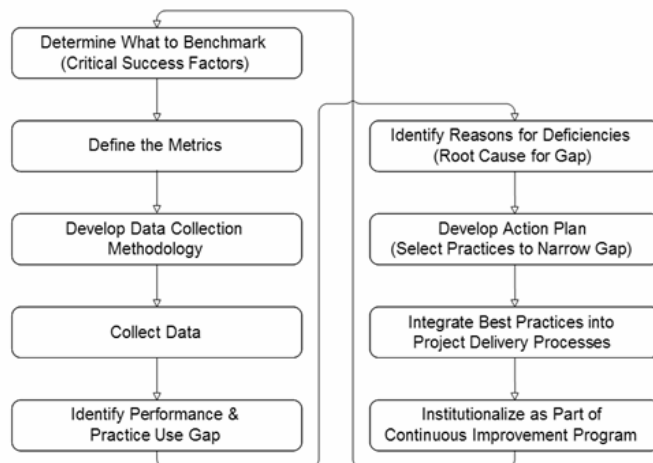
While most benchmarking procedures include external comparisons, benchmarking can be either an internal or an external process. Internal benchmarking typically involves different departments or units within an organization, is relatively easy to conduct and to make comparisons between participating departments, and can take place at the corporate and operational levels of the organization (Wood, n.d.). While internal benchmarking can lead to small, incremental improvements, it rarely leads to a major breakthrough (Maintenance Resources, n.d.). External benchmarking features the selection of a high-performing organization willing to grant access to its practices and procedures to another organization that wishes to emulate its performance. External benchmarking takes time and often involves periodic data collection and performance sampling that can extend across several years.

Corporations can choose from several types of benchmarking processes. Benchmarking can occur at a macro level, comparing a wide range of operations and procedures, or at a micro level, conducting an in-depth

analysis of a specific process. Some types of benchmarking, such as process and operational, can be internal or external while other types of benchmarking, such as best-in-class, world class, and competitive comparisons, can only be an external process.

## The Benchmarking Process

Benchmarking begins with an internal analysis of an organization's current practices and procedures, followed by comparing the identified practices with those of a known high-performing organization. The comparison creates a gap analysis that identifies the differences in practices between the two organizations and the distance the benchmarking organization has to move to reach or exceed the industry standard (Maintenance Resources, n.d.). In 2005, the National Research Council adapted Camp's 1989 roadmap (see Figure 1) for planning a benchmarking activity and identified steps to follow in implementing a benchmarking process.



*Figure 1. Benchmarking roadmap (Committee for Oversight and Assessment of U.S Department of Energy Project Management, National Research Council, 2005, adaptation from Camp, 1989).*

Regardless of the type of benchmarking, organizations need a well-structured and systematic process (Wood, n.d.). While there are similarities in the recommended benchmarking steps, portions of the process are essential to achieving successful results. For example, in describing the Juran benchmarking process, Wood indicated that the positioning analysis conducted in phase 1 was critical to determining appropriate benchmark comparisons, and without a comprehensive study of the practices identified as enablers in the benchmarked organization in phase 2, the organization might draw misinformed conclusions, which could lead to emulating practices that would not result in similar improvements. Similarly, successful organizational benchmarking is predicated on a thorough gap analysis (Maintenance Resources, n.d.), comparing the company's process in quantifiable terms to the results observed in the benchmarked organization and setting a goal and timeline to achieve the current level of performance of the benchmarked organization. The gap analysis and established goal must recognize that the benchmarked organization will improve during this period and will still be at a higher level of performance. Finally, the PHS Management Training process (2004) emphasized the importance of proto-planning, indicating that benchmarked areas should be those with the potential to add sustainable competitive advantages to the organization; in other words, not every process needs to be world class, only those that will deliver sustainable competitive advantages.

# **Application of Benchmarking**

## **Benchmarking in Education**

Currently, the best known example of education benchmarking is the comparisons made between the United States content standards and the curriculum standards used in nations with high student performance on international assessments (National Governors Association, Council of Chief State School Officers, & Achieve, Inc., 2008). The benchmarking resulted in identification of practices and curriculum content incorporated into a set of internationally benchmarked common core standards. The common core standards, now adopted by more than 40 states, provide the foundation for reforms in nationally funded grants and programs (Council of State School Officers, n.d.). In another example, international benchmarking in 19 countries participating in the World Education Indicators (WEI) program determined progress in the number of students who completed compulsory education, and enrolled in and completed a university education (UNESCO, 2007).

Districts have conducted benchmarking studies between schools to determine the factors that create high student performance and increased ratings on a total academic index (Lanier & Saltzman, 1999). Other studies between 25 higher education institutions in five states benchmarked the graduation success of Latina and Latino students in attaining a degree in science, technology, engineering, or mathematics (Dowd, Malcom, & Bensimon, 2009). At a specific process level, benchmarking has examined assessment procedures, such as essay grading methods (Boring, Hendrickson, Forester, Tran, & Erasmia, 2008), and compared instructional practices in engineering design programs at institutions of higher education (Fridley, Jorgensen, & Lamancusa, 1997).

## **Benchmarking Human Behavior**

Compared to industry and corporate activities, benchmarking human behavior is a relatively new endeavor. Two interesting examples of benchmarking human behavior come from a U.S. Department of Energy (DOE) study on measuring performance and a Sandia National Laboratories study on human reliability. In the DOE research (Committee for Oversight and Assessment of U.S Department of Energy Project Management, National Research Council, 2005), the agency worked with projects from external companies and separate program offices to identify performance improving processes and gain an understanding of what constituted good performance. The benchmarking resulted in the development of a quantitative process examination and the integration of performance measures into a common data collection system. The DOE's Project Manager Career Development Program incorporated the new measures beginning at the first level of certification, providing specific training in the application of DOE specific benchmarking procedures. The results of the study aided training and enabled project directors to benchmark their projects as well as contributed to the development and improvement of the measures and project management processes in the Department.

The second example from the Sandia National Laboratories reported (Boring et al., 2008) on benchmarking human reliability analysis methods. In the report, the Lab discussed benchmarking efforts to understand human reliability when performing routine procedures and the role of cognitive modeling in performance. Human reliability benchmarking analyzed human errors through a diagnosis activity and the corresponding selection of a response strategy. When the procedures and human errors were benchmarked across organizations, the data showed that employees made different assumptions about task characteristics and identified different tasks as starting points in the analysis of how to respond to situations. Benchmarking

pointed out that the tasks often decomposed at different levels and that the most important human errors in a given task were the result of the failure to attend to detail that would signal the need to take significant actions. The benchmarking in cognitive modeling demonstrated that analysis of human error should not focus solely on the result but also analyze the methods and processes that led to the end state, including performance shaping factors, assumptions made, and task decomposition.

## **Benchmarking Capacity Building**

Benchmarking capacity building analyzes the degree to which an individual or organizational increases its level of knowledge, skill, and capacity to perform a task or function. It is a dynamic, moving array of capabilities with a focus on analyzing the process and resultant state of acting. Benchmarking capacity building compares the behaviors or stages in a growth process.

## **The ARCC Approach to Capacity Building**

In the work of the ARCC at Edvantia, the capacity-building technical assistance provided to each State Education Agency (SEA) in a five-state region is preceded by the development of a detailed annual plan based on specific needs of the SEA and the multiple dimensions of capacity needed to conduct and sustain the work (Harsh, 2010). The ARCC plans are built on the premise that technical assistance needs to be carefully constructed and customized to address the complex work of the organization. Furthermore, the complexity of the work necessitates a multidimensional approach to capacity building.

A multidimensional approach to capacity building recognizes that organizational systems (e.g., state education agencies) are complex and require multifaceted analysis, implementation, and monitoring of the various dimensions involved in the capacity building process. Land, Hauck, and Baser (2009) posited that organizations were not only complex, but also adaptive systems that adjusted and changed in order to build sustainability. As a result, approaches to capacity building that do not acknowledge the complexities and adaptive nature of the system are insufficient for complex organizational transformation and renewal. Instead, capacity building should be viewed as an emergent property that involves multiple processes and attends to the complex interactions among the components of the system. Land et al.(2009) indicated that external intervention could provide significant support for capacity building if the basis of the intervention was on the right approach—one in which goals were clear to all involved, needs were integrated into the solution, and multiple aspects of capacity were considered when designing and implementing the initiative.

The basis of the ARCC's multidimensional approach to capacity building is Banathy's (1996) three-dimensional model of designing and implementing organizational change. The Banathy model addresses the critical dimensions that need consideration when designing and refining systems that must function in a changing world. The model is comprehensive in that it incorporates all possible configurations of change initiation, and inclusive in that it considers designs that change parts of the system, change the whole system, or create a new system. Banathy approaches system design and capacity building as a multidimensional inquiry that yields multiple perspectives on the status of the organization. He notes that, because complex systems have many interacting elements, the use of multiple perspectives is essential to ensuring that judgments, decisions, and design choices are viable for making the desired organizational changes. The ARCC's

multidimensional approach to capacity building addresses the same need to gain a thorough understanding of the system needs to design and implement appropriate capacity building technical assistance.

In the ARCC, Levels of Use (LoU) established in the Concerns Based Adoption Model (Hall & Hord, 2011; Hall, Dirksen, & George, 2006; Loucks, Newlove, & Hall, 1975) has been used to conceptualize capacity building as a development process. The LoU portray eight discrete developmental levels of behaviors in which individuals engage as they become more familiar with and skilled at using an innovation (see Figure 2). These levels range from nonuse to “an active, sophisticated, and highly effective use” (Hall, Loucks, Rutherford, & Newlove, 1975, p. 5) of the innovation or beyond, in which an individual explores major modifications to increase client impact. Levels of Use include nonuse, orientation, preparation, mechanical, routine, refinement, integration, and renewal. In addition to the levels, seven categories “represent the key functions that users carry out when they are using an innovation” (Hall et al., 1979, p. 6). Category descriptions at each level represent typical user behaviors. The seven categories include knowledge, acquiring information, sharing, assessing, planning, status reporting, and performing.

Users	VI	<b>Renewal:</b> State in which the user re-evaluates the quality of use of the innovation, seeks major modifications of or alternatives to present innovation to achieve increased impact on clients, examines new developments in the field, and explores new goals for self and the system.
	V	<b>Integration:</b> State in which the user is combining own efforts to use the innovation with related activities of colleagues to achieve a collective impact on clients within their common sphere of influence.
	IVB	<b>Refinement:</b> State in which the user varies the use of the innovation to increase impact on clients within immediate sphere of influence. Variations are based on knowledge of both short- and long-term consequences for clients.
	IVA	<b>Routine:</b> Use of the innovation is stabilized. Few if any changes are being made in ongoing use. Little preparation or thought is being given to improving innovation use or its consequences.
	III	<b>Mechanical Use:</b> State in which the user focuses most effort on the short-term, day-to-day use of the innovation with little time for reflection. Changes in use are made more to meet user needs than client needs. The user is primarily engaged in a stepwise attempt to master the tasks required to use the innovation, often resulting in disjointed and superficial use.
Nonusers	II	<b>Preparation:</b> State in which the user is preparing for first use of the innovation.
	I	<b>Orientation:</b> State in which the user has recently acquired or is acquiring information about the innovation and/or has recently explored or is exploring its value orientation and its demands upon user and user system.
	0	<b>Nonuse:</b> State in which the user has little or no knowledge of the innovation, no involvement with the innovation, and is doing nothing toward becoming involved.

Figure 2. Levels of Use (Hall & Hord, 2011).

Each level of use entails the cultivation and employment of increased knowledge and skill along with the ability to implement the innovation in advanced ways. The CBAM model is especially helpful in conceptualizing a framework for organizational change, while the Levels of Use provide a useful frame for identifying the three types of capacity that need development at the individual level. In Level I (orientation) and Level II (preparation), the individual acquires information and knowledge about the change innovation. The capacity to know and

understand the characteristics and components of the proposed change addresses the first individual capacity requiring development. Level III (mechanical) and Level IVA (routine) address the initial implementation of the proposed change and describe the second essential capacity—the ability to apply or use the information and skills acquired. Level IVB (refinement), Level V (integration), and Level VI (renewal) move the individual to a higher level of use, one that requires the capacity to create or act on opportunities to use the innovation. The capacity to be proactive encompasses the ability to refine, integrate, and renew the use of the change innovation.

As ARCC staff design and implement capacity building initiatives, the skill acquisition at each level of use is an important variable that helps technical assistance providers construct appropriate activities. The degree to which individual capacity increases is the area of interest and inquiry for the ARCC's benchmarking initiative.

## **ARCC Model for Benchmarking Capacity Building**

During Year 5 of the ARCC, staff met to discuss the inception and operationalization of a benchmarking initiative for specific ARCC initiatives to gauge the SEA staff levels of implementation. After several discussions, staff decided to pilot using the CBAM Levels of Use methodology (Hall et al., 1975; Loucks et al., 1975) to frame this effort.

### **ARCC Benchmarking Methodology and Validation Process**

ARCC staff trained and certificated in the LoU methodology used the tool to identify SEA levels of use. As a preliminary step, staff members used the ARCC Capacity Building Monograph (Good, Harsh, & Bumgardner, 2010), which describes successful technical assistance initiatives completed during the 2005-2010 period, to glean evidence for assigning retrospective LoU ratings to 16 initiatives (one in Kentucky, three in North Carolina, five in Tennessee, three in Virginia, and four in West Virginia). For this activity, staff reviewed the monograph narrative for evidence aligning with the seven categories and eight levels, and determined an overall LoU rating based on the category ratings. To ensure inter-rater reliability, staff rated levels of use for the SEA initiatives independently and then met to discuss their findings. Further, at least two staff members rated each initiative. When rating discrepancies occurred, staff discussed their ratings and reached consensus. These discussions helped ensure congruence in future ratings.

To validate LoU results from the monograph-based rating activity, staff used the LoU protocol to conduct telephone interviews with the key SEA staff member from a sample of three initiatives (one in North Carolina, one in Tennessee, and one in Virginia). In the interviews, staff posed a standardized set of questions aligned with the categories. Responses to the questions determined LoU ratings for each category, which led to an overall LoU rating for each interviewee based on the category ratings. Staff selected the North Carolina e-Coaching, the Tennessee District Appraisals, and the Virginia Assisting Low-Performing Schools initiatives for the pilot. A key SEA staff member for each initiative received an e-mail invitation to participate in a telephone interview. Each interview was audio recorded with two staff members rating each interview and discussing similarities, differences, and areas of uncertainty before producing an overall rating.

Findings from the LoU interviews were triangulated with those from the monograph-based rating activity. The graphs (Figures 3-5) and narrative below summarize the ratings based on the monograph narrative and interview data, as well as a reflective summary of the staff members' experiences with this activity.



## North Carolina e-Coaching Initiative

**Monograph rating.** Professional development for North Carolina Department of Public Instruction (NCDPI) staff is the foundation for building capacity to assist in implementing No Child Left Behind (NCLB). To that end, ARCC's work with North Carolina focused on e-Coaching to support the Sheltered Instruction Observation Protocol (SIOP), a model used to address instruction for English Language Learner (ELL) students. E-Coaching provides a range of opportunities that can support state initiatives in meeting NCLB requirements. The purpose of e-Coaching is to use available technologies to support state coaching efforts. E-Coaching provides access to skills, knowledge, and expertise that might otherwise not be feasible, available, or appropriate. ARCC's e-Coaching project began in Year 3 and continued through Year 5 as ARCC built the NCDPI's capacity to support its own e-Coaching programs.

In Year 3, the NCDPI needed to provide follow-up training and support to district and school staff trained in SIOP within the constraints of time and funds. ARCC staff recommended that the NCDPI consider electronic means to provide professional development; however, NCDPI staff did not possess the knowledge or experience to develop and deliver online professional development. They were at the non-use LoU of e-Coaching as a form of professional development. With ARCC's modeling and technical assistance, online book studies became a method for follow-up training and support for the SIOP coaches. By the end of ARCC's fifth year, the NCDPI had developed the capacity to enhance district and school leadership and systems of support via a cost effective, accessible, and user friendly means of professional development. The NCDPI successfully planned and facilitated an e-Coaching book study. This reflected a mechanical LoU of the innovation.

**Interview rating.** As a result of the North Carolina LoU interview on the e-Coaching initiative, researchers identified a number of issues for deliberation and reflection. First, the importance of defining the innovation within common parameters to provide both Edvantia researchers and NCDPI staff members with the same concept of the initiative and its boundaries. Second, both researchers and NCDPI staff needed a common understanding of the overall initiative as well as the specific components embedded within. Third, the pilot activity was not without limitations—a key limitation was the time lag between data collection and reportage that informed the documentation used for the original rating activity and the time of the actual NCDPI LoU interview. The initiative became a moving target given the amount of time that lapsed between the two data points, and the e-Coaching effort was no longer at the same place. As a result, the NCDPI interviewee's perceptions colored the current understanding and involvement with the initiative as it exists, not as it was when reporting earlier data. Therefore, the overall rating for the North Carolina e-Coaching initiative, based on the LoU interview, indicated a movement from the non-use LoU prior to ARCC involvement to the preparation LoU at the end of Year 5.

The monograph rating resulted in a mechanical LoU while the interview rating resulted in the preparation LoU. This slight decrease illustrates how perceptions differ over time, how initiatives and components within evolve, how both researchers and NCDPI staff may view the work through different lenses, and the subsequent difficulty in reconciling differences in ratings. See Figure 3 for a visual depiction of the difference between the monograph-based and interview-based ratings for the initiative at the end of Year 5.

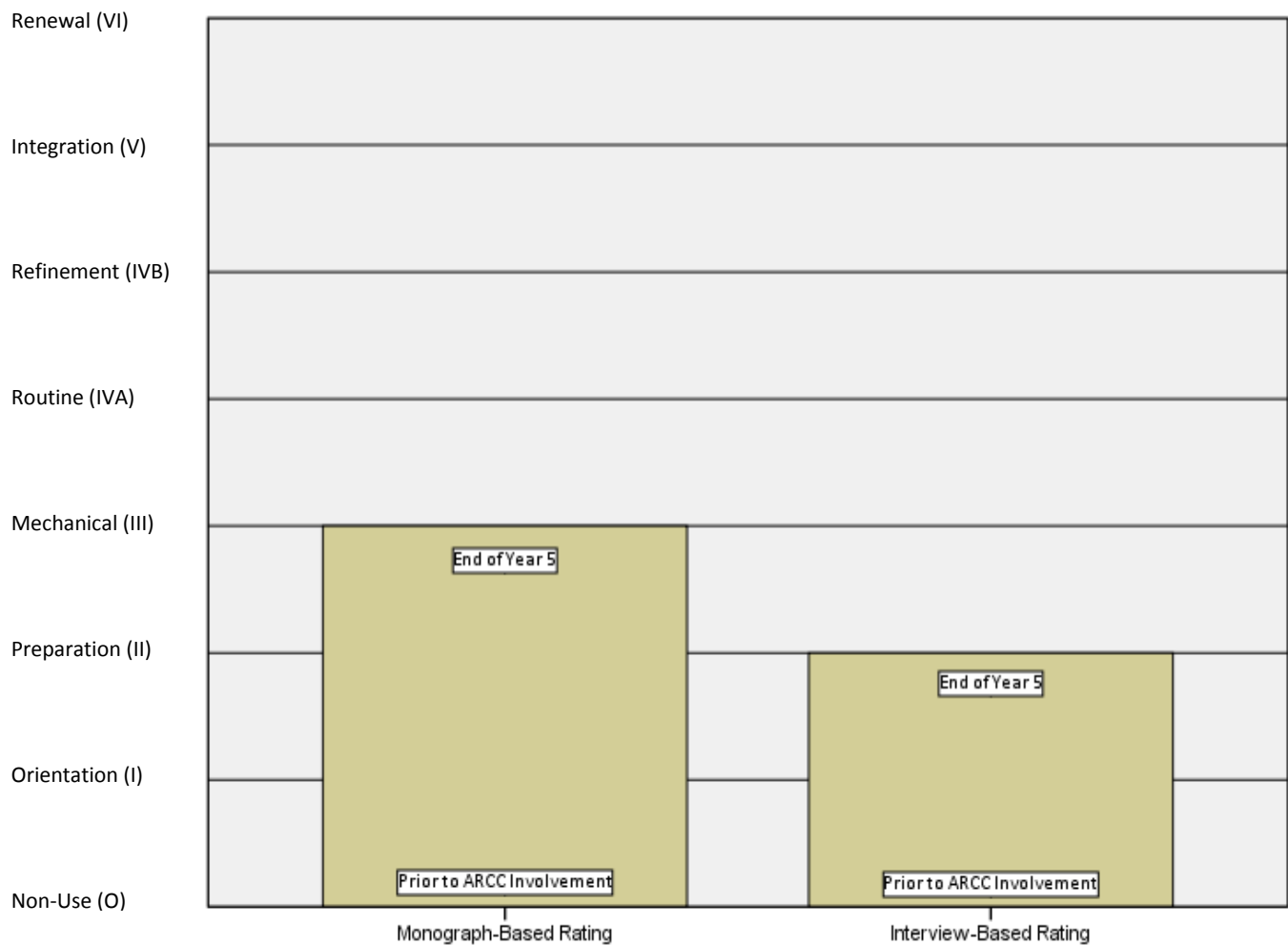


Figure 3. North Carolina e-Coaching Initiative.

### Tennessee District Appraisals Initiative

**Monograph rating.** The ARCC worked closely with the Tennessee Department of Education (TDE) to refine the Tennessee Statewide System of Support (TSSS) for the past five years. The overall goal of this initiative was to build the capacity of the TDE to provide assistance to targeted districts and schools through a high quality state system of support. One important element of the TDE SSOS was the district appraisals. The TDE wanted to create a districtwide appraisal process for high-priority school systems and piloted the District Audit Tool, developed jointly by the Council of Chief State School Officers (CCSSO) and Edvantia. The TDE, a team of recently retired educators, and ARCC staff piloted the tool with four districts in Year 1. The group customized the district audit process to meet the needs of the TDE. In Year 3, the ARCC and TDE conducted an audit in the Metro Nashville Public School District. Considering that the innovation focused on only the district appraisal component of the TSSS, the TDE was at the non-use LoU prior to the ARCC's involvement. At this point, the TDE has conducted, with the assistance of the ARCC, several district audits including the initial pilots and attained the mechanical LoU.

**Interview rating.** Review of extant data sources for the Tennessee District Appraisals initiative resulted in a mechanical LoU rating. However, after analyzing the data collected from a telephone interview using the LoU protocol, it was apparent that the TDE was moving to the refinement LoU. The main reasons for the difference in rating were the way in which the innovation was defined and the intervening time between ratings. ARCC researchers narrowly defined the innovation as the CCSSO district appraisals piloted in four districts in Year 2 and conducted in a third district in Year 3. TDE staff conceptualized the CCSSO district appraisals as a part of a larger statewide system of support initiative and moved on to the next stage of the initiative—developing their own district appraisal process. There was evidence throughout the LoU interview that Tennessee was refining an existing process to meet their needs and did not intend to replicate the CCSSO district appraisals. However, it is important to note that, although the LoU interview with the SEA client showed evidence of moving to refinement of the innovation, the work was on hold and has been for the last several months due to major leadership changes within the TDE. Given the intention of the TDE to continue the project, it seems appropriate to consider the TDE at the refinement LoU; however, this does raise the question of whether this is an appropriate classification on the LoU matrix given the *hold* status. See Figure 4 for a visual depiction of the difference between the monograph-based and interview-based ratings for the initiative at the end of Year 5.

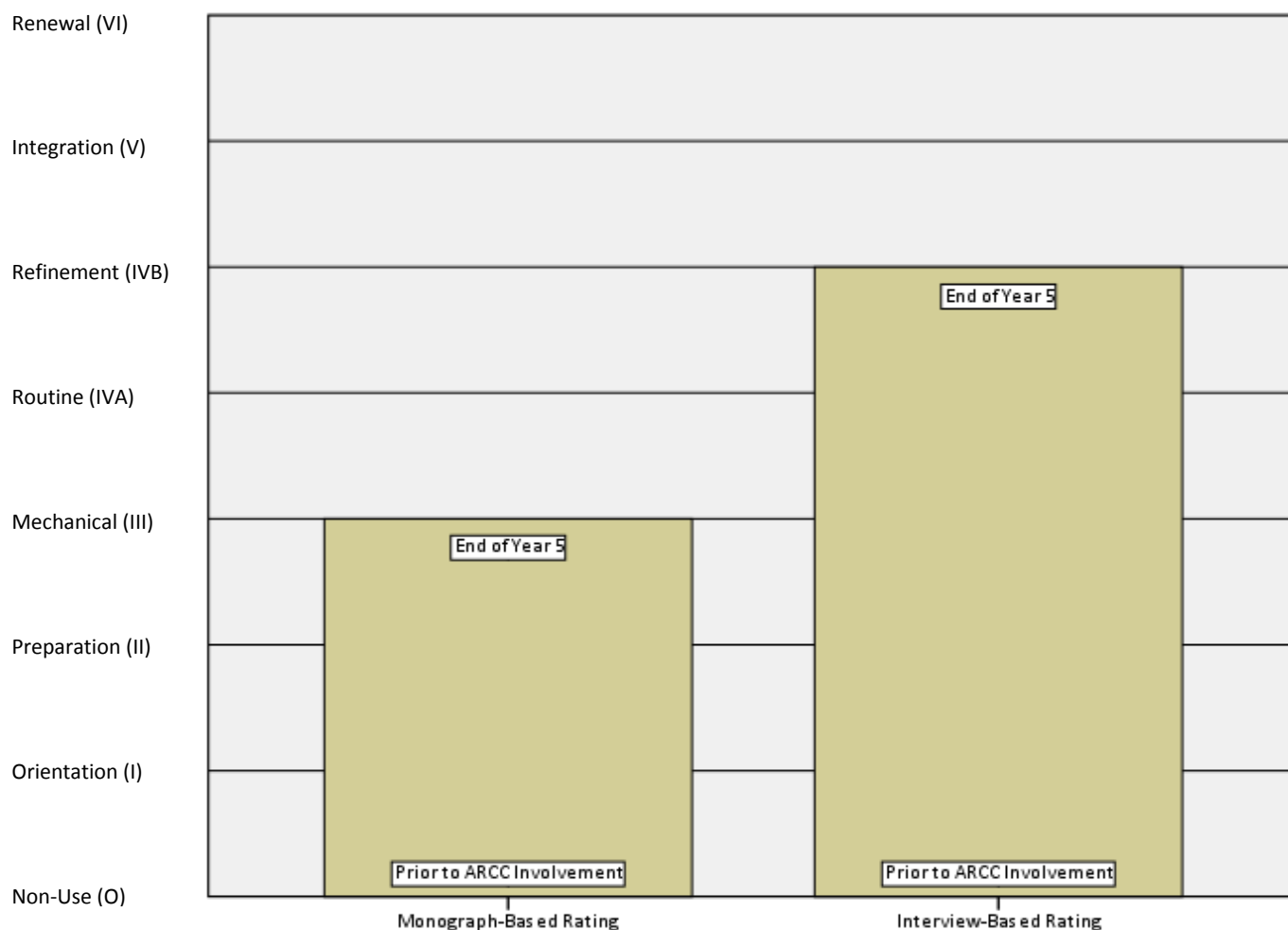


Figure 4. Tennessee District Appraisals Initiative.

## Virginia Assisting Low-Performing Schools Initiative

**Monograph rating.** The Virginia Department of Education (VDE) determined that it needed a broader based and systemic approach to school improvement. VDE leaders wanted to shift to a statewide system of support to achieve wider and deeper improvement. As a result, the VDE—with support from the Virginia Board of Education and in active collaboration with the ARCC, the Center on Innovation and Improvement (CII), and the Virginia Foundation for Education Leadership—initiated an effort to support schools under sanctions imposed by the state or NCLB. The expanded expertise, experience, resources, and perspective brought to the table by the ARCC and the other external partners enriched and accelerated the quality, quantity, and depth of Virginia’s capacity to create a statewide school improvement initiative. In so doing, the state demonstrated the viability of the regional comprehensive centers and content center partners working directly with SEAs to build capacity.

This initiative focused on developing the VDE’s capacity to assist division schools that were in or approaching restructuring under NCLB or VDOE sanctions. In Year 3, the VDE became aware of CII’s *The Handbook on Restructuring and Substantial School Improvement* and wanted to use it as a base for building the VDE’s capacity to assist low-performing school divisions identified by the state. Prior to receiving the handbook, the VDE had no knowledge of this information (the non-use LoU). During Year 3, a collaborative relationship emerged between the VDE, ARCC, and CII. Year 3 efforts focused on building the capacity of VDE staff to train targeted divisions in using the handbook and associated processes via a web-based approach. The collaboration focusing on building SEA capacity to support low-performing division schools continued and expanded into Years 4 and 5. By Year 5, the Virginia Assisting Low-Performing Schools initiative grew significantly, targeting 73 schools in 41 districts. The plan for 2009-2010 emerged from considerable reflection on Year 4 and careful reviews of VDE and ARCC evaluations of each component and training session. VDE staff refined and modified materials and processes from the preceding two years into the third year (refinement LoU).

**Interview rating.** Comparison of LoU ratings derived from analysis of extant data and a VDE telephone interview for the Virginia Assisting Low-Performing Schools initiative exhibited no differences in ratings. Analysis of data from both sources indicated that the VDE was at the refinement LoU. Despite the consistency in ratings, there were still some interesting observations. As the interviewer proceeded to the set of questions concerning coordination and collaboration (LoU V probes), it became apparent that the VDE was not changing their use of the innovation (the innovation in the way that ARCC researchers defined it) based on input of and in coordination with what colleagues were doing. Collaboration occurred (i.e., through the involvement of Virginia in the CII Pacesetter’s Group) among VDE team members and across the SEAs in the Pacesetter’s Group. But this collaboration was tangential to the initiative as defined by ARCC researchers. Examining the data collected via the LoU interview around sharing indicated the VDE was at a lower level of use for this particular category; however, when reviewing the totality of the interview data, the VDE was at the refinement LoU. As with the other two initiatives showcased, the criticality of defining and specifying the boundaries of the initiative so that both the interviewer and the interviewee have the same view was apparent. Similar to the other two initiatives, the Virginia initiative expanded and evolved since review of the extant data. The initiative became a moving target and measurement of the innovation through review of extant data differed from the innovation characterized in the mind of the VDE staff during the LoU interview. See Figure 5 for a visual depiction of the monograph-based and interview-based ratings for the initiative at the end of Year 5.

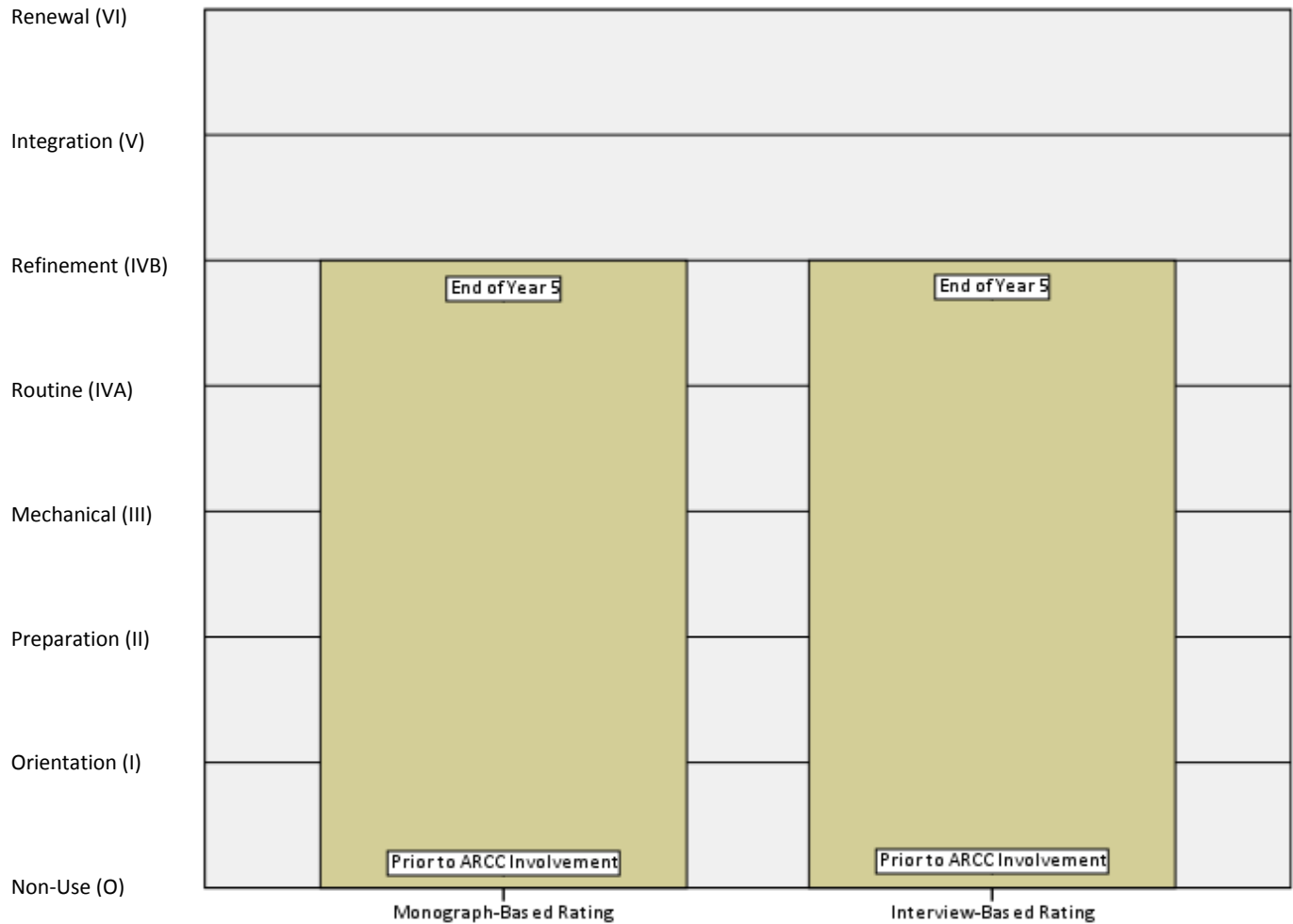


Figure 5. Virginia Assisting Low-Performing Schools Initiative.

## Limitations and Lessons Learned

ARCC researchers recognize the importance of clearly defining the *it* (the specific capacity building initiative) in order for ARCC and SEA staff to share an understanding of the initiative. In some instances during the pilot, the SEA interviewees had different parameters of the initiative, which, in turn, colored their perceptions of their levels of implementation. Discrepancies existed in two of the three LoU interview ratings compared to the ratings assigned based on the monograph review—in both situations, it related to differences in the definition of the *it*, or the initiative. In one situation (e-Coaching), ARCC researchers viewed the initiative to be much larger in scope than did the SEA staff member. As a result, ARCC researchers rated the LoU higher. In the second situation, ARCC evaluators narrowly defined the district appraisal initiative as well as the SEA's intended use. This resulted in a lower LoU rating from the interview compared to the one assigned from the monograph.

We recognize that, in a one-year period, the levels of use should be incremental because the objectives/activities (whatever the initiative or unit of analysis is) are complex. This has implications for tracking

the levels of use over an extended period; however, it is important to note that the initiative sometimes evolves as the SEA moves to the next level in the activity/objective (not LoU levels) with regard to the need for information, developing structures or processes, doing the work, etc. As a result, the levels of use might not follow a linear path. As the capacity building initiative undergoes adaptation, the implementers may revert to the orientation, preparation, or even non-use levels without progressing through all subsequent levels.

We recognize the time lag between when ARCC staff created their ratings from the monograph and the pilot LoU interviews. As a result, ratings may differ due to ongoing progress of specific initiatives. Further, during the interviews, SEA staff members focused on the most recent or upcoming activities, rather than looking back to where they might have been at the time at which the monograph ratings were generated—end of Year 5 or even earlier for some projects.

Throughout this pilot, it has become clear that the LoU ratings can focus on multiple levels. Pilot LoU interviews with SEA staff revealed both individual and organizational level behaviors.

## **Recommendations**

We recommend continuing the pilot of the benchmarking initiative in Year 6. We suggest selecting one initiative from each state. Further, because changes in priorities often occur in SEAs, we will identify an alternate for each state.

In addition to conducting LoU interviews with key SEA staff, we suggest adding a second activity to the benchmarking methodology. In order to triangulate data across sources, we suggest including LoU interviews with the ARCC state liaison for each respective project (state liaisons are those ARCC staff members who serve as the primary point of contact with a SEA and who manage the technical assistance initiatives). We recommend that these LoU interviews be conducted at two points in time during Year 6—near the middle of the year (March 2011) and near the end of Year 6 (August 2011).

During the Year 6 pilot, we should prepare and share brief descriptions of each project/initiative *it* with each SEA staff member and state liaison in advance of their respective interviews. This will ensure that ARCC staff, SEA staff, and state liaisons have the same understanding of the parameters of the project.

The initial benchmarking pilot and its proposed continuation into ARCC's Year 6 of operation will focus on benchmarking individual capacity building of SEA staff using the LoU framework. On further discussion and reflection on our work, it became apparent that capacity building initiatives or the *it*, which we are benchmarking, involves not only individuals, but often the intent is for either collective or organizational capacity building around the particular innovation (i.e., initiative or project). As originally conceptualized and validated, Levels of Use should be at the individual level. Thus, an area for future study and development is expansion of the LoU model and determining LoU categories for the collective and organizational levels. Drawing on what we know about systems change, there is a need to focus on all three dimensions (i.e., individual, collective, and organizational) of capacity building.

## Collaboration with Consultants

The ARCC's benchmarking team had the opportunity to meet with Dr. Gene Hall, co-author of the Concerns Based Adoption Model and lead researcher for the Levels of Use, and Dr. Jeanne Century, developer of the framework for the four types of capacity building (1999), one of the dimensions used in the ARCC's multidimensional capacity building model, to discuss the benchmarking initiative and current findings.

Dr. Hall expressed interest and support for the ARCC's benchmarking work, and in follow-up correspondence, noted *"Rarely do I have the opportunity to spend the day with colleagues who have in-depth knowledge about CBAM in general and especially Levels of Use. I am honored that you and your colleagues have built Levels of Use into your organizing framework, plans, and activities. All the best as you continue your work on the ground and developing conceptual frameworks that are based in the literature and grounded in real world practices."* Both consultants critiqued several iterations of this brief and served as a sounding board for initial efforts. Further, Dr. Hall and Dr. Century are continuing their collaborating with ARCC at Edvantia staff to offer insight and guidance toward future efforts.

## References

- Banathy, B. (1996). *Designing social systems in a changing world*. New York: Plenum Press.
- Boring, R, Hendrickson, S., Forester, J., Tran, T., & Erasmia, L. (2008, April). *Issues in benchmarking human reliability analysis methods: A literature review*. Albuquerque, NM: Sandia National Laboratories.
- Camp, R.C. (July 1, 1993). *A bible for benchmarking, by Xerox*. Financial executive. Retrieved from <http://www.allbusiness.com/finance-insurance/399130-1.html>
- Camp, R. C. (1989). *Benchmarking, the search for industry best practices that lead to superior performance*. Portland, OR: ASQC Quality Press.
- Century, J. R. (1999). *Determining capacity within systemic educational reform*. (ERIC Document Reproduction Service No. ED434162).
- Committee for Oversight and Assessment of U.S Department of Energy Project Management, National Research Council. (2005). *Measuring performance and benchmarking project management at the Department of Energy*. Washington, DC: National Academies Press. Retrieved from [http://www.nap.edu/catalog.php?record\\_id=11344](http://www.nap.edu/catalog.php?record_id=11344)
- Council of Chief State School Officers (CCSSO). (n.d.) *Common core state standards initiative: In the states*. Retrieved from <http://www.corestandards.org/in-the-states>
- Dowd, A. C., Malcom, L. E., & Bensimon, E. M. (2009). *Benchmarking the success of Latino and Latina students in STEM to achieve national graduation goals*. Los Angeles: University of Southern California.
- Fridley, J., Jorgensen, J., & Lamancusa, J. (1997, November). *Benchmarking: A process basis for teaching design*. Paper presented at the Frontiers in Education Conference, Pittsburgh, PA.

- Fullan, M. (2008). *The six secrets of change: What the best leaders do to help their organizations survive and thrive*. San Francisco: Jossey-Bass.
- Good, K., Harsh, S., & Bumgardner, S. (2010, July). *Capacity building monograph: 2005-2010 ARCC initiatives*. Charleston, WV: Appalachia Regional Comprehensive Center at Edvantia.
- Hall, G. E., Dirksen, D. J., & George, A. A. (2006). *Measuring implementation in schools: Levels of Use*. Austin, TX: Southwest Educational Development Laboratory.
- Hall, G. E., & Hord, S. M. (2011). *Implementing change: Patterns, principles, and potholes* (3rd ed.). Boston: Allyn & Bacon.
- Hall, G. E., Loucks, S. F., Rutherford, W. L., & Newlove, B. W. (1975). Levels of use of the innovation: A framework for analyzing innovation adoption. *Journal of Teacher Education*, 26(1), 5-9.
- Harsh, S. (2010, April). *Gaining perspective on a complex task: A multidimensional approach to capacity building*. Charleston, WV: Appalachia Regional Comprehensive Center at Edvantia.
- Land, T., Hauck, V., & Baser, H. (2009). *Capacity change and performance: Capacity development: Between planned interventions and emergent processes: Implications for development cooperation* (Policy Management Brief No. 22). Maastricht: European Centre for Development Policy Management. Retrieved from [www.ecdpm.org/pmb22](http://www.ecdpm.org/pmb22)
- Lanier, L., & Saltzman, E. (1999). *Public education and Greenwood County: A benchmarking tool*. Greenwood Partnership Initiative and Greenwood Area Chamber of Commerce. Clemson, SC: The Strom Thurmond Institute of Government and Public Affairs, Clemson University.
- Loucks, S. F., Newlove, B. W., & Hall, G. E. (1975). *Measuring levels of use of the innovation: A manual for trainers, interviewers, and raters*. Austin, TX: Southwest Educational Development Laboratory.
- Maintenance Resources. (n.d.). The language of benchmarking. In *Benchmarking Fundamentals*. Retrieved from <http://www.maintenanceresources.com/bookstore/maintenancemanagement/wiremanpdf.pdf>
- National Governors Association, Council of Chief State School Officers, & Achieve, Inc. (2008). *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Washington, DC: National Governors Association.
- PHS Management Training. (2004). *Benchmarking for competitive advantage*. Retrieved from <http://www.training-management.info/PDF/benchmarking-training.pdf>
- UNESCO. (2007). Education counts. *Benchmarking Progress in 19 WEI Countries, World Education Indicators–2007*. Montreal: UNESCO Institute for Statistics. Retrieved from <http://www.uis.unesco.org>
- Wood, B. (n.d.). *7 steps to better benchmarking*. Retrieved from <http://bpmmag.net/mag/7-steps-better-benchmarking-0507/>