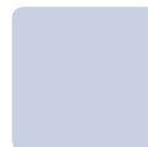


Is It Live, Or Is It Memorex?

**Dr. Michael Bunch, Sr. Vice President of
Research & Development**

2013



MI  **MEASUREMENT
INCORPORATED**
EXCELLENCE IN ASSESSMENT™
www.measurementinc.com

EXCELLENCE IN ASSESSMENT

IS IT LIVE, OR IS IT MEMOREX?

DR. MICHAEL BUNCH

The Educational Testing Service Invitational Conference of 1985 had as its theme “The Redesign of Testing for the 21st Century.” It has been a little over 25 years since the conference proceedings were published, and it is amazing to see that the future still looks a lot like it did then: cognitive-based assessment, online assessment, widespread use of computer-adaptive testing, universal access to technology, and instantaneous reporting of test results. So many wonderful things, still within our view but just beyond our grasp!

Although the projections of the 1985 ETS conference were fairly modest in scope, [Bob Linn](#) sounded a note of caution. He predicted that the educational, psychometric, and technological breakthroughs of the 1980s and beyond would not be sufficient to cause fundamental change in educational assessment because they would have to overcome human reluctance to abandon a system that actually worked rather well, in terms of efficiency and cost effectiveness. We face the same issues today.

A recent watershed moment illustrates not only the current state of technology but the resistance to it as well. Dr. Ellis Page pioneered a technology that allowed computers to read and score essays and demonstrated its viability nearly 50 years ago. For decades, Dr. Page’s Project Essay Grade (PEG) remained an academic curiosity.

Fast forward to 2012. Automated scoring of essays by artificial intelligence (AI) has been commercialized. Nine vendors (eight companies and one university) participated in a demonstration of the viability of AI scoring of a wide variety of essay types over several grades and genres. Measurement Incorporated (MI) was one of those companies. The conclusion of the independent evaluation of all these AI scoring programs was that, overall, computers could score essays as well as humans could. Five of the companies (with MI in the lead) actually outperformed the human readers. In a follow-up demonstration in the summer of 2012, computers scored content-based responses to open-ended questions as well. Once again, MI took the lead.

The immediate response of many academics and journalists was that computers just can’t appreciate a fine piece of writing or give counsel to a budding Steinbeck. Their objection to computer scoring of essays had nothing to do with the technical adequacy of the process or product; it was simply that the machine is not human.

The recording industry has faced similar criticisms for more than a hundred years. No matter how good recording became or how faithfully it reproduced the sound of the human voice (or instruments played by humans), people objected, simply because it was not human. An iconic set of television ads in the 1980s asked, “Is it live, or is it Memorex?” Today, having moved from

cylinders to wax to vinyl to 8-track to mini-cassette to CDs to MP3 formats and beyond, few people worry whether the sound they are hearing is live or a recording. They don't care because they know they can listen to their music in some form wherever they go. They can't do that with live music. Convenience, cost, and fidelity have won.

In the same way, we will eventually embrace automated scoring of essays, not because AI scoring is so superior to human scoring but because AI scoring can provide instantaneous and reliable scores. Humans can't do that. Or if they did, it would be terribly expensive. MI is already providing scores for one statewide writing assessment as well as instantaneous scores for commercially available tests. Other companies are providing writing assessments scored by computers for admissions purposes. Companies throughout the world are using AI to score the writing of their employees and prospective employees. Once again, convenience, cost, and fidelity will win.

Throughout Connecticut and North Carolina, MI provides online writing exercises that are scored by PEG. Students can write as many essays as they like, on a variety of topics, and submit them to PEG for scoring. Scores on six dimensions of writing come back in two to three seconds. In addition, the scoring system directs students to tutorials designed to help them improve one or more of the six dimensions of their writing. The system also allows teachers to monitor the entire process, check the various drafts, and leave notes for the students. This system has become quite popular because it teaches as it tests and allows teachers and students to interact along the way.

The future turns out to be one in which human beings thoughtfully use available technology to achieve fundamental human desires – to educate their children and prepare them for an even more challenging and exciting future. We are pleased to be a part of that future and to make this useful technology available.

ABOUT ROBERT L. LINN

Robert L. Linn, is distinguished professor emeritus at the University of Colorado at Boulder, past president of the American Educational Research Association, editor of *Educational Measurement, 3rd Edition*, and technical advisor to several state education agencies and national testing organizations. He has published extensively in the field of educational assessment.

ABOUT DR. MICHAEL BUNCH

Dr. Michael B. Bunch is Senior Vice President of Research and Development at Measurement Incorporated. His expertise in the development of criterion-referenced, high-stakes testing programs is well known among assessment professionals. He contributes regularly to these programs through national organizations such as the American Educational Research Association, the Council of Chief State School Officers, and the National Council on Measurement in Education. Dr. Bunch published (along with Dr. Gregory Cizek), the graduate-level text, *Standard*



Setting (Sage Publications, 2007) and is in great demand as a reviewer and discussant for standard-setting research and as a leader of standard-setting activities.