DOCUMENT RESUME

ED 445 008 TM 031 545

AUTHOR Hickey, Daniel T.; Holbrook, Jennifer
TITLE PALS-Supported Performance Assessments in the Learning by Design Project.
PUB DATE 2000-04-00
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Item Banks; Middle School Students; Middle Schools; *Performance Based Assessment; Problem Solving; *Student Evaluation; *Test Construction; Testing Programs

ABSTRACT

This paper describes an assessment and evaluation effort occurring in the context of a middle school science curriculum development effort known as Learning by Design (LBD). The LBD approach builds on a body of cognitive science and professional education research that emphasizes learning from problem-solving experience. The assessment and evaluation effort uses assessment instruments obtained from the Performance Assessment Links in Science (PALS) Web site, as well as some more conventional multiple choice items, some also obtained from the World Wide Web. Following a brief description of the LBD learning environment, the paper provides an overview of the specific tools being used and a description of the refinements and extensions that the study is making. Results from the work in progress are presented to help illustrate the approach. In the 1998-1999 school year, 179 students in 12 classes taught by 4 LBD teachers and 51 students in 2 comparison classes completed a multiple choice content test before and after physical science instruction. In the current year, a revised content test was being completed before and after instruction in eight LBD and three comparison physical science classrooms. It is expected that this approach will provide valid information about the degree to which inquiry-oriented rituals and learning-oriented participatory structures are established in the different classrooms. These studies will help design the LBD environment and aid in the modification of PALS items for the assessment program. (Contains 26 references.) (SLD)

ED 445 008

PALS-Supported Performance Assessments in the Learning by Design Project

Daniel T. Hickey
Jennifer Holbrook

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in New Orleans
April 2000

PALS-Supported Performance Assessments in the Learning by Design Project

Daniel T. Hickey
Georgia State University

Jennifer Holbrook
Georgia Institute of Technology

This paper describes an assessment & evaluation effort occurring in the context of a middle school science curriculum development effort known as Learning by Design (LBD, Kolodner, Crismond, Gray, Holbrook, & Puntembakar, 1998). The assessment & evaluation effort employs assessment instruments obtained from the Performance Assessment Links in Science (PALS) website, as well as more conventional multiple choice items, some also obtained from the WWW. These assessment tools are being implemented within a approach to program evaluation that is (1) akin to the pragmatic model advanced by Pogrow (e.g., 1998, 1999), (2) is organized around the three views of knowing and learning described by Greeno, Collins, & Resnick (1996), and (3) uses the competitive approach described by Greeno, & Moore (1993) to reconciling the differences between these three views. Reflecting contemporary views on assessment practices, the performance assessments are being extended to provide the sort of formative feedback called for by leading theorists (e.g., Bransford, Brown, and Cocking, 1999; Torrance & Pryor, 1998). This extension presents the challenge of balancing concerns regarding evidential validity with the desire to use our assessment practices to directly enhance learning—what Fredericksen and Collins (1989) called *systemic* validity. Our efforts to address this challenge follow from the example presented in Hickey, Wolfe, & Kindfield (2000). Following a brief description of the LBD learning environment, we provide an overview of the specific tools we are using and description of the refinements and extensions we are making, presented in the context of our assessment & evaluation framework. Results from this work-in-progress are presented to help illustrate our approach.

The *Learning by Design Environment*

The Learning by Design curriculum is being developed by a team at Georgia Tech, under the direction of Janet Kolodner, with the support of the National Science Foundation[1]. The LBD approach builds on a body of cognitive science and professional education research that emphasizes learning from problem-solving experience. This includes case-based reasoning (Kolodner, 1993), problem-based learning (e.g., Barrows, 1985), and analogical reasoning (Holyoak and Thagard, 1997). In the context of extended design problems that feature various activities structured to facilitate meaningful collaboration, students grapple with and learn the content knowledge and problem solving skills associated with the domains of physics and earth science as well as more general scientific inquiry skills in the context of those domains. Six LBD curricular units have been or will be completed. This includes one "Launcher" unit that introduces students to the LBD approach and three physical science units that are largely completed, and two earth sciences units that are still under development. [2]

[1] Materials Development Program Grant ESI-9818828

[2] For more information about Learning by Design, visit http://www.cc.gatech.edu/edutech/projects/lbdview.html

A Pragmatic Evaluation Framework

The LBD team has been developing materials and implementing them in Atlanta-area middle school schools since 1996. In 1999, a research team at Georgia State University began working with LBD project researchers at Georgia Tech to assess learning in LBD and comparison classrooms and provide program evaluation information. During the 1998-1999 school year, learning outcomes were documented for the Physical Sciences curriculum in the classrooms of five LBD teachers and two comparison teachers. During the present school year (1999-2000), learning outcomes are being documented in the classrooms of 8 LBD and 2 comparison Physical Sciences classrooms and 6 LBD Earth Sciences classrooms.

Three aspects of the framework we are using to evaluate LBD are particularly noteworthy. First, this framework embraces Pogrow's (e.g., 1998, 1999) pragmatic view of evaluation research that emphasizes outcomes across different implementations of the same innovation (e.g, comparing gains in *strong, weak*, and *failed* implementations). We concur with Pogrow that such within-group comparisons yield the most valid evidence of a program's effectiveness, and that the information derived from such designs is essential for program improvement. Furthermore we share Pogrow's concerns about traditional quasi-experimental evaluation designs, because such designs often compare only the strongest implementation of an innovation to some comparison treatment of unascertained relative quality--as was apparent in the various evaluations of Slavin's (e.g., 1999) *Success for All* direct instruction programs (see also Venezky, in press). To this end, the research team is devoting substantial attention to observing and documenting differences in the nature and relative quality of the learning environment in the various LBD and comparison classrooms. Only in light of this information are we interpreting gains for individual teachers and comparing pairs of LBD/comparison teachers. Reflecting the emerging "design experiment" model (Brown, 1992; Collins, 1999), our approach emphasizes the need for effective models of practice while de-emphasizing traditional goals of theoretical coherence and parsimony.

The second noteworthy aspect of our evaluation framework is its deliberate incorporation of the comparative model of knowing and learning best exemplified in handbook chapters by Greeno, Collins, & Resnick (1996) and Case (1996). These authors draw on a range of educational research to show that what means to "know" something like the laws of motion or be "able to" engage in scientific inquiry depends entirely ones assumptions about knowledge itself. One's assumptions about knowledge, in turn, support additional assumptions about the nature of learning. More to the point of the present paper, one's assumptions about knowing and learning have direct implications for assessing what students know and evaluating the degree to which given learning environments have contributed to that knowledge. Like Greeno, et. al. and Case, our framework distinguishes between three perspectives that follow generally from the theories associated with Skinner, Piaget, and Vygotsky. We chose to label these perspectives *empiricist, rationalist,* and *sociohistoric*. Following is a description of how we are using this distinction to help organize our efforts, establish the utility of the different assessment tools we used, and interpret those results.[3]

Knowing as having associations. While most clearly associated with behaviorism, empiricist views of knowing and learning continue to be influential in many sectors. This includes of cognitive psychology (e.g., Anderson, Reder, & Simon, 1996) and instructional design (e.g., Gagne,1985). Empiricist views are consistent with the "folk psychology" models

[3] For more information about these three perspectives and various WWW resources consistent with each, visit KLTI (Knowing, Learning, & Teaching on the Internet) at http://education.gsu.edu/epedth/KLTI

that are implicit in the arguments of many parents, policy makers, and media commentators. From this perspective, knowledge is construed as a repertoire of patterns that the mind has learned to detect--behavioral or cognitive associations that represent fragments of an objective external reality-- and operations that can be executed on those patterns. This perspective is inherently *reductionist* (assuming that complex behavior or concepts consist of smaller elements) and *additive* (assuming these smaller elements readily assemble into an accurate representation of the more complex entity). From this perspective, "knowledgeable" activity is the result of "bottom up" processing of lower-level components of knowledge and skill.

When knowledge is construed in terms of many small associations, learning is seen as the process of forming, strengthening, and adjusting those associations, and those associations are presumed to transfer readily from the learning environment to subsequent transfer environments where they might be needed. By demonstrating that students have made associations that are arguably useful in some transfer environment, the students are presumed to have learned transferable knowledge. Thus, assessment of an individual's knowledge and evaluation of whether or not transferable learning occurred in a particular environment is quite straightforward. The traditional multiple-choice test works quite well for this purpose. Because this perspective construes a domain of knowledge as a stable, objectifiable body of associations, it is possible to specify a large number of test items from which a representative sample can be drawn. With the addition of powerful psychometric techniques such as item response theory, these assumptions support the use of standardized tests as is commonplace in many aspects of education.

One can certainly assess knowledge of physics and evaluate learning environments like LBD using standardized content tests. We did just that, by constructing a multiple choice test consisting of items taken from a variety of sources, including the released NAEP items and items from the TIMMS study.[4] Of course, evaluating an innovation such as LBD in this fashion is problematic, because the environment was designed with an entirely different set of assumptions about learning in mind. Empiricist assumptions about learning suggest that the ideal environment keeps learners engaged in the routines needed to build and strengthen the relevant associations from part-to-whole, as epitomized in conventional drill and practice and direct instruction environments. These settings make it possible to identify the entire range of associations that students might be tested on, and ensure that students have mastered as many of those associations as possible through repeated practice and testing.

While inconsistent with the assumptions behind LBD, empiricist assessment approaches offer important advantages and we intend to continue using them. Multiple choice tests are very simple to implement and yield reliable scores about individual proficiency and learning; by examining specific groups of items one can make strong inferences. When coupled with sophisticated psychometric methods such as Rasch scaling, these tests can provide remarkably precise information about how individual students compare to other students. Furthermore, such tests may yield the kind of evidence that some observers and stakeholders are looking for. In this era of increased accountability on standardized measures, including such measures in ones evaluation plans make it possible to "fine tune" implementations to ensure whatever degree of coverage of such content is considered appropriate.

[4] For information about and items from the Third International Math and Science Study, visit http://www.timss.bc.edu

In the 1998-1999 school year, 179 students in 4 LBD teachers' classrooms and 51 students in 2 comparison teachers' classrooms completed a 30-item multiple choice content test before and after physical sciences instruction. Rasch scaling of posttest scores was then used to provide an index of each student's relative proficiency and each item's relative difficulty. For the two validly matched LBD/comparison teachers (both highly regarded science teachers who taught gifted students in advantaged suburban schools), a significantly larger gain on scaled scores was found for the LBD students, $F(1,121) = 13.42$, $p < .001$. However, a potential ceiling effect may have contributed to these outcomes. Particularly promising was the large gains found the LBD teacher who taught relatively disadvantaged students. While these students' posttest scores were still below the pretest scores of the more advantaged students, these findings and the findings of the other LBD classrooms show that the curriculum was effective at enhancing students' knowledge of relatively discrete physical science concepts. In the current year a revised content test (with more difficult items) is being completed before and after instruction in the 8 LBD and 3 comparison Physical Science classrooms. A separate Earth Science test is also being completed instruction in 6 LBD earth sciences classrooms.

Knowing as general concepts and abilities. Rationalist perspectives emerged as a major focus of psychological research on learning in the 1970s. Within cognitive science, the rationalist perspectives was most evident in the emergence of schema theories (e.g., Shank & Abelson, 1977) and the associated case-based reasoning models (e.g., Kolodner, 1993). Perhaps most readily understood as the antithesis of empiricism, rationalist perspectives view the mind as a uniquely human organ whose innate function is making sense of information in the environment. Knowledge is viewed in terms of structures of information and processes that the mind constructs in order recognize and make sense of (i.e., "rationalize") symbols in order to understand concepts and exhibit general abilities. "Knowledgeable" activity is construed as "top down" because the individual is presumed to be marshalling the various higher-level schema needed to construct a solution for the problem represented by any particular task. Learning is viewed as a natural outcome of intrinsic sense-making processes (as in Piaget's assimilation and accommodation).

When learning is viewed in terms of conceptual schema constructed by each individual (rather than objective lower-level associations), transfer is analyzed in terms of those same structures. From this perspective, assessments should examine whether students can employ schema presumably constructed solving problems in the learning environment to solve new problems in different contexts. This focus on the transfer of higher-level reasoning structures is complicated, and is at the heart of the tension between empiricist and rationalist approaches to assessment and evaluation. In the extreme view, rationalism's top-down view of proficiency assumes that the lower-level associations presumed by empiricists to drive bottom-up proficiency don't really exist—so that when students are solving simple multiple-choice items they are still using higher-level schema to make sense of that particular feature of the environment.

As described above, the instantiation of rationalist perspectives in cognitive science provides much of the basis for the LBD curriculum. As detailed by Greeno, Collins, & Resnick (1996) students who are expected to construct an understanding of a domain must be given opportunities to interact with material aspects of the domain and be presented with problems and activities that engage their relevant interests, initial understanding, and general problem solving ability. From this perspective, the ideal learning guides learners towards a more complete understanding, paying particular attention to the sequences of conceptual development and generality of the concepts that students develop. Because these are essentially the same principles used to design the LBD curricula, we expect that assessment practices that follow from rationalist perspectives should provide the most valid evidence of LBD's effectiveness.

Our efforts to assess student's development of transferable conceptual understanding in the LBD environment has employed performance assessments selected from the PALS (Performance Assessment Links in Science, see Quellmalz, Schank, Hinojosa, & Padilla, 1999) item bank maintained by SRI.[5] Assembled with the support of the National Science Foundation, the WWW site makes available various assessments gathered from prior large-scale performance assessment efforts. In general, these items feature multi-step investigations around authentic scientific problems, generally involving some sort of hands-on activity and written description. Because both PALS performance assessments and the LBD curriculum are referenced to the National Research Council (1996) science standards, we were able to readily identify the set of middle-school items that targeted the physical science content standards we were targeting in LBD. We then selected a "near transfer" item *(Speeding in a School Zone)* that required students to reason about Newton's laws in a similar context as the LBD curriculum (e.g., the velocity of a toy car), and a "far-transfer" item (*Where the Rubber Meets the Road*) that presented the concept in relatively different context (the force needed to overcome sliding friction for different kinds of rubber under different surface conditions).

During the 1998-1999 school year, these two assessments were administered to students in the LBD physical science classrooms and to similar students in more conventional classrooms following physics instruction. While certainly more resource-intensive than the multiple choice test described above, the PALS items proved relatively easy to implement and score. The results showed that some LBD classrooms outperformed comparison classrooms. However, we lacked pretest information and sufficient information about the curriculum in the comparison classrooms to fully interpret these results. In the current implementation, students in the 8 LBD and 3 comparison physical sciences classrooms completed *Sand in the Bottle* before instruction and are completing *Speeding in a School Zone* and *Where the Rubber Meets the Road* after instruction. We will also administer all three of the assessments during the same time period to another sample of middle school students who are not otherwise participating in the study. These scores will be scaled to provide the item difficulty indices needed to fully interpret the gains shown by the other students from pretest to posttest.

Knowing as participation. The relatively newer sociohistoric perspective views knowledge as a cultural entity that is distributed across the physical and social environment in which that knowledge is developed and used. Thus, an individual's knowledge of a domain is distributed, or "stretched across" the people, books, computers, classrooms, worksheets, etc. present in the context where presumably knowledgeable participation can occur. From this perspective, knowledge is represented in the regularities of successful activity in particular context, and knowledgeable activity is presumed to indicate that the individual has become familiar with (i.e., "attuned" to) the constraints (that bound participation) and affordances (that scaffold participation) of the environment in which successful activity occurs. In other words, knowledge is represented in the individual's ability to use all of the tools available to overcome the limits of mind in order to maximize successful participation. Thus, learning is presumed to take place within the construction of socially defined knowledge and values, and occurs as the individual co-constructs understanding of that domain in whatever context it is encountered.

[5] The PALS website is located at http:// www.ctl.sri.com/pals/

Through this participation, individuals strengthens their ability to further participate in this activity. An analysis of transfer, then, must consider the constraints and affordances that support activity in the learning environment and in the transfer environment then consider "transformations" that relate to a given pair of learning and transfer situations. For transfer to occur, some constraints and affordances must be the same (be "invariant") across both situations, and the learner must learn (become "attuned" to) these *invariants* in the initial learning situation (Greeno, Smith, & Moore, 1993).

Greeno, Collins, & Resnick (1996) outline four principles for designing environments that follow from sociohistoric perspectives, including (1) fostering participation in social practices of inquiry and learning, (2) providing support for positive epistemic identity, (3) developing disciplinary practices of discourse and representation, and (4) providing practice in formulating and solving realistic problems.. While the theoretical core of the LBD curriculum follows most clearly from rationalist perspectives, the broader context in which that curriculum is embedded was designed to be generally consistent with sociohistoric perspectives. Indeed, as described by Kolodner, Crismond, Fasse, Gray, Holbrook, & Puntembakar (in preparation), the primary changes to the project involve new or newly emphasized activities that are designed to enhance inquiry-oriented participatory rituals in the LBD classrooms. In a modest departure from the somewhat mechanistic language of the schema theoretic approach, the LBD framework now characterizes the many collaborative features of the learning environment as "LBD rituals" Nearly every activity is carried out in a collaborative fashion, and many are designed to ritualize the social aspects of authentic inquiry. The most public of these rituals are the "pinup" and "gallery walk" that will look and feel remarkably familiar to anyone who as attended a poster session at an academic conference.

Given these aspects of the LBD curriculum, we have every reason to believe that the rituals and participatory structures in the LBD classroom should be more organized around learning and inquiry, relative to students in classrooms featuring more conventional science instruction. Evaluation of learning when learning is characterized as enhance participatory rituals can be best understood within Vygotsky's notion of the zone of proximal development, or ZPD. As individuals become attuned to the constraints and affordances of an environment, the upper bound of the competency where they can participate successfully in that environment (and therefore other transfer environments that share invariant aspects) increases. According to Greeno, Collins, & Resnick (1996), applying this perspective to assessment calls for (1) observing individuals participate in the development and use of knowledge, (2) allowing students to participation in designing the assessment system, and (3) taking account the effects of the assessment practice on the larger educational system. Sociohistoric instructional principles are at the heart of now-familiar portfolio assessment practices and other assessment-oriented educational reforms, and provide the motivation for several aspects of the LBD effort. This include ongoing ethnographic observations that are documenting inquiry oriented rituals, as well as the *LBD Student Success Handbook* (Gray, Groves, & Kolodner, 2000). This guidebook provides LBD students with some of tools for self-assessment and reflection, including clear performance criteria, examples of quality work, etc.

A major component of the effort describe here is gathering relatively more systematic evidence of LBD's influence on inquiry-oriented social structures and participatory. One of the PALS performance assessment described above (*Where the Rubber Meets the Road*) was actually completed by students as a collaborative activity. Groups of 3-5 are videotaped while they construct the solution to the assessment problem. These tapes are then scored according to dimensions of collaboration derived from Pomplum (1996). Specifically, each of the three phases of the task (setup, experiment, writeup) is scored on a 1-5 scale on each of the 10 dimensions shown on Table 1. We presume the kind of participatory practices required to get a high score on this activity are arguably similar to rituals of practice and patterns of participation that are ostensibly established during the LBD activities. In other words, many of the constraints and affordances of the LBD environment that students must become attuned to in order to participate successfully are invariant across both the LBD environment and this aspect of the assessment environment. Because these participatory practices are arguable desirable in any science learning environment, we believe that higher collaboration scores are valid evidence of LBD's effectiveness. Additionally, because the effectiveness of this collaboration relates to the quality of the groups responses to the actual problem, we also consider the assessment score as additional evidence about the participatory practices that have been ritualized in each individual classroom.

During the 1998-1999 school year, we pilot tested this method with just the one assessment, in 12 classrooms taught by the in 4 LBD teachers and 2 classrooms taught by 2 comparison teachers. While scoring the resulting 47 tapes was a relatively laborious process, we were able to reach acceptable levels of reliability and assemble an apparently valid "index" of the participatory practices in the various classrooms. We found apparent differences between teachers; encouragingly, the scores for different classrooms taught by the same teacher were relatively smaller than the differences between teachers. We also found the expected association between participatory quality and individual group scores, with the higher-scoring groups within classrooms, and the higher scoring classrooms overall showing higher quality collaboration. However, the one LBD teacher with a valid comparison teacher did not record noticeably higher scores on either participation or performance. Given the exploratory nature of this pilot, we have yet to gather sufficient relative information about the curriculum in both the LBD and comparison classrooms to interpret the differences more completely. In the current implementation cycle, we are again administering *Where the Rubber Meets the Road* following instruction in both LBD and comparison Physical Science classrooms, and hope to pilot an additional assessment in the Earth Science classrooms in the same manner. In the event we are able to streamline the rather laborious process, we may implement a similar pretest assessment as well.

We are confident that this method will provide valid evidence of the degree to which inquiry-oriented rituals and learning-oriented participatory structures are established in the different classrooms. However, we acknowledge that the constraints imposed by the need to gather large scale comparative data prevents the method from being wholly consistent with a sociohistoric view. Another essential component of our framework is extensive ethnographic field observations and interpretive ethnographic analyses of videotaped class sessions. We believe that these methods will provide the most valid evidence regarding the rituals and participatory practices that emerge in both the LBD and comparison classrooms.

This ethnographic study of LBD's influence on the classroom culture supports the third (and most esoteric) noteworthy aspect of our assessment and evaluation framework. We anticipate conflicting interpretations of LBD's effectiveness across content-test, individual performance assessments, and group performance assessments. As described by Greeno, Smith, and Moore (1993), there are two ways to reconcile the tension between the three views of knowing and learning. One approach that is implicit in many efforts to use individual-oriented empiricist and/or rationalist models to understand broader sociocultural activity can be described as the *levels of aggregation approach.* In this approach, sociocultural activity is understood as merely an aggregation of behavioral contingencies (readily explained within an empiricist framework) and/or aggregated patterns of cognitive information processing (readily explained within an rationalist framework). However, as argued by Hickey (1999, 2000), this approach is problematic for several reasons, including its failure to reconcile the tension between empiricist and rationalist perspectives (for example, in the ongoing debate of the effects of "extrinsic" rewards on "intrinsic" motivation). An alternative approach to reconciliation embraces a Hegelian cycle of thesis-antithesis-synthesis. This so-called *competitive* approach characterizes the rationalist perspective as an antithesis to the thesis of empiricism, and characterizes sociohistoricism as a higher-order synthetic perspective which can be used to reconcile the strengths and weakness of the two prior approaches. Whereas the first approach views sociocultural activity as aggregations of behavioral contingencies and/or cognitive information processes, the second approach represents patterns of both behavior and information processing as just special cases of a broader forms of human activity that is most readily explained within a sociohistoric framework.

When applied to our assessment and evaluation effort, this competitive approach implies that the detailed understanding of the sociocultural activity in the classroom gained by ethnographic observations will be used to interpret the entire set of results. No presumption is made regarding the relative value of different kinds of learning outcomes we are documenting. Rather, performance on each assessment is construed within a contextualist, situative perspective that views all activity in terms of both the individual and the group becoming attuned to the constraints and affordances on participation in that particular activity. Thus, rather than viewing scores on the content test as evidence "knowledge" of the domain, these scores are indicative of the degree to which the learning environment prepared members of that classroom community to participate in the activity of individually recognizing the correct response to discrete questions in a highly structured context. We believe that this aspect of our framework will be invaluable for interpreting the entire body of results, and for helping us refine the LBD environment as needed reconcile the conflict between the calls for broader inquiry and participatory skills represented in the science education standards, and the pressure for students to register higher scores on standardized content tests.

Extension and Refinements of the PALS Performance Assessments

The PALS items come directly from the website in remarkably consistent and clear form. However, not surprisingly, constraints on the project and the desire to directly enhance student learning have led us to modify the tasks and our use of them substantially.

New items. One of the changes involves inserting new items on the tasks. For example, because LBD aspires particularly to help students learn how to design experiments, we have added additional items at the start of both the *Speeding in a School Zone* and *Where the Rubber Meets the Road* that ask students to actually design an experiment. We are just now beginning to score these items and it remains to be seen whether they can be scored reliably, how they compare to the other items, and how they reflect on LBD. Because such items target the aspects of domain reasoning covered in LBD classrooms only, it will be necessary to examine scores on these items separately from the other items. We also have yet to ascertain the whether or not these items interact with other items on the assessments, and whether the resulting longer items can be completed within a standard class period. Resolution of these questions is a major task facing the research team once the present round of assessment and evaluation is completed.

Formative feedback. The third component of a sociohistoric assessment practices according to Greeno, Collins, & Resnick (1996) is an accounting for assessment's effects on the larger educational system. Reflecting the contextualist orientation of sociohistoric perspectives, assessment practices are increasingly viewed as a salient feature of the learning environment, rather than as a isolated summative activity. As such the value of assessment practices for directly supporting students learning is a central theme of contemporary sociohistoric views of learning and instruction. For example, the guidelines for implementing the practices suggested in the recent NRC report entitled *How People Learn* (Bransford, Brown, & Cocking, 1999) suggest that:

> Formative assessments–ongoing assessments designed to make students' thinking visible to both teachers and students–are essential. They permit the teacher to grasp the students' preconceptions, understand where the students are in the "developmental corridor" from informal to formal thinking, and design instruction accordingly. In the assessment-centered classroom environment, formative assessments help both teachers and students monitor progress. (Donovan, Bransford, & Pellegrino, 1999, p. 21).

In a groundbreaking article, Frederiksen and Collins (1989) emphasized the consequences of assessment practices by introducing the notion of *systemic validity:*

> A systemically valid test is one that induces in the educational system curricular and instructional changes that foster the development of the cognitive skills that the test is designed to measure. Evidence for systemic validity would be an improvement in those skills after the test has been in place within the educational system for a period of time (p 27).

Frederiksen and Collins propose a set of principles for the design of systemically valid assessment systems, including the *components* of the system (a representative set of tasks, a definition of the primary traits for each subprocess, a library of exemplars, and a training system for scoring tests), *standards* for judging the assessments (directness, scope, reliability, and transparency) and *methods for fostering self-improvement* (practice in self-assessment, repeated testing, performance feedback, and multiple levels of success).

When referenced to established science education standards and embedded in a learning environment such as LBD, the PALS assessment tasks provide most of what is needed to create a systemically valid assessment system. Starting with the present round of posttest, we are providing each student a copy of his or her completed assessment, along with a revised version of the scoring rubric and guidelines for the teacher on how to use these materials to maximize learning. It seems to us that having teachers guide students through the process of scoring their own completed assessment will maximize opportunity to learn the targeted concept. Because consequences were attached to student activity on the activity in the first place, it is likely that students were motivated to do the best they could on the activity. As was documented in research reported by Hickey, Wolfe, & Kindfield (2000), these motivational factors along with the shared experience on a specific well-designed task allow students and teachers to participate in a much more sophisticated discussion of domain reasoning than seems otherwise possible. An additional valuable outcome of the formative assessment practices is that teachers and administrators are much more willing to set aside class time for assessment. In the current climate of increased accountability on standardized measures, teachers are reluctant to commit many class periods to purely summative assessment activities—when such activities are not seen as enhancing performance on the mandated measures used to judge overall teaching effectiveness. In our experience this reluctance turns to enthusiasm when the assessment activities are seen as a powerful way to ensure that all students understand the specific key concepts. .

Formative assessment feedback raises important issues regarding the validity of evidence collected within assessment practices. If students in the LBD classrooms are given formative feedback that is not provided to comparison students, then any evidence from subsequent similar assessments is confounded. Our efforts to balance concerns between *evidential validity* and *consequential validity* draws on the methods reported by Hickey, Wolfe, & Kindfield (2000). This includes focusing on specific reasoning targets and carefully sequencing of assessments and instruction. Because we are not currently assessing students again following formative feedback, our summative scores are not confounded. However, in the next years' implementation we expect to extend the scope of the implementation to include multiple rounds of assessment. By selectively incorporating formative feedback in some classrooms but not others, we also expect to be able to judge the systemic validity of our practices (by comparing final outcomes in LBD classrooms with and without formative feedback) while still maintained the necessary degree of evidential validity (by only comparing learning in the LBD classrooms without formative feedback to the comparison classrooms).

Formatting changes. The final types changes and revision we have made concern the formatting of the assessments themselves. In addition to the inclusion of the new items described above, our desire to provide formative feedback has required substantial revision. In particular, the need to quickly and efficiently duplicate large numbers of completed assessments has required us to find ways to avoid stapling sheets together and the like. While these are seemingly mundane aspects of the challenge we face, they have turned out to be essential to both gathering the performance data and providing feedback. This is one of the many areas in which we see enormous value in the community of practice that is developing around the PALS websites. The discussion list maintained at the site seems like an ideal vehicle for users to distributed new elements and extensions (such as our revised tasks and formative feedback materials). We look forward to participating in such a worthwhile activity.

Authors Note

Daniel T. Hickey, Department of Educational Psychology and Special Education, Georgia State University, Atlanta, GA 30303, dhickey@gsu.edu; Jennifer Holbrook, EduTech Institute, Georgia Institute of Technology, Atlanta, GA, holbrook@cc.gatech.edu.

We acknowledge the contributions of many other individuals to this ongoing assessment and evaluation effort. This includes Janet Kolodner, Barbara Fasse, Jackie Gray, Paul Camp, Sarah Hyser, David Crismond at GA Tech, and Joy Mordica, Brooke Brimm, Laura Vandewiele, Rodia Vance, Art Russell, Enxin Deng, and Bing Wang at GSU. We also gratefully acknowledge the participation and patience of the many teachers and students in the LBD and comparison classrooms. Regrettably, we are not allowed to acknowledge them by name.

References

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25* (4), 5-11.

Barrows, H. S. (1985). *How to design a problem-based curriculum for the preclinical years*. NY: Springer.

Bransford, J. B., Brown, A. L. Cocking R. (1999) (Eds.). *How people learn: Brain, mind, experience, and school.* Committee on Learning Research and Educational Practice. Washington, DC: National Academy Press. (Full text online at www.nap.org)

Brown, A. L. (1992). Design experiments: Theoretical and methodlogical challenges in creating complex intervations. *The Journal of the Learning Sciences,* 141-178.

Case, R. (1996). Changing views of knowledge and the impact on educational research and practice. In D. R. Olson and N. Torrance (Eds.). *The handbook of education and human development*, (pp. 75-99). Blackwell: Cambridge.

Collins, A. (1999). The changing infrastructure of educational research. In E. C. Lagemann & L. S. Schulman (Eds.), *Issues in educational research: Problems and possibilities.* (pp. 289-298). San Francisco: Jossey-Bass.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18 (9), 27-32.

Gagné, R. (1985). *The conditions of learning and a theory of instruction* (4th ed.). NY: Holt, Rinehart, & Winston.

Gray, J., Groves, H., & Kolodner, J. (2000, June). *A survival guide: The student success handbook for learners in project based science environments.* Presentation at the Fourth International Conference for the Learning Sciences. Ann Arbor, MI.

Greeno, J. G., Collins, A. M., & Resnick, L. (1996). Cognition and learning. In D. Berliner and R. Calfee (Eds.) *Handbook of Educational Psychology,* (pp. 15-46). New York: MacMillan.

Hickey, D. T. (1999, August). *Epistemological reconciliation and the future of motivation research.* Paper presented at the 7th meeting of the European Association for Research on Learning and Instruction. Goteborg, Sweden.

Hickey, D. T. (2000). *Assessment, engagement, & epistemological reconciliation in a technology-supported genetics learning environment.* Research award #REC99-027 from the National Science Foundation's Program on Research in Educational Policy and Practice (2000-2003)

Hickey, D. T., Wolfe, E. W., & Kindfield, A. C. H. (2000). Assessing learning in a technology-supported genetics environment: Evidential and consequential validity issues. *Educational Assessment*, 6 (3), 155-196.

Holyoak, K. J. & Thagard, P. (1997). The Analogical Mind. *American Psychologist,* Vol. 52, No. 1

Kolodner, J. L. (1993). *Case-based reasoning*. San Mateo, CA: Morgan Kaufman.

Kolodner, J. L. (1997). Educational implications of analogy: A view from case-based reasoning. *American Psychologist, 52,* 57-66.

Kolodner, J. L., Crismond, D., Fasse, B., Gray, J., Holbrook, J., & Puntembakar, S (in preparation). *Putting a learner-centered learning-by-design curriculum into practice: Lessons learned.* EduTech Institute of Georgia Institute of Technology.

Kolodner, J. L., Crismond, D, Gray, J., Holbrook, J., & Puntembakar, S. (1998). Learning by Design: From theory to practices. In M.Guzdial, J. Kolodner, A. Bruckman, & A. Ram (Eds.), *Proceedings of the Third Annual International Conference of the Learning Sciences*. Charlottesville, VA: Association for the Advancement of Computers in Education.

Pogrow, S. (1998). What is an exemplary program and why should anyone care? A reaction to Slavin & Klein. *Educational Researcher, 27,* (7), 22-28.

Pogrow, S. (1999). Rejoinder: Consistently large gains and high levels of achievement are the best measures of program quality: Pogrow responds to Slavin. *Educational Researcher, 28,* (8), 24-26, 31.

Pomplum, M. (1996). Cooperative groups: Alternative assessment for students with disabilities. *The Journal of Special Education, 30* (1), 1-17.

Quellmalz, E., Schank, P., Hinojosa, T., & Padilla, C. (1999). *Performance assessment links in science* (ERIC/AE Digest Series EDO-TM-99-04). ERIC Clearinghouse on Assessment and Evaluation: University of Maryland.

Slavin, R. (1999). Rejoinder: Yes, control groups are essential to program evaluation: A reponse to Pogrow. *Educational Research, 28* (3), 36-38.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning, and assessment in the classroom.* Buckingham: Open University Press.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Erlbaum.

Venezky, R. L. (in press). An alternative perspective on Success for All. In K. Wong (Ed.), *Advances in Educational Policy* (Vol. 4).

Participatory Rituals Rating Scale
(Adapted from Pomplan, 1996)

1	2	3	4	5	Rater	
					Group	

A. Group members demonstrate collaborative participation.

1	3	5	
Work is done by individuals; No teaming is evident.	Moderate amount of cooperation; some members not involved.	Teamwork, participation and collaboration are high; decision making is shared	*Set-up* / *task* / *questions*

B. Group members engage in inquiry, problem solving, and co-construction of knowledge.

1	3	5	
Very little questioning or brainstorming in the group.	Members question and inquire; offer ideas, information and solutions.	Members elaborate, modify, and justify own and others' ideas.	*Set-up* / *task* / *questions*

C. Group members use group-oriented language.

1	3	5	
Talking is self-oriented; don't answer questions.	Both task-oriented and self-oriented talk occurs ("I").	Talking is task-oriented and group-oriented ("we").	*Set-up* / *task* / *questions*

D. Group members listen to others.

1	3	5	
Speak when others speak; don't listen; interrupt; ignore others.	Members listen only part of the time or only some members listen.	Listen and look at speaker; ask questions to clarify.	*Set-up* / *task* / *question*

E. Group members support each other.

1	3	5	
Physically turn away; negative comments; put-downs, sarcasm.	Listen but no nod when listening; neutral comments ("OK").	Listen and nod when listening; positive comments ("Good idea").	*Set-up* / *task* / *question*

F. Group members understand project and plan tasks.

1	3	5	
Unclear goals; unclear directions; no plan or schedule.	Only some of group understands goals and understands plan.	All understand goal and directions; group makes a plan.	*Set-up* / *task* / *question*

G. Group members handle conflicts.

1	3	5	
Differences act to greatly; restrict progress on activities.	Conflicts interfere with quality of activity and project.	Conflicts and differences are resolved and are not an issue.	*Set-up* / *task* / *question*

H. Group members stay on task.

1	3	5	
Group unfocused; not working on task.	Group partially on task; on task some of the time.	Group stays on task; completing work.	*Set-up* / *task* / *question*

I. Group members use "science talk".

1	3	5	
No use of science terminology and/or principles in group discussions.	Some use of science terminology or principles but many missed opportunities.	Frequent use of science terminology or principles when appropriate.	*Set-up* / *task* / *question*

1 2 3 4 5

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM031545

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: PALS-Supported Performance Assessments in the Learning by Design Project

Author(s): Daniel T. Hickey, Jennifer Holbrook

Corporate Source: Georgia State University

Publication Date: April, 2000

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: [signature]	Printed Name/Position/Title: Daniel T. Hickey, Asst Prof.
Organization/Address: Georgia State University	Telephone: 404 651 0127 — FAX: 404 651 4901
	E-Mail Address: dhickey@GSU.EDU — Date: Jun 6 2000

(over)

ERIC® Clearinghouse on Assessment and Evaluation

University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

Tel: (800) 464-3742
(301) 405-7449
FAX: (301) 405-8134
ericae@ericae.net
http://ericae.net

May 8, 2000

Dear AERA Presenter,

Hopefully, the convention was a productive and rewarding event. As stated in the AERA program, presenters have a responsibility to make their papers readily available. If you haven't done so already, please submit copies of your papers for consideration for inclusion in the ERIC database. We are interested in papers from this year's AERA conference and last year's conference. If you have submitted your paper, you can track its progress at http://ericae.net.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the **2000 and 1999 AERA Conference**. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

Please sign the Reproduction Release Form enclosed with this letter and send **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can mail your paper to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: AERA 2000/ERIC Acquisitions
University of Maryland
1129 Shriver Laboratory
College Park, MD 20742

Sincerely,

Lawrence M. Rudner

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

ERIC is a project of the Department of Measurement, Statistics & Evaluation