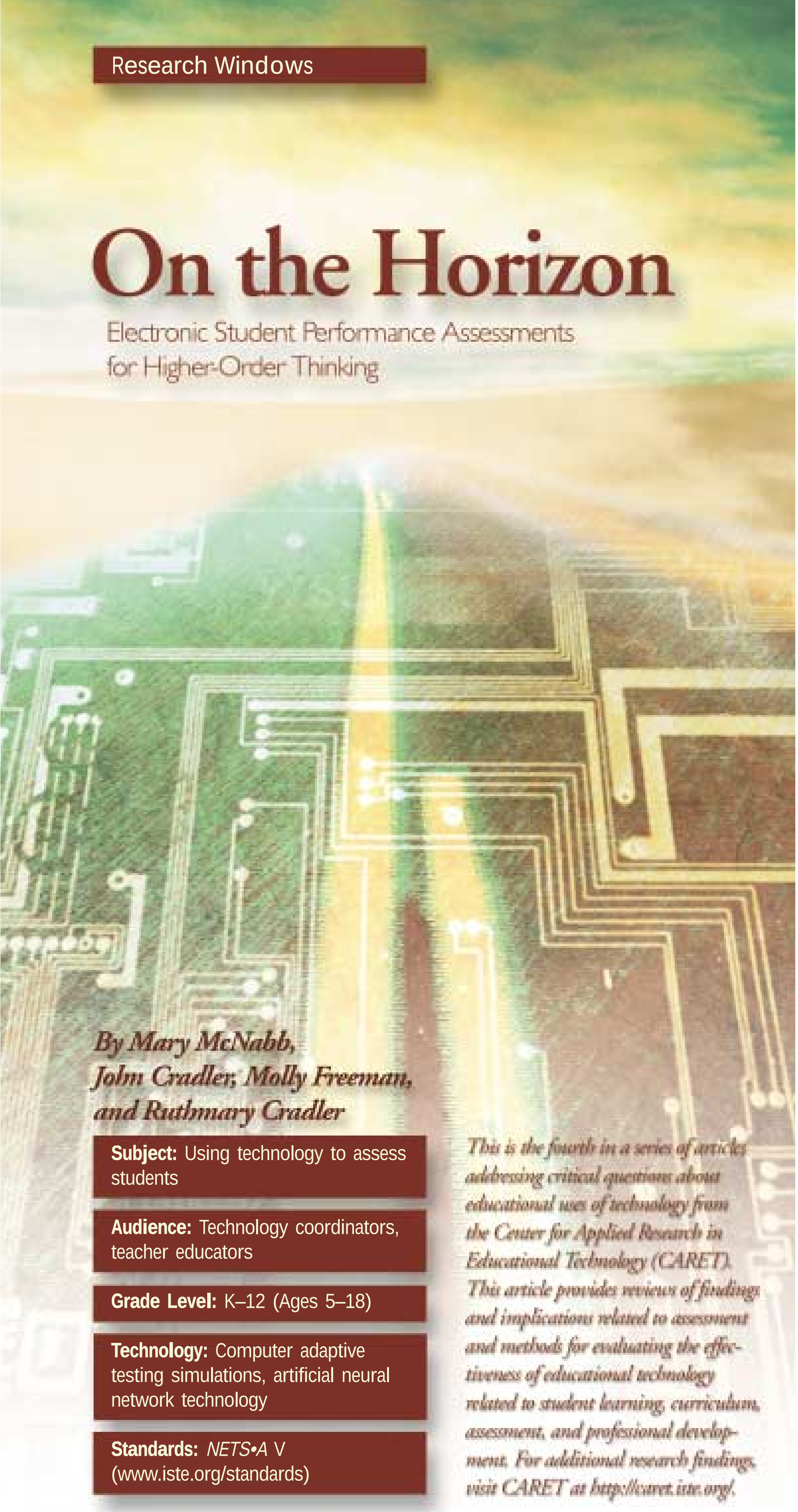

On the Horizon

Electronic Student Performance Assessments for Higher-Order Thinking

By Mary McNabb, John Cradler, Molly Freeman, and Ruthmary Cradler

Subject: Using technology to assess students

Audience: Technology coordinators, teacher educators

Grade Level: K–12 (Ages 5–18)

Technology: Computer adaptive testing simulations, artificial neural network technology

Standards: *NETS•A* V (www.iste.org/standards)

This is the fourth in a series of articles addressing critical questions about educational uses of technology from the Center for Applied Research in Educational Technology (CARET). This article provides reviews of findings and implications related to assessment and methods for evaluating the effectiveness of educational technology related to student learning, curriculum, assessment, and professional development. For additional research findings, visit CARET at http://caret.iste.org/.

The No Child Left Behind (NCLB) Act has made the assessment of student performance a national priority. NCLB establishes financial incentives for schools to improve students' scores on standards-based academic assessments. *(Editor's note:* See the October Research Windows column, *L&L 30*[2], p. 46, for details on NCLB and technology. Also, see Resources at the end of this article for the NCLB and other Web addresses.) Many educators are selecting curriculum and electronic learning resources and applications that are most likely to improve students' test scores on high-stakes standardized measures. The fear is that this movement will result in a back-to-basics trend. Several studies identified by CARET, however, suggest that a variety of technology applications can promote development of higher-order skills and knowledge.

For at least the past two decades, researchers have stressed the need for correlated supplements to standardized tests that assess higher-order thinking and problem solving along with student performance and knowledge directly related to specific learning experiences (Costa, 1985). A process was developed to assess the development of thinking skills and problem skills resulting from the specific use of instructional computing applications (Cradler, 1985; Pogrow, 1985). The value of a balanced measurement approach and advancements toward reliable and efficient measures of higher-order thinking skills associated with technology use has become a primary focus of recent efforts by researchers and educators (McNabb, Hawkes, & Rouk, 1999; Pellegrino, 2002; Pellegrino, Chudowsky, & Glaser, 2001).

To help educators understand new developments in assessment practices relating to higher-order thinking, CARET staff has reviewed research that addresses these two questions:

1. How can technology be used to enable the instruction and assessment of higher-order thinking skills?
2. How can technology be used to effectively assess and monitor student performance?

Assessment as an Instructional Intervention

The Committee on the Foundations of Assessment, sponsored by the National Academy of Sciences (NAS), describes an important role for electronic assessments in what is taught and how it is taught in the 21st century (Pellegrino et al., 2001). The role for technology in the assessment process parallels developments in learning theory that support assessment-centered curricula. For example, Riel (2000) reports that

> Effective learning environments are assessment-centered because knowing what students are learning and what they need to know is critical information for shaping learning environments. When students don't learn, the community suffers. High stakes testing only makes clear the dimensions of the problem. Assessment that is ongoing and prescriptive can help make every student a valued member of society without the negative effects of ranking schools.

Assessment-centered lessons, units, or projects require teachers to identify learning outcomes and to explicitly link all lesson components, criteria, and rubrics for assessing what students are learning. Additionally, programs with embedded assessments can provide students and teachers with immediate feedback about what is being learned. This feedback can then inform changes in lesson difficulty, presentation, modality, and approach. Embedded performance assessments help teachers modify activities and teaching strategies as needed to facilitate student understanding.

Technology can play a key role in supporting assessment-centered teaching and learning practices. A recent application of technology to assessment is the State2State Assessment Exchange being developed by the U.S. Open e-Learning Consortium (USOeC) (Clements, Ligon, Burt, & Mangino, 2002). The USOeC intent is "to provide a pool of well-designed standards-related materials to teachers in the classroom to be used to assess student understanding on an as-needed basis" (Clements et al., p. 8). The computer-based assessments provide:

- Audio and visual components not available with paper-and-pencil tests
- Materials for specific levels of students
- Instant feedback on student performance
- Instant identification of where students are having trouble
- Resources for the student or teacher to address student learning needs

Furthermore, test items from the database are:

- Adaptable to students with special needs
- Interactive in ways that are familiar to students
- Accessible to teachers for designing their own tests (Clements et al., 2002, p. 11)

The value of technology applications for assessment purposes has substantial corroboration. Pellegrino (2002) says,

> By enriching task environments through the use of multimedia, interactivity, and control over the stimulus display, it is possible to assess a much wider array of cognitive competencies than has heretofore been feasible. … A significant contribution of technology has been to the design of systems for implementing sophisticated and classroom-based formative assessment practices. Technology-based systems have been developed to support individualized instruction by extracting key features of learners' responses, analyzing patterns of correct and incorrect reasoning, and providing rapid and informative feedback to both student and teacher. (p. 10)

And, according to Bransford, Brown, and Cocking (1999), assessment-centeredness provides multiple opportunities to make students' thinking visible so they can receive feedback and be given chances to revise their thinking and to learn about their own learning. This type of feedback requires embedding formative assessments into instructional practices, which benefits learners by:

- Identifying prior knowledge and tailoring instruction to draw on it
- Acknowledging and building on achievement levels
- Recognizing misconceptions or knowledge gaps that instruction can address
- Stimulating future learning by engaging and motivating learners in deeper levels of analysis on an individualized basis (Bransford et al.; Chung & O'Neil, 1997; Pellegrino et al., 2001)

School improvement efforts across America are focusing on fostering students' deeper content area understanding, advanced reasoning and inquiry skills, and collaboration/teaming abilities. These are often referred to as *21st-century skills.* Rote memory tests, used for external auditing purposes, were not designed to measure these types of learning outcomes.

Several studies identified by CARET, however, suggest that a variety of technology applications can promote development of higher-order skills and knowledge.

These more complex assessments, however, require the robust electronic information processing capacity of today's technologies to administer, record, and analyze data efficiently and to report to teachers and learners on a timely basis.

Technology-Assisted Assessment

Among the research organizations taking a lead in developing specifications and methodologies for electronic assessments is the Center for Research on Evaluation, Standards, and Student Testing at the University of California, Los Angeles (CRESST/UCLA). Baker (2000) explains that most electronic testing research and development efforts to date have used technology primarily for improving the efficiency of existing testing practices. Methods for using technology to expand the domains of what can be measured, however, are emerging. For example, networked environments provide ways to measure individual and team performance strategies in problem-solving simulations. Students can use digital authoring tools as simple as Inspiration and HyperStudio to demonstrate their mental maps of relationships among facts, concepts, and processes. More sophisticated electronic tools are being developed to:

- Score students' concept maps based on comparison to others' maps on the same topic (O'Neil & Schacter, 1997)
- Facilitate scoring of open-ended response assessments such as essays (Burstein, Marcu, Andreyev, & Chodorow, 2001; Chung & O'Neil, 1997; Foltz, Gilliam, & Kendall, 2000; Kintsch, Steinhart, Stahl, & LSA Research Group, 2000)
- Assess complex problem-solving skills (Stevens, Ikeda, Casillas, Palacio-Cayetano, & Clyman, 1999; Chung, de Vries, Cheak, & Stevens, 2002).

Interactive Multi-Media Exercises (IMMEX) has been identified as one of the most promising electronic assessments for problem-solving skills currently available to K–12 students (Pellegrino, 2002). IMMEX allows teachers to quickly evaluate students' understanding of content, and to apply that information in real-world situations by mapping a student's progress through a problem-solving scenario. A series of IMMEX studies found that the same information is not processed or interpreted the same way by students whose levels of experience and knowledge understanding vary (Stevens et al., 1999; Vendlinski & Stevens, 2002). The instant feedback from IMMEX provides teachers with data about the range in students' problem-solving abilities, which then can inform subsequent instruction. The IMMEX library contains K–12 problem sets that can be accessed by filling out a New User Registration at the IMMEX Web site. (*Editor's note:* See *L&L 28*[7], p. 26, for an article about IMMEX.)

Technology-Based Assessments for Instructional Decision Making

Kingsbury (2002) reports that in recent years, some school districts and state departments of education have adopted the computer adaptive test (CAT) as a primary measure of student achievement and growth to inform instructional decisions. CATs can be administered as measures independent of instruction and/or actually embedded in adaptive instructional systems (Fletcher, 2002; Park, 1996). Testing and monitoring procedures embedded within adaptive instructional systems are designed to provide feedback to teachers and students for instructional decision making and to guide learners in their own learning paths (Park). With the CAT approach, a student who performs well on a set of intermediate-level questions will be presented with advanced-level questions. If a student performs poorly on the intermediate-level questions, the test presents a set of novice-level questions until the student's highest sustainable performance level is established as the basis for his or her test score.

Adaptive instructional systems designed to present content according to learners' prior knowledge usually embed multiple-choice assessments of content area knowledge. Systems that use interaction strategies such as Socratic questioning are aimed at scoring learners' open-ended responses and providing immediate feedback using artificial intelligence techniques such as latent semantic analysis (LSA). Automated scoring techniques that use LSA compare essay content to other scored essays or known reference materials (such as a textbook) on the topic and assign scores based on the scores of other most-similar essays. This technique produces scores highly correlated with human raters' scoring of open-text responses, and its efficiency makes essay assessments more feasible on a large scale (Burstein et al., 2001; Chung & O'Neil, 1997; Foltz et al., 2000; Kintsch et al., 2000). Find an example of an essay assessment tool at Colorado University's LSA Web page.

The Horizon Is Closer

Certainly, a shift in thinking about assessment is afoot. The infusion of electronic assessments into daily classroom routines is immediately foreseeable for the assessment of basic academic and higher-order thinking skills. Large-scale assessments currently used in the United States are undergoing enormous pressure to change because of outmoded cognitive-scientific bases for test design, misalignment to curriculum standards, the performance differential for learner groups, and lack of useful feedback to help individuals improve. Bennett (2001) claims that technological advancements, primarily the Internet, will provide means for making assessments interactive, broadband, switched, networked, and standards-based. These more complex assessments, however, require the robust electronic information processing capacity of today's technologies to administer, record, and analyze data efficiently and to report to teachers and learners on a timely basis. According to Bennett (2002a), at least six states are exploring and/or implementing assessments that can be administered electronically. South Dakota is the first to administer its state assessment exclusively on the Internet—the paper version is no longer used (Bennett, 2002b). These states, however, are using electronic assessments primarily to assess basic academic skills, with minimal emphasis on higher-order skills and knowledge. Examples of large-scale, electronically administered, basic skills tests are available at the Oregon Department of Education and Virginia Department of Education Web sites.

As availability of and interest in electronic assessments increase, teachers will need professional development in using formative data to guide students' learning. Educators who understand the nature and purpose of new electronic assessments and methods for interpreting the data will be better able to select tests that align with the content and performance skills their students are required to learn. These teachers will be better able to interpret the test results in terms of the specific needs of their students and to adjust their instruction accordingly.

Educators who understand the nature and purpose of new electronic assessments and methods for interpreting the data will be better able to select tests that align with the content and performance skills their students are required to learn.

Resources

CARET: http://caret.iste.org
Colorado University's LSA Web page: http://lsa.colorado.edu
CRESST/UCLA: www.cse.ucla.edu
IMMEX: www.immex.ucla.edu
NCLB: www.nochildleftbehind.org
Oregon Department of Education: www.ode.state.or.us/asmt/development/rfpCAM.htm
U.S. Open e-Learning Consortium (State2State Assessment Exchange): www.ctlt.org/projects/us_open_e_learning
Virginia Department of Education: www.pen.k12.va.us/VDOE/Technology/soltech/rfp/rfpweb2000.pdf

References

Baker, E. (2000). *Understanding educational quality: Where validity meets technology* [Online]. Fifth annual William H. Angoff Memorial Lecture presented at Educational Testing Service, Princeton, NJ. Available: www.ets.org/research/pic/angoff5.pdf.

Bennett, R. E. (2002a). *Issue brief: Using electronic assessment to measure student performance* [Online]. Washington, DC: NGA Center for Best Practices. Available: www.nga.org/center/divisions/1,1188,C_ISSUE_BRIEF%5ED_3126,00.html.

Bennett, R. E. (2002b, April). *Inexorable and inevitable: The continuing story of technology and assessment.* Panel presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Bennett, R. E. (2001). How the Internet will help large-scale assessment reinvent itself. *Education Policy Analysis Archive* [Online], *9*(5). Available: http://epaa.asu.edu/epaa/v9n5.html.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school* [Online]. Washington, DC: National Academy Press. Available: www.nap.edu/html/howpeople1/.

Burstein, J., Marcu, D., Andreyev, S., & Chodorow, M. (2001, July). *Towards automatic classification of discourse elements in essays.* Paper presented at the 39th annual meeting of the Association for Computational Linguistics, Toulouse, France. Available: www.ets.org/research/dload/burstein.pdf.

Chung, G. K. W. K., de Vries, L. F., Cheak, A. M., & Stevens, R. H. (2002, April). *Cognitive process validation of an online problem solving assessment.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Chung, G. K. W. K., & O'Neil, H. F., Jr. (1997). *Methodological approaches to online scoring of essays.* Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing. Available: www.cse.ucla.edu/CRESST/Reports/TECH461.PDF.

Clements, B. S., Ligon, G. D., Burt, W., Mangino, E. (2002). *Why a State2State Assessment Exchange?* A policy framework for ESEA and online assessment & proposed S2S Exchange Usage Model prepared for the U.S. Open e-Learning Consortium by Evaluation Software Publishing, Inc.

Costa, A. (1985). *Developing minds: A resource book for teaching thinking.* Arlington, VA: Association for Curriculum and Supervision.

Cradler, J. (1985). CompTHINK and criteria for selecting software. In A. Costa (Ed.), *Developing minds: A resource book for teaching thinking* (pp. 256–265). Arlington, VA: Association for Curriculum and Supervision.

Cradler, J. (2002). New federal programs suggest an expanded role for technology. *Learning & Leading with Technology, 30*(2), 46–49, 56–57.

Fletcher, J. D. (2002). Is it worth it? Some comments on research and technology in assessment and instruction. In National Academies Center for Education (Ed.), *Technology and assessment: Thinking aloud—Proceedings from a workshop* (pp. 26–39). Washington, DC: National Academy Press. Available: http://books.nap.edu/books/0309083206/html/26.html.

Foltz, P. W., Gilliam, S., & Kendall, S. (2000). Supporting content-based feedback in online writing evaluation with LSA. *Interactive Learning Environments, 8*(2), 111–129.

Kingsbury, G. G. (2002, April). *An empirical comparison of achievement level estimates from adaptive tests and paper-and-pencil tests.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Kintsch, E., Steinhart, D., Stahl, G., & LSA Research Group. (2000). Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments, 8*(2), 87–109.

McNabb, M. L., Hawkes, M., & Rouk, U. (1999). *Critical issues in evaluating the effectiveness of technology.* Washington, DC: U.S. Department of Education, Office of Educational Technology.

O'Neil, H., Jr., & Schacter, J. (1997). *Test specifications for problem-solving assessment* [Online]. Los Angeles: National Center for Research on Evaluation, Standards, and Student

Assessment continued on page 59.

Assessment continued from page 53.

Testing. Available: www.cse.ucla.edu/CRESST/Reports/TECH463.PDF.

Park, O. (1996). Adaptive instructional systems. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 634–664). New York: Simon & Schuster Macmillan.

Pellegrino, J. W. (2002, April). *Advances in cognition, assessment, and technology: Deep structure issues.* Panel presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment* [Online]. Washington, DC: National Academy Press, National Research Council, Center for Education, Board on Testing and Assessment, Committee on the Foundations of Assessment. Available: www.nap.edu/books/0309072727/html/.

Pogrow, S. (1985). HOTS: A computer-based approach. In A. Costa (Ed.), *Developing minds: A resource book for teaching thinking* (pp. 239–240). Alexandria, VA: Association for Supervision and Curriculum Development.

Riel, M. (2000, September). *White paper: New designs for connected teaching and learning* [Online]. Paper presented at the Secretary's Conference on Educational Technology 2000. Available: www.gse.uci.edu/mriel/whitepaper/.

Stevens, R., Ikeda, J., Casillas, A., Palacio-Cayetano, J., & Clyman, S. (1999). Artificial neural network-based performance assessments. *Computers in Human Behavior, 15*, 295–313.

Vendlinski, T., & Stevens, R. H. (2002) A Markov model analysis of problem-solving performance and transfer. *Journal of Technology, Learning and Assessment* [Online], *1*(3). Available: www.bc.edu/research/intasc/jtla/journal/v1n3.shtml.

Mary McNabb, EdD, works as a consultant for CARET focusing on investigating effective uses of technology for teaching, learning, and assessment purposes. She began using networked computers to teach English in 1993, which fostered her interest in earning a doctorate in educational technology in 1996. She worked at the North Central Regional Educational Laboratory (1996–2000) and has served on several Preparing Tomorrow's Teachers to Use Technology (PT^3) grant projects, including ISTE's NETS for Teachers leadership committee and the PT^3 core group of evaluators.

John Cradler is the co-director of the CARET project and president of Educational Support Systems (ESS). He first became involved with technology when working with the South San Francisco Unified School District. In 1974, while looking for a more efficient way to provide reading tutoring for special education students, he discovered a Stanford University project on computer-assisted instruction, which was adapted to support a phonics tutoring project called Success Controlled Optimal Reading Experience (SCORE).

Molly Freeman, PhD, currently conducts research with ESS and, since 1996, has consulted with the Internet Institute of Santa Clara County Office of Education. Molly's experience with technology began with punch-card coding responses to interviews of children for one of the first studies of school integration in Riverside, California. As Chairperson of the Division of Education and Human Development at Holy Names College from 1980 to 1986, Molly participated in the early stages of computer integration with teacher preparation programs.

Ruthmary Cradler works as a consultant, specializing in evaluation of educational technology program implementation and development. She is a member of the PT^3 national advisory committee for program evaluation. She holds an MA from San Francisco State University. Thirty years ago, she was excited about the electric (versus hand-cranked) ditto machine. She was also excited about her first Cuisinart food processor, which had a greater long-term effect on her life than the ditto machine.