Invitational Research Symposium on
Through-Course
Summative Assessments

Four Years of Cognitively Based Assessment of, for, and as Learning (CBAL): Learning About Through-Course Assessment (TCA)

John P. Sabatini, Randy Elliot Bennett, and Paul Deane

Educational Testing Service

April 2011

Center for K – 12 Assessment
& Performance Management at ETS

FOUR YEARS OF COGNITIVELY BASED ASSESSMENT OF, FOR, AND AS LEARNING (CBAL): LEARNING ABOUT THROUGH-COURSE ASSESSMENT (TCA)

John P. Sabatini, Randy Elliot Bennett, and Paul Deane
Educational Testing Service

Overview

The goal of this paper is to describe some of the many lessons learned from a multiyear research and development (R&D) program aimed at building a model for innovative summative and formative assessment at the K-12 level. The program, Cognitively Based Assessment of, for, and as Learning (CBAL), is intended to generate new knowledge and capability that can be used in the near-term for the design, administration, and scoring of innovative assessments like those intended for use by the Common Core State Standards Assessment consortia. As of this writing, the project is in its fourth year, and its operation has required that the research team grapple with issues of through-course assessment (TCA) because such assessment is a central feature of the CBAL conceptualization. Because the lessons learned are derived from this conceptualization and experience, we will first describe some of the key elements of the CBAL program and the theory of action that guides the R&D agenda.[1] This description will set the groundwork for the discussion of lessons learned, in which we detail some of the reasoning, challenges, and design decisions we reached in designing TCA.

[1] For more information about the CBAL initiative, including full papers, see http://www.ets.org/research/topics/cbal/initiative

Cognitively Based Assessment of, for, and as Learning

The CBAL program intends to produce a model for a system of assessment that: documents what students have achieved (*of learning*); helps identify how to plan instruction (*for learning*); and is considered by students and teachers to be a worthwhile educational experience in and of itself (*as learning*). Towards achieving these goals, CBAL consists of multiple, integrated components: summative assessments, formative assessments, professional development, and domain-specific cognitive competency models (Bennett 2010; Bennett & Gitomer, 2009).

Each of the CBAL components is informed by, and aligned with, domain-specific cognitive competency models (which we describe in some detail in the next section). The summative assessments consist of multiple events distributed across the school year. The intention is that results will be aggregated for accountability purposes. The formative assessments consist of componential item sets, classroom tasks, and extended activities, as well as associated teacher guides that provide insights into techniques for integrating formative tasks into instructional units. In some cases, these formative tasks are based on developmental or learning progressions (Harris & Bauer, 2009; Heritage, 2008). Finally, the professional development component consists both of organized teacher communities of practice and a collaborative social networking website that further elaborates relationships among assessments, instruction, and the cognitive competency models.

Domain-Specific Cognitive Competency Models

Domain-specific competency models are developed with the goal of integrating learning sciences research, including learning progressions (where such progressions are available), with content standards. The competency models help not only in the specification of knowledge, processes, strategies, and habits of mind to be assessed, but also in identifying instructional-practice principles for use in assessment design. In CBAL, the models also serve as a common conceptual foundation for both summative and formative assessments.

Each model is derived from reviews of the cognitive and learning sciences literature in mathematics and in English language arts (ELA) reading and writing. That literature speaks to both student development and effective instructional practice. Recent versions of the models for middle-grades students can be found in Deane (2010), Graf (2009), and O'Reilly and Sheehan (2009). The models have been linked to the Common Core State Standards (CCSS). They are iteratively refined via collaborations with teachers and classroom pilot data as the CBAL project progresses through its multiyear research agenda.

Figure 1 illustrates the central role of the competency models in the overall assessment system. In CBAL, the competency models help to integrate learning science research with content standards such that the amalgam becomes the driver of assessment design (both formative and summative), the basis for an evidence-based curriculum, and the starting point for professional support to aid teachers in building their repertoire of research-based pedagogical practices. By grounding assessment, curriculum and instruction, and professional development in the same learning sciences and content standards foundation, we hope to facilitate the intended outcome of improved classroom practice. This approach also represents a deliberate attempt to ensure that learning sciences research, itself often informed by and encapsulated in the wisdom of practice, is better disseminated throughout the educational enterprise.

While carefully constructed and thoughtful content standards are important for setting appropriate targets for instruction, they often remain abstract, too far removed from informing good instructional practice. Consequently, assessments that are defined solely with respect to the content standards run the risk of having limited instructional relevance, and, additionally, may fail to account for results from decades of learning science research that can serve as a principled guide to implementing sound instruction. As we will discuss in more detail later, CBAL takes advantage of the opportunities afforded by through-course assessments to

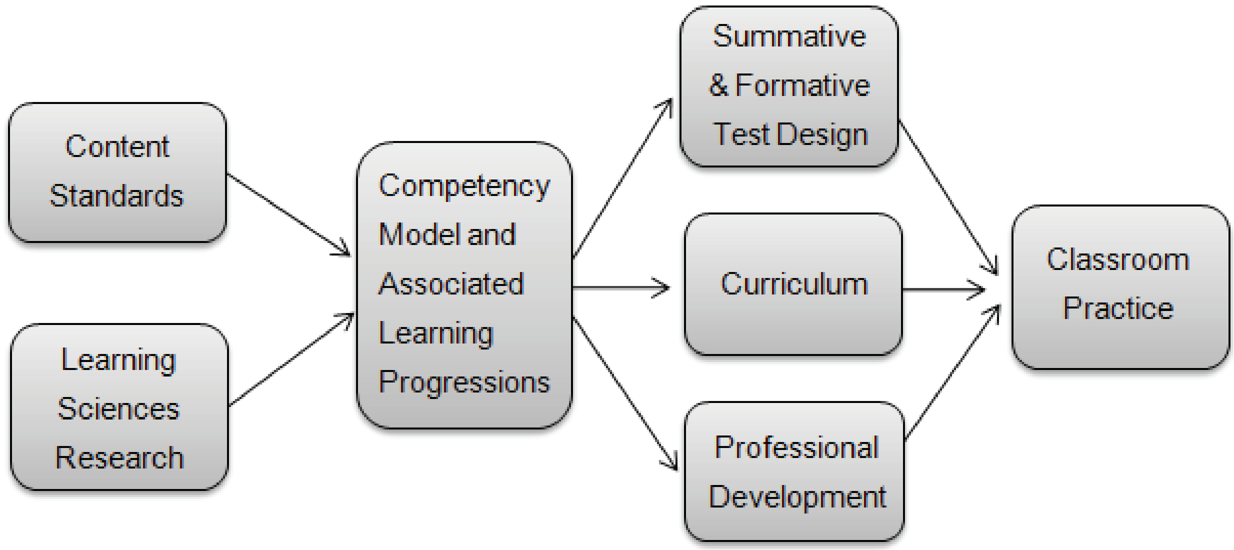


Figure 1. The role of competency models in CBAL.

Note. From "Cognitively Based Assessment of, for, and as Learning: A Preliminary Theory of Action for Summative and Formative Assessment," by R. E. Bennett, 2010, Measurement: Interdisciplinary Research and Perspectives, 8, pp. 70-91. Copyright 2010 by Educational Testing Service. Reprinted with permission.

instantiate design principles derived from this competency-model foundation that would otherwise be difficult to achieve in a single, comprehensive end-of-year examination.

Why Through-Course Assessments?

As noted, TCAs have been a foundational attribute of the CBAL initiative since its inception. Three primary aims justify the decision to use through-course assessments: first, the importance of any one assessment occasion is diminished; second, tasks can be more complex and more integrative because more time is available for assessment in the aggregate; and, third, the assessments can provide prompt interim information to teachers while there is time to take instructional action.

For the CBAL research program, one might say that the latter two aims are the most critical, as the first aim would generically be true of any through-course assessment system. However, one's goals for designing a TCA system need not be to deploy more complex, integrative tasks; likewise, one does not need to aim for providing interim information to teachers. TCAs serve CBAL precisely because they create the opportunity to deliver more complex, integrative tasks and to feed back information to teachers and learners. However, as we describe in detail later, these goals entail a set of design decisions and complexities that must be coordinated and managed.

Theory of Action

The CBAL system model is designed as an educational intervention, as well as an indicator of student achievement. As such, Bennett (2010) has described a theory of action for CBAL in terms of components, hypothesized action mechanisms, intended intermediate effects, and intended ultimate effects. The logic model summarizing this theory of action is shown in Figure 2. The theory of action guides assessment design and validation, but not just in the sense of the evaluation of score claims, but also in the evaluation of the intended impact of the assessment system on individuals and institutions.

Several ideas are important to note in Figure 2. First, the logic model makes clear that the ultimate goals (or intended effects) of CBAL as an assessment system are to provide more meaningful information to policy makers *and* to contribute to improved student learning. Second, these intended effects are caused by a set of intended intermediate effects. The latter effects target changes in teacher competency, classroom practice, and student engagement. Last, these intermediate effects are, in turn, caused by action mechanisms, each of which is associated with a particular CBAL component.

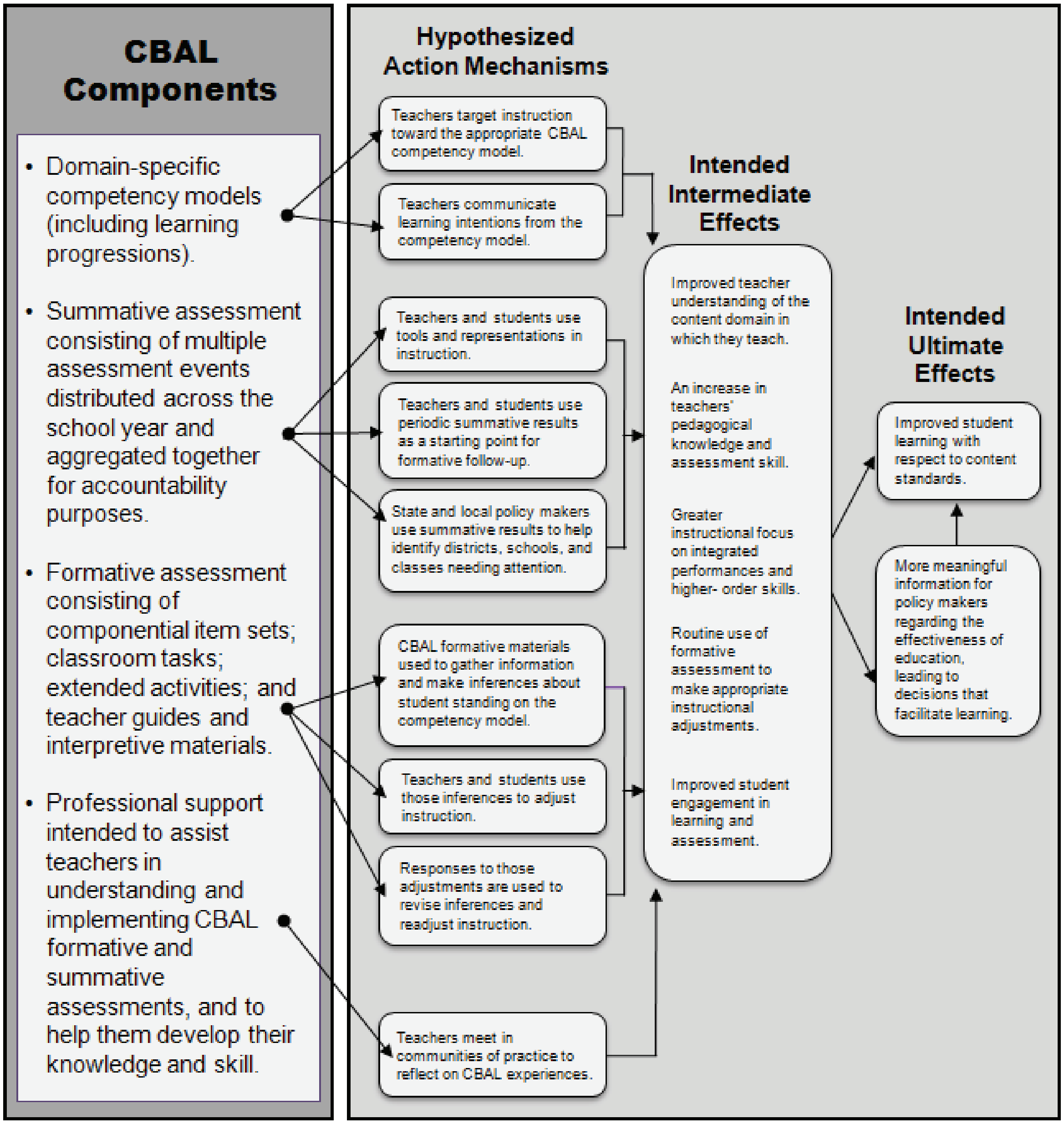


Figure 2. A logic model summarizing the CBAL theory of action.

Note. From "Cognitively Based Assessment of, for, and as Learning: A Preliminary Theory of Action for Summative and Formative Assessment," by R. E. Bennett, 2010, *Measurement: Interdisciplinary Research and Perspectives, 8*, pp. 70-91. Copyright 2010 by Educational Testing Service. Reprinted with permission.

Two action mechanisms are associated with the CBAL competency model and concern teacher use of that model to guide instruction and communicate learning goals. Three action mechanisms are linked to the CBAL summative assessments. For students and teachers, these mechanisms entail the instructional use of the tools and representations contained in the summative assessments and the use of summative results as a starting point for formative follow-up. For state and local policy makers, the action mechanism is the use of summative results to identify classes, schools, and districts needing administrative attention. The CBAL formative components have three associated action mechanisms concerning the making of inferences about student standing, the use of those inferences to adjust instruction, and the use of student responses to those adjustments to revise inferences and readjust. Finally, the professional support component has as its action mechanism participation by teachers in communities of practice to reflect upon their experiences with using CBAL to understand and improve student performance.

Some Lessons Learned

In the remainder of the paper, we describe a few of the many lessons learned from CBAL research that may be helpful to designers of through-course assessments. The unifying theme concerns careful consideration of the specific purposes to be achieved by using a TCA approach (and there are likely to be more than one) and how those purposes influence the actual content and design of the assessments themselves. We would not advocate targeting more than a few purposes, as optimizing to the set reduces the effectiveness of achieving each individual purpose. Tradeoffs are inevitable, and a manageable set of purposes, especially at the onset of a complex project with high stakes for all involved, is a prudent course to take.

Evidence Sources for Lessons Learned

To test out the CBAL theory of action would require not only the delivery of TCA at different points in time, but also the implementation of all of the CBAL components in authentic settings, including the use of using results for accountability purposes. Such a scenario is well beyond what an assessment research program like CBAL can hope to achieve. In keeping with the idea of creating a system *model*, only parts of the CBAL system have been developed and studied. Parts of the model that have been developed include the extensive reviews of the cognitive science literature that constitute the basis for the cognitive competency models; the creation of prototype assessments through collaborative activities with talented educators to help ground design in the wisdom of practice; iterative pilots in field sites to learn about fit in the types of environments within which the full system might operate; and linking of CBAL assessment prototypes and competency models to the Common Core State Standards. In 2009, the CBAL team conducted a multistate trial in which two reading or two writing summative assessments were administered to the same (seventh and eighth grade) students close together in time. Across 16 CBAL pilots, nearly 10,000 online tests have been administered. Psychometric results from those pilot administrations are reported by Bennett (in press). Finally, a second multistate study is being conducted with through-course assessments being administered at two points in time—winter and spring semesters of 2011. This experience provides several important lessons about the design of through-course assessments.

Lesson 1: Clearly articulate the intended purpose(s) for through-course assessment and use those purposes to drive assessment design. In the CBAL program, we prioritized two key aims for summative assessments. First, we sought to design summative and formative assessments that would function as useful measurement tools. Second, we designed those assessments so that they would be considered by students and teachers to be worthwhile educational experiences in and of themselves, experiences that would promote the development of higher-order thinking skills demanded for success in the English language arts and mathematics content domains. Setting these as our primary goals guided a variety of decisions and ultimately led us to specific design principles.

Our examination of the learning sciences research in each domain helped identify models of effective learning and instructional approaches and how those models might lead to the development of proficiency in critical knowledge, processes, strategies, and habits of mind. That examination also produced examples of the kinds of activities and tasks that require students to reason and problem solve in the domains in question. The insights gained provided a lens by which we could observe skilled practitioners to better appreciate variations of good instruction. Seeking to capture these practices in assessments led us to two key design principles that now undergird most all of our assessment designs, and that fit well with the TCA approach.

Scenario-based task sets. Such task sets are composed of a series of related tasks that unfold within an appropriate social context. The goals include: to communicate how the tasks fit into a larger social activity system; to set standards for performance; to give test takers a clearer idea of how to allocate attention and give focus to their deliberations; to provide opportunities to apply strategic processing and problem solving; and to have learners evaluate and integrate multiple sources of information in a meaningful, purpose-driven context. The scenarios are created to focus on targeted nodes in the respective competency models, but also to permit the integration of other knowledge and skills that may be prerequisite or co-requisite in performing tasks in the domain. Below we provide brief examples of scenarios from the three domains which CBAL has explored: mathematics, reading, and writing (see Appendix for example screen shots of scenarios).

As Harris and Bauer (2009) explain, the CBAL mathematics prototypes utilize scenario-based task sets that draw from at least two content areas in the competency model. The scenario functions not as simply a setting but, rather, drives the design of the task set (see also Harris, Bauer, & Redman, 2008). For example, one scenario involves a region experiencing drought, with particular focus on a lake whose receding water levels may no longer be high enough to exit the dam. The focus of this task set is the cross-cutting mathematical process of argument. The content strands of linear functions and statistics are also drawn upon. The introductory activity sets up the big question or idea that the students will have to address at the end of the task set, "Does action need to be taken about the water crisis?" Students are provided with an explanation of why the lake is important to the community, that is, because it is used to produce electricity and provides water for crops. A series of tasks is then presented that leads students through the problem to a culminating task calling for a judgment about whether action needs to be taken and evidence to back that assertion.

A similar scenario drives one of the ELA reading assessment designs (O'Reilly & Sheehan, 2009; Sheehan & O'Reilly, in press). Students are introduced to a scenario-based task set in which a wind farm has been proposed for their community, and their class has decided to create a website to help members of the community to be more informed about wind power. The scenario unfolds across a series of tasks addressing the questions:

- How does wind power work?
- What are some possibilities and challenges of using wind power as an energy source?
- Is the proposed idea good for the community?

Related readings drive each subsection, with a combination of selected- and constructed-response items to which the student must respond.

Finally, a scenario-based writing assessment covering the writing competencies of summarization and argumentation addresses the question of whether there should be a ban on advertisements directed at children (Deane, Sabatini, & Fowles, 2011). In conjunction with several readings on the topic, students work their way through tasks that require them to: apply the points in a rubric to someone else's summary of an article about children's advertisements; read and summarize two articles about the issue; determine whether statements addressing the issue are presenting arguments pro or con; determine whether specific pieces of evidence will weaken or strengthen particular arguments; critique someone else's argument about the issue; and, finally, write an argumentative essay taking a position on children's advertising.

While the CBAL scenario-based task sets share a heritage with earlier performance assessments in education, there are important design distinctions. Specifically, earlier performance assessments tended to be composed of a smaller number of more highly interdependent tasks delivered under less formal conditions and without the use of technology (e.g., conducting a specific science experiment and writing up the results). Like these assessments, CBAL scenario-based task sets share in the goal of creating a more authentic, meaningful, and purposeful context for deploying one's knowledge and skills. Furthermore, scenario-based tasks provide an opportunity to assess understanding of key related content at deeper levels than discrete questions can; therefore, it is essential to target content identified as critical to assess. However, the CBAL task sets are designed to gather information about particular constellations of skill in the competency models, as well as how those skills are integrated into a more complex performance. Thus, in each of the CBAL scenario-based task sets described above, there are items that test discrete skills, as well as more complex interrelated tasks most appropriately scored with a holistic rubric. The TCA design allows for the assessment of a broad range of content when aggregated across multiple administrations.

Scenario-based task sets help in achieving a foundational learning sciences principle of contextualizing skill and knowledge as they are applied by expert practitioners in a domain, rather than asking students to recall isolated facts or execute procedures absent any meaningful context. In this way, CBAL assessments can better serve as worthwhile learning experiences because they can help students connect knowledge, processes, strategies, and habits of mind to conditions of use. The tradeoff is that engaging students in a scenario requires that assessment time be spent setting up the purpose and allowing students to deliberate, reason, and reflect on the tasks with respect to that purpose. In general, however, the TCA design allows for simultaneously achieving depth by using a focused problem set within an individual TCA and breadth by covering the broader set of required domain-competencies across TCAs.

Tools and representations. The goal of including innovative tools and representations derived from domain practice is to get the most accurate estimate of the student's achievement *and* to model good teaching and learning. In the category of tools and representations, we include rubrics and guidelines providing explicit information about how the performance will be judged; tips, checklists, and graphic organizers providing direct models of what kinds of strategies are deployed by successful performers; appropriate reference materials and devices to support comprehension and thinking; and simulations that encourage exploration and understanding of conceptual relationships.

As an example, Harris and Bauer (2009) describe simulation tools used in the math scenario-based task set described above. To assist students in becoming familiar with inflow and outflow in the context of a dam and lake, a simulation was developed that allowed students to experiment with inflow and outflow rates and their effect on the volume of water in a sink. The simulation presents a familiar setting where students can set the rate of inflow from the faucet and outflow by manipulating the drain plug. Questions accompanying the sink simulation require students to interpret graphs of inflow/outflow rates and describe the effects on water volume. After several such tasks, the problem context returns to the main question around the viability of the lake.

In the wind power reading example discussed above, a student must complete graphic organizers designed to probe her or his understanding of scientific text explanations, for example, the difference between windmills used to generate electricity from wind versus the operation of household fans that use electricity to generate wind. In another question, the student must complete a graphic organizer, which helps in probing his or her understanding of the organization and structure of information in a text, both in aligning details with main ideas and in inferring or inducing topical categories.

The children's advertising writing scenario includes rubrics for evaluating a summary, as well as activities in which students are asked to use those rubrics to examine simulated peer summaries to identify whether they adhere to or violate specific rubric elements (e.g., inserting one's own opinion into a summary). In another portion from the same scenario-based task set, students evaluate a series of statements as pro or con and judge whether specific claims are warranted. These elemental skills of argumentation are highly predictive of performance on the culminating essay task, but also model thinking that is foundational to the formulation of a persuasive argument.

In each of these examples, the learning sciences literature reveals insights into the cognitive strategies that skilled individuals use in proficiently performing complex tasks in a domain. Including these tools and representations in the assessment calls upon the student to demonstrate strategic processing using devices common in domain practice. Further, that inclusion encourages the student and teacher to incorporate such tools and representations into classroom practice and, more generally, to develop the reasoning and strategic behavior required to successfully use similar tools and representations more broadly in domain performance. Further opportunities to use such tools and representations are provided in the CBAL formative assessments, which offer elaborated task sequences that cover the competency models more deeply than a summative assessment could.

Lesson 2: Use a theory of action to guide the design and evaluation of through-course assessments*.* In the CBAL theory of action (Bennett, 2010), the assessment system as an intervention becomes a key part of what it means to demonstrate *technical quality*. Technical quality as such is not just instrument functioning; it is also the impact (negative and positive) of instrument use on students, teachers, classroom practice, school functioning, and the larger education system as a whole (Bennett, Kane, & Bridgeman, 2011). Thus, the theory of action becomes a key component in assessment design and in evaluating the success of that design through its implementation and impact. Following is a select set of examples of how a theory of action (as depicted in Figure 2) fits into the design of TCA.

Theory-of-action states as an intermediate outcome: Teachers and students use periodic assessment results as a starting point for formative follow-up. This outcome has several very specific design implications. First, it suggests that results from assessments given during the year must be scored and reported with a reasonable turnaround for student and teacher use. Selected-response items are most efficient, as they can be scored nearly immediately when administered electronically. Constructed-response turnaround time can also be rapid with the strategic use of automated scoring. One of the active research areas in the CBAL program concerns the development and evaluation of natural language processing (NLP) approaches to the scoring of essay and other writing tasks (e.g., Deane, in press). It may be that some, but not all, of the responses to tasks composing an assessment can be scored immediately, allowing some types of instructionally relevant results to be provided quickly to teachers and learners. Results requiring greater levels of quality control and statistical postprocessing would be reported later. Such a phased approach to reporting may serve the purpose of providing instructionally actionable information in a reasonable time period.

Second, the outcome obviously suggests that score reports must be designed to encourage valid inferences about performance. Valid inferences may need to be couched as qualified interpretive claims, that is, *formative hypotheses* (Bennett, 2010). A formative hypothesis is a qualified statement suggesting that the teacher collect follow-up evidence to confirm or refute the hypothesis. This idea is rooted in the fact that it is not often possible to derive from a summative test sufficient information to support a reliable inference about an individual’s skill strengths and weaknesses. Expecting summative assessments to provide individual diagnostic information is often a bridge too far. For groups, it may be more feasible to make test-based inferences regarding relative mastery or deficiency of subskills (e.g., when nearly all students in class answer correctly or incorrectly all items in a specific node of the competency model or standard), but even this inference may be weakened by the underrepresentation of certain skill areas on the test such that teachers may still need to do additional informal data gathering to confirm the suggestion.

Formative tools and processes designed to generate additional student or classroom information might be used by teachers to carry out the needed follow-up. These formative assessments may be designed to simply add more items or tasks targeting a specific subskill, to sample a wider range of knowledge and skills in the subdomain, or to probe at a finer grain size a progression of skills that comprise performance. In the CBAL program, formative assessments are designed to serve each of these purposes.

The use of summative tests to generate formative hypotheses for teacher follow-up has obvious implications for test security and confidentiality. Presumably, those hypotheses will be most actionable if teachers have access to the item responses and tasks that generated the hypotheses—that is, examples of student work. However, to give teachers access is to reveal content that can no longer be reused. This disclosure has implications for the number of tasks in the TCA item pool and puts pressure on the testing program to continuously refresh the item pool.

Theory-of-action states as an intermediate outcome: Teachers and students can use tools and representations in instruction. We have previously described how CBAL uses tools and representations in designing the summative (and formative) assessments. One challenge is to guard against picking and choosing tools and representations that are too specific, and therefore cannot be used more generally in domain performance. For example, the five-paragraph essay, while perhaps a useful heuristic for introducing students to one basic organizational structure, can generate unintended consequences when used repeatedly in high-stakes assessment. The unintended consequence is students (and teachers) may focus too much attention on the lower level features of this structure without addressing deeper writing and thinking skills. The challenge for assessment design is to select a variety of general tools and representations that are legitimately part of the domain so that students learn to use them in various settings, adapting their thinking as necessary to effectively use those tools and representations in task performance.

Lesson 3: Decide how achievement is to be conceptualized and use that conceptualization in through-course assessment design. How TCA scores are aggregated depends on how one conceptualizes achievement, with different conceptualizations implying different designs and different approaches to aggregating TCA scores. The purpose here is not to consider the (many) technical complexities, but rather to focus on how decisions might influence the content and design of TCA. For example, if one wants to measure student *growth* across the TCA in a year, then there must be considerable overlap in what is measured each time to ensure something comparable to assess growth with. If one's primary goal is to document a student's *final status* at year end, then the culminating TCA might comprehensively cover the year's work, with each preceding TCA used simply to refine the estimate provided by that final measurement. Last, if one's goal is to measure *accomplishment*, the individual TCAs might each be constructed to measure different content and skills, probing those content and skills in some depth, with the summary score across TCA taking the form of a composite.

CBAL designs have thus far primarily explored an *accomplishment* conceptualization of achievement. In mathematics, key conceptual and developmental competencies have guided the design of each TCA (e.g., development of proportional reasoning; understanding of the concepts of variable and equality; and functions). In reading, broad text types (e.g., literary, informational, persuasive) are used to focus the scenario-based task sets in a TCA, but each TCA also includes a discrete task set to broaden coverage and potentially support longitudinal linking. In writing, each TCA targets a specific writing genre (e.g., persuasive writing, critical interpretation, appeal building).

Ongoing and Future Directions

One of the research foci of the current CBAL agenda is to understand the development of competency across time. A useful way to operationalize this understanding is to postulate developmental sequences—roughly, learning progressions (e.g., Heritage, 2008). While the CCSS emphasize increasing skill sophistication across grades, often it is not clear precisely how to interpret differences in standards across grades; and where the differences are clear, it is not always clear how these descriptive claims are empirically grounded. Progressions, by contrast, are often built around clearly defined qualitative shifts reflecting the emergence of new cognitive capacities; and these, in turn, can be related to empirical observations from the developmental literature.

In the CBAL mathematics strand, Harris and Bauer (2009) note that a rich research-based understanding of mathematical competency is not sufficient to connect summative assessment, formative assessment, and professional development in ways that can deeply support learning. They argue that it is also necessary to consider how competency develops. As such, the CBAL mathematics team has organized its assessments around developmental models that: define stages of competency through which students are proposed to progress from a cognitive perspective; are explicit about changes that occur as a consequence of learning; provide a basis for defining a meaningful scale of measurement; and offer a road map for supporting teaching and learning. In mathematics, there is a foundation of empirically based models of learning progression that the team draws upon.

In reading and writing, there is less agreement about empirically based learning progressions (Heritage, 2008). Some research is available, for example, with respect to the development of children's understanding of narrative (McKeough, 2007; Nicolopoulou & Bamberg, 1997; Nicolopoulou, Blum-Kulka, & Snow, 2002) and argumentative writing (Felton & Kuhn, 2001; Kuhn, 1999; Kuhn & Udell, 2003). In other cases, the research literature is quite sparse or inconclusive, and we have had to glean information from various sources, including curricula and standards, in order to propose progressions that make sense in terms of what is known about child development and the progression of standards, even if they cannot yet be validated directly. Thus, the developmental sequences embedded in the CBAL reading and writing model constitute hypotheses that we intend to verify and revise as research proceeds.

Conclusion

This paper reviewed some of the lessons learned from four years of work on CBAL, a research and development activity centered around creating a model for innovative K-12 assessment. Among the more general lessons learned from our experience is the importance of focusing on a small number of clearly articulated assessment purposes, since the purposes of an assessment, and their relative priority, has a major impact on its design. In CBAL, our primary purposes have been (a) to measure student achievement effectively and (b) to create assessments that also function as worthwhile educational experiences. We are attempting to fulfill that second purpose by grounding our assessment design in learning sciences research, as well as in content standards, relying heavily in design on such devices as scenario-based task sets, and on tools and representations modeling good teaching and scaffolding effective learning practice.

A second lesson learned was that a theory of action can be an indispensible tool in guiding the design of through-course assessments and in evaluating the extent to which validity and impact claims for the overall assessment system can be supported. The purposes that drive an assessment are constrained by the role that that assessment is supposed to play in the theory of action; thus, our experience suggests that significant thought should be given early on as to exactly what role the TCAs in a particular assessment system will play in the theory of action. This conclusion implies a prerequisite action very early in the test design process, namely, specifying the theory of action in enough detail to make it useable for assessment design and evaluation purposes.

Finally, we learned that much depends on how achievement is conceptualized. If achievement is viewed in terms of growth, through-course assessments must be designed to support measurement of change, which implies similar content across TCAs. On the other hand, if achievement is viewed in terms of accomplishment, the contents of specific TCAs may be strongly linked to curricular decisions and provide much less support for growth modeling (but may provide rather more coverage of the full construct).

All of these considerations imply that the design of through-course assessment is not a straightforward process, since we must think through how each design decision will play out across every each node in the theory of action.

References

Bennett, R. E. (in press). *CBAL: Results from piloting innovative K-12 assessments*. Princeton, NJ: Educational Testing Service.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice, 18*, 5–25.

Bennett, R. E. (2010). Cognitively Based Assessment of, for, and as Learning: A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives, 8*, 70–91.

Bennett, R. E., & Gitomer, D. H. (2009). Transforming K-12 assessment: Integrating accountability testing, formative assessment and professional support. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century* (pp. 43–61). New York, NY: Springer.

Bennett, R. E., Kane, M., & Bridgeman, B. (2011). *Theory of action and validity argument in the context of through-course summative assessment.* Princeton, NJ: Educational Testing Service.

Deane, P. (in press). NLP methods for supporting vocabulary analysis. In J. P. Sabatini & E. R. Albro (Eds.), *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences*. Lanham, MD: Rowman and Littleford Education.

Deane, P. (2010). *The skills underlying writing expertise: Implications for K-12 writing assessment*. Princeton, NJ: ETS.

Deane, P., Sabatini, J., & Fowles, M. (2011, February). *Rethinking K-12 writing assessment to support best instructional practices*. Paper presented at the Writing Research Across Borders II conference, Fairfax, VA.

Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skill. *Discourse Processes, 32*(2/3), 135–153.

Graf, A. E. (2009). *Defining mathematics competency in the service of cognitively based assessment for grades 6 through 8* (ETS Research Report No. RR-09-42). Princeton, NJ: ETS.

Harris, K., & Bauer, M. I. (2009, September). *Using assessment to infuse a rich mathematics disciplinary pedagogy into classrooms.* Paper presented at the 35th International Association for Educational Assessment (IAEA) Annual Conference, Brisbane, Australia.

Harris, K., Bauer, M. I., & Redman, M. (2008, September). *Cognitive based developmental models used as a link between formative and summative assessment.* Paper presented at the 34th International Association for Educational Assessment (IAEA) Annual Conference, Cambridge, England.

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment. Paper prepared for the Formative Assessment for Teachers and Students (FAST) State Collaborative on Assessment and Student Standards (SCASS) of the Council of Chief State School Officers (CCSSO).* Retrieved from the CCSSO website: http://www.ccsso.org/Documents/2008/Learning_Progressions_Supporting_2008.pdf

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28*(2), 16–46.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*(5), 1245–1260.

McKeough, A. (2007). Best narrative writing practices when teaching from a developmental framework. In S. Graham, C. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 50–73). New York, NY: Guilford.

Nicolopoulou, A., & Bamberg, M. G. W. (1997). Children and narratives: Toward an interpretive and sociocultural approach. In M. Bamberg, (Eds.), *Narrative development: Six approaches* (pp. 179–215). Mahwah, NJ: Lawrence Erlbaum Associates.

Nicolopoulou, A., Blum-Kulka, S., & Snow, C. E. (2002). Peer-group culture and narrative development. In S. Blum-Kulka & C. E. Snow (Eds), *Talking to adults: The contribution of multiparty discourse to language acquisition* (pp. 117–152). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Reilly, T., & Sheehan, K. M. (2009). *Cognitively Based Assessment of, for, and as Learning: A framework for assessing reading competency* (ETS Research Report. No. RR-09-26). Princeton, NJ: ETS.

Sheehan, K. M., & O'Reilly, T. (in press). The case for scenario-based assessments of reading competency. In J. P. Sabatini & E. R. Albro (Eds.), *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences.* Lanham, MD: Rowman and Littleford Education.

Appendix

Screen Shots of Mathematics, Reading, and Writing Assessments

Dams and Drought | Question # | Timer | CBAL MATH | Calc | STOP | Back | Next

Dams and Droughts

This picture shows a lake that had a water crisis in 2007.

When the lake is full, you can reach the water from the dock.

The lake has a dam at one end. Water flowing from the lake past the dam is used to create electricity and provide water for crops. The lake may become so shallow that there will not be enough water to generate electricity.

This activity will help you figure out whether water conservation actions need to be taken.

Dams and Drought | Question # | Timer | CBAL MATH | Sink | Calc | STOP | Back | Next

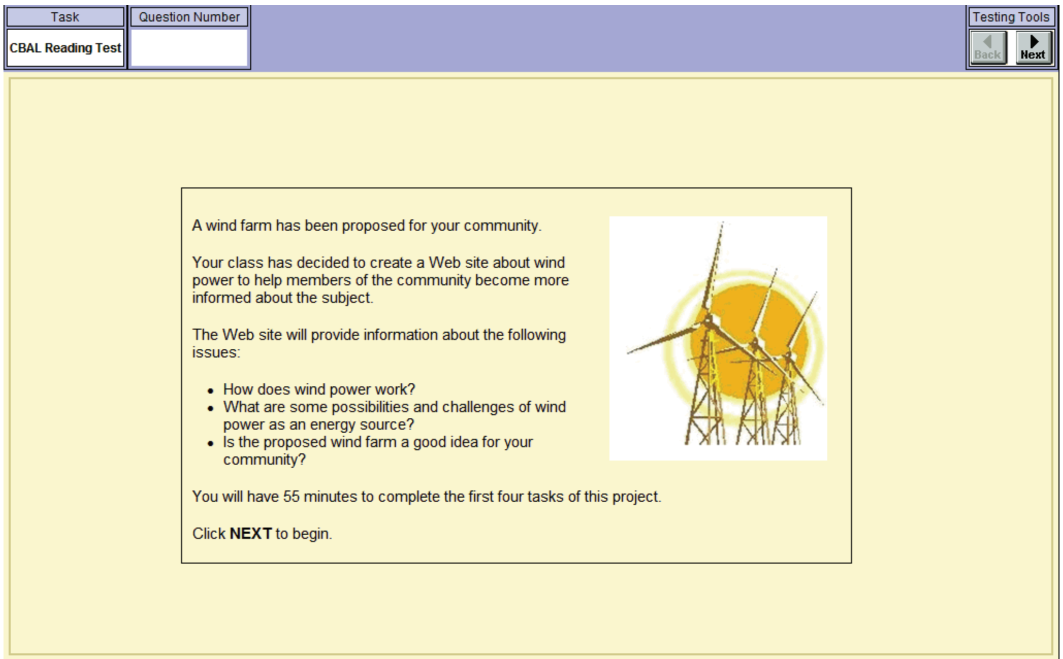
How the Sink Tool Works

- Water flows from the faucet into the sink.
- You can set the rate at which the water flows into the sink.
- You can set whether or not the water flows out of the sink by choosing to have the plug in or out.
- The full sink holds 18 gallons of water.

The rate at which the water flows in or out under certain conditions is shown below:

Water Flows In			Water Flows Out
Faucet is turned on all the way	Faucet is turned on halfway	Faucet is turned on one-quarter of the way	Plug out
3 gallons per minute	1.5 gallons per minute	0.75 gallons per minute	1.5 gallons per minute

Practice using the Sink Tool by clicking on the button above: Sink



Task: CBAL Reading Test | Question Number | Testing Tools | Back | Next

A wind farm has been proposed for your community.

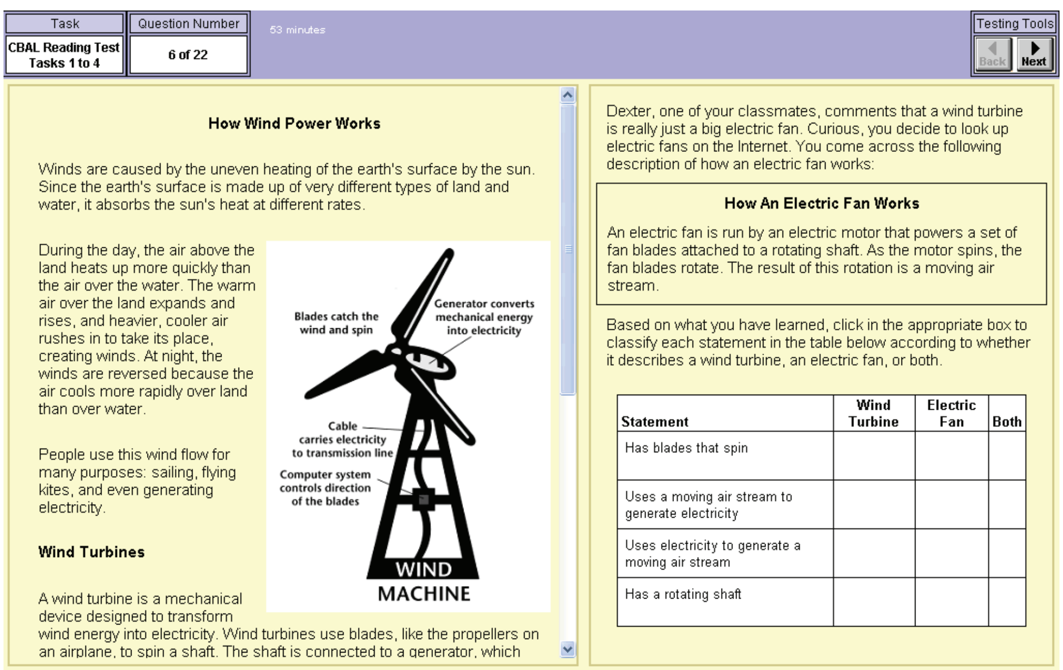
Your class has decided to create a Web site about wind power to help members of the community become more informed about the subject.

The Web site will provide information about the following issues:

- How does wind power work?
- What are some possibilities and challenges of wind power as an energy source?
- Is the proposed wind farm a good idea for your community?

You will have 55 minutes to complete the first four tasks of this project.

Click **NEXT** to begin.



Task: CBAL Reading Test Tasks 1 to 4 | Question Number: 6 of 22 | 53 minutes | Testing Tools | Back | Next

How Wind Power Works

Winds are caused by the uneven heating of the earth's surface by the sun. Since the earth's surface is made up of very different types of land and water, it absorbs the sun's heat at different rates.

During the day, the air above the land heats up more quickly than the air over the water. The warm air over the land expands and rises, and heavier, cooler air rushes in to take its place, creating winds. At night, the winds are reversed because the air cools more rapidly over land than over water.

People use this wind flow for many purposes: sailing, flying kites, and even generating electricity.

Wind Turbines

A wind turbine is a mechanical device designed to transform wind energy into electricity. Wind turbines use blades, like the propellers on an airplane, to spin a shaft. The shaft is connected to a generator, which

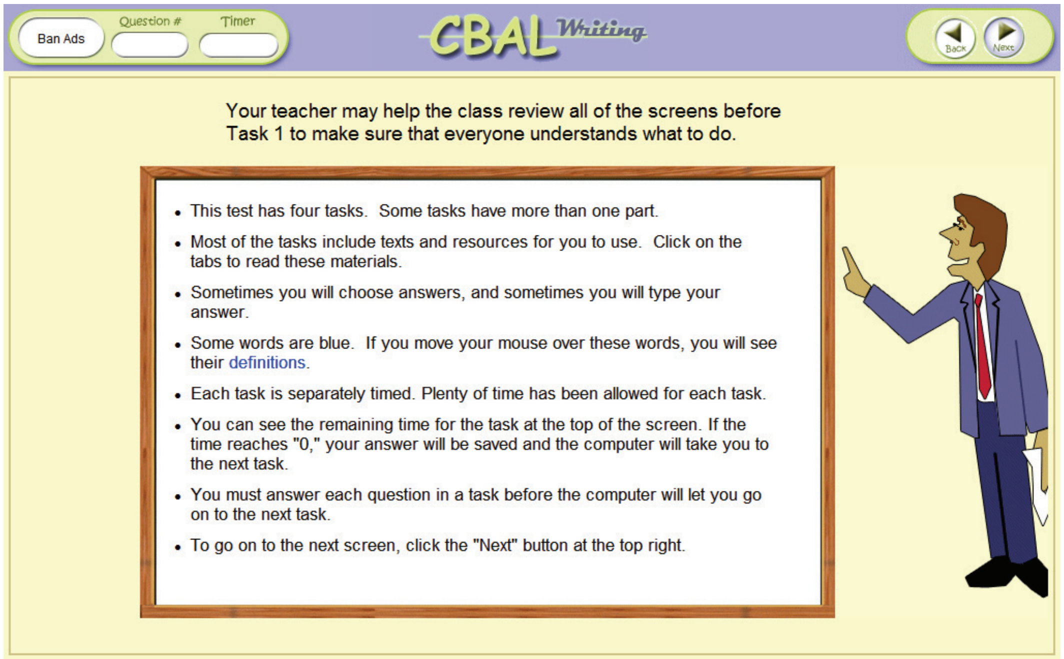
Dexter, one of your classmates, comments that a wind turbine is really just a big electric fan. Curious, you decide to look up electric fans on the Internet. You come across the following description of how an electric fan works:

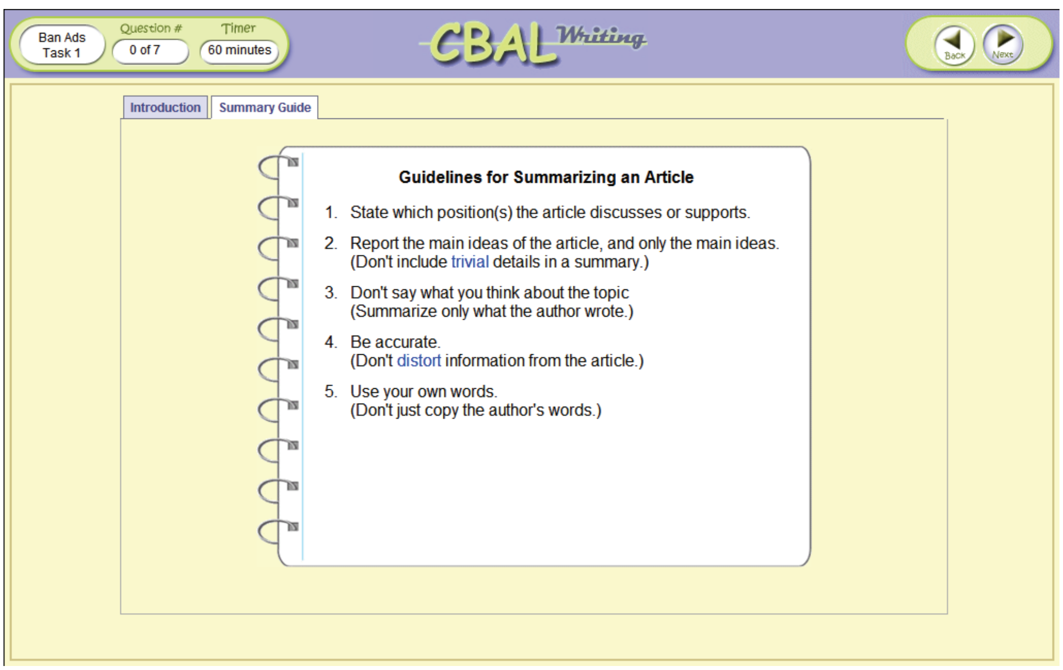
How An Electric Fan Works

An electric fan is run by an electric motor that powers a set of fan blades attached to a rotating shaft. As the motor spins, the fan blades rotate. The result of this rotation is a moving air stream.

Based on what you have learned, click in the appropriate box to classify each statement in the table below according to whether it describes a wind turbine, an electric fan, or both.

Statement	Wind Turbine	Electric Fan	Both
Has blades that spin			
Uses a moving air stream to generate electricity			
Uses electricity to generate a moving air stream			
Has a rotating shaft			


Ban Ads
Question #
Timer
CBAL Writing
Back
Next
Your teacher may help the class review all of the screens before Task 1 to make sure that everyone understands what to do.
• This test has four tasks. Some tasks have more than one part.
• Most of the tasks include texts and resources for you to use. Click on the tabs to read these materials.
• Sometimes you will choose answers, and sometimes you will type your answer.
• Some words are blue. If you move your mouse over these words, you will see their definitions.
• Each task is separately timed. Plenty of time has been allowed for each task.
• You can see the remaining time for the task at the top of the screen. If the time reaches "0," your answer will be saved and the computer will take you to the next task.
• You must answer each question in a task before the computer will let you go on to the next task.
• To go on to the next screen, click the "Next" button at the top right.


Ban Ads Task 1
Question #
0 of 7
Timer
60 minutes
CBAL Writing
Back
Next
Introduction
Summary Guide
Guidelines for Summarizing an Article
1. State which position(s) the article discusses or supports.
2. Report the main ideas of the article, and only the main ideas. (Don't include trivial details in a summary.)
3. Don't say what you think about the topic (Summarize only what the author wrote.)
4. Be accurate. (Don't distort information from the article.)
5. Use your own words. (Don't just copy the author's words.)