K–12 EDUCATION TESTING:

K–12 education is one of the government's most challenging responsibilities, and the K–12 testing process is one of education's most poorly managed activities.

BETTER, FASTER, AND CHEAPER

C. JACKSON GRAYSON, JR.

The United States, propelled by the 2001 No Child Left Behind Act (NCLB), has embarked on a course of testing K–12 students on a mammoth scale never before required in U.S. history.

Whether these expanded testing requirements are good or bad for student learning is debatable. However, there is no doubt that the new testing requirements under NCLB will strain the capacity and capability of schools, districts, states, test publishers, and scoring organizations to administer and use the tests in a timely, accurate, and cost-efficient way.

Even before this ramp-up, there were already three big problems in K–12 student standardized testing:

- *Delays.* It takes too long to get the test scores back. The total "cycle time" from test taking to return can range from one to eight months.
- *Errors.* There are far too many errors in testing, scoring, and reporting. These mistakes can have significant consequences.
- *Costs.* The entire testing and scoring process costs too much. Errors and inefficient processes along the entire supply chain inflate the costs.

C. JACKSON GRAYSON, JR. is Chairman of the nonprofit American Productivity & Quality Center, located in Houston, Texas. He can be contacted at jgrayson@apqc.org

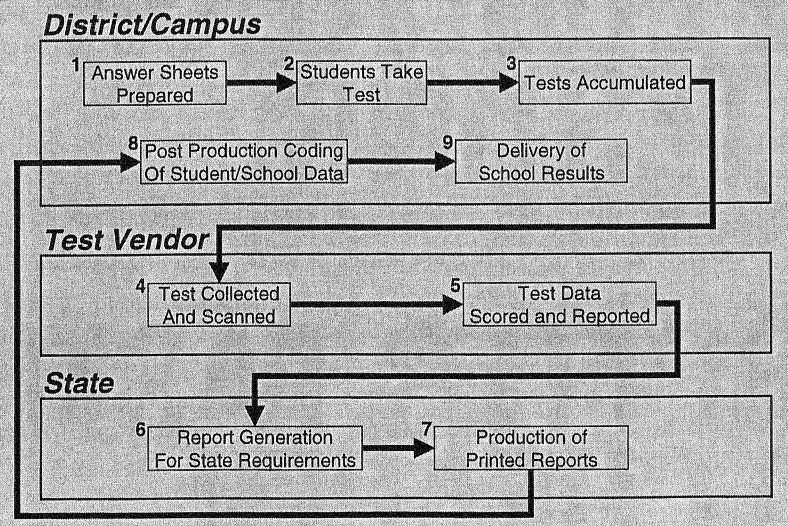
This article describes why these problems have arisen, why they are continuing, and why they are likely to get worse if changes are not made. It also recommends a way to turn the situation around through the use of Six Sigma in an Integrated Supply Chain. Used by business for more than 15 years, the Six Sigma method holds promise for making the entire testing cycle better, faster, and cheaper.

Potential disaster

These problems—if not solved—are going to get worse. Not only is the sheer volume increasing, there is a growing demand for faster and more accurate test scoring, and an increased demand for costly customized tests to meet new state standards. Each customized test that a state orders must be designed, written, edited, reviewed by state educators, field-tested, checked for validity and bias, and calibrated to earlier tests. It requires a battery of people trained in test development, educational statistics, and psychometrics, all of which strains capacity, and increases expenses and delays. The stakes are high. Decisions made on the tests will determine:

- Whether students will graduate
- Whether they will be retained or promoted to the next grade
- Who goes to summer school
- Whether the school or district will be labeled a "school needing improvement" (also called "low performing") under NCLB, which car-

EXHIBIT 1 The Testing Cycle



ries with it certain significant consequences for the school, district, and state

These problems come at a time when the system already has delays, errors, and costs, and when state budgets are shrinking. All this has the potential for disaster unless changes are made in how these tests are taken, scored, and returned so that they become more accurate, timely, and affordable.

Who's to blame?

The truth is . . . everyone. Everyone in the entire testing cycle—schools, districts, states, test publishers, scoring organizations, legislators, and the U.S. Department of Education. However, as Dr. W. Edwards Deming, the quality guru, colorfully—and correctly—reminded us: "Don't fix the blame. Fix the system!"

So, what does the testing "system" look like? Examine the diagram in Exhibit 1 for the interrelationships of the key elements of the testing cycle detailed below.

- Schools and districts typically administer standardized student tests in the Spring. A few give them in the Fall.
- Most multiple-choice questions are answered with pencil on "bubble sheets." Those requiring an open-ended answer are handwritten by the students.
- If the state uses "pre-coding," the student information is "coded" into the answer booklets. If the tests aren't pre-coded, the district codes the student information by hand.
- The answer booklets are bundled and sent directly by mail to the test publishers and scoring organizations (hereinafter publishers/scorers).
- The publishers/scorers resolve the data and student matching issues, compare the answer keys with the students' answers, and compute a score for each student.
- The publishers/scorers then send the results to both the districts and the state.
- The state aggregates the individual results, then sends reports to the U.S. Department of Education, as well as to the districts.
- The districts pass the results to the principals, teachers, parents, and community. Confidentiality is maintained for individual students' scores.

The process flow in Exhibit 1 looks straightforward. Why doesn't it work well? Where does the system break down?

School and Districts. When schools and districts give tests and send them to the publishers/scorers, the answer sheets often contain (1) inaccurate or missing student data, (2) erasures or smudges, (3) damaged or torn forms, and (4) illegible handwritten data.

Also, schools and districts sometimes send incorrect or incomplete student information so that test results can't be matched with the correct stu-

dents. Many phone calls, e-mails, and faxes fly between the publishers/scorers and the districts to reconcile these problems. Results: delays, errors, and cost.

Test Publishers/Scorers. Test publishers/scorers also have inefficient, error-generating processes of their own. Over the years, they have created processes to "find and fix" these problems—numerous check points, multiple reviews, time in handling, reviewing, and moving paper—time spent in firefighting and expediting to speed delivery or fix errors that lead to large spikes in temporary manpower to handle peak loads. Despite all this effort and cost, significant errors are sometimes made that have negative consequences for schools, districts, and students:

- A testing company made errors in grading exams that were given to nearly 340,000 Georgia elementary school students.
- A flawed answer key at a test publisher incorrectly lowered multiple-choice scores for 12,000 Arizona students.
- In New York City, nearly 9,000 students attended summer school because they had mistakenly been told they had failed the state assessment.
- Based on erroneous scores calculated by a contractor, one state sent thousands of children to summer school in the mistaken belief that their performance was poor enough to meet the criteria for summer intervention.

These errors are not only bad for districts and students, they are also bad for the publishers/scorers—leading to loss of customers, penalties when performance clauses in contracts are not met, and sometimes lawsuits. One test scoring company paid $7 million in damages to affected students and their families in Minnesota because of such errors.

States. States are responsible for managing the process. However, they are often caught in the middle between the legislatures, the districts, the publishers/scorers, and the federal government. States can add to the delays and costs when they make frequent changes to the testing process, administrative procedures, and reports. States frequently lack the staff resources to manage the process or handle the workload. The General Accounting Office (GAO) reported last year that 16 states didn't even monitor the scoring done by the contractors.

All parties are trying hard—schools, districts, publishers/scorers, states—but unless changes are made, and soon, the testing cycle as a train wreck in the making will persist—similar to where business was about 25 years ago.

THERE IS HARDLY A MAJOR SUCCESSFUL BUSINESS FIRM IN THE UNITED STATES THAT DOES NOT USE SUPPLY CHAIN MANAGEMENT, LINKING CUSTOMERS AND SUPPLIERS IN UNPRECEDENTED WAYS TO ACCOMPLISH SOMETHING TOGETHER THAT WAS IMPOSSIBLE WORKING INDEPENDENTLY.

Learning from business history

Education today is about where U.S. businesses were in the 1970s and early 1980s—with rising costs, dissatisfied customers, huge waste, large quantities of defects, and silo organizations. Japan and Germany were taking over U.S. markets.

American businesses reacted predictably. First came denial: "We're No. 1!" Then came excuses: "Nothing wrong with us. It's the Japanese government subsidies, their culture, their cheap labor." They also blamed the U.S. unions, their own employees, their suppliers, even their customers: "Customers don't *care* about quality. If you want to them quality, give them leather seats. That's something they can smell."

When these excuses failed, businesses finally got the message and began to change. They began listening to customers, involving employees, using quality tools, got on airplanes to benchmark the Japanese, and came back, as David Kearns, then-Chairman of Xerox put it, "Scared, and determined to change."

Change, they did. Among many things, they

- Created a major focus on productivity and quality: TQM, Baldrige, employee involvement, empowerment, teams, union–management cooperation, statistical process control, to name a few.
- Paid attention to customer needs and satisfaction.
- Knocked down silo departments, focused on processes, especially cross-functional processes.
- Focused on speed, time-to-market, just-in-time, cycle time reduction, and benchmarking.
- Created integrated supply chains.

Education can learn from business to improve the education testing cycle to make it better, faster, and cheaper. A first step is to treat the entire testing cycle as an *integrated supply chain*, composed of schools, districts, states, and publishers/scorers. The second step is to use *Six Sigma*.

Integrated supply chain

An integrated supply chain is a linkage between all organizations working in a common process. In the education testing cycle, schools, districts, states, and publishers/scorers are all links

in a supply chain from test taking to test return. Right now, they are loosely connected, and operating their processes independently. However, if time, errors, and cost are to be reduced throughout the total supply chain, and be tightly connected, then all links in the chain need to coordinate and work together in collaborative and coordinated ways—not just conversationally, but in process design and operation.

One of the most visible and successful examples of this in business was in the textile/fabric industry. Roger Milliken of Milliken, Inc., convened a meeting of the textile/fabric mills, apparel manufacturers, and retail stores in 1985, and they all agreed to work together as a supply chain to reduce errors, speed up the entire process, decrease costs, and increase profits. This was called "Quick Response" and the gains exceeded all their expectations. The turnaround time in Milliken alone declined from six weeks to *one* week. The response time from fabric ordering to receipt in department stores was cut from 18 weeks to *three* weeks. A West Coast department store cut its purchase order time from seven to 10 days to *three* days.

From the fabric and textile industry, the idea moved quickly to other industries and sectors—food, health industries, housewares, toys, hardware, consumer electronics, pharmaceuticals, home building, etc. This approach sparked the entire business sector to examine, shorten, and improve the entire cycle time in every process—internal and external—in an integrated supply chain mode.

There is hardly a major successful business firm in the United States that does not use supply chain management, linking customers and suppliers in unprecedented ways to accomplish something together that was impossible working independently. They produce a win-win situation for all: suppliers, manufacturers, and customers.

To our knowledge, no one in education testing has yet focused on the test taking and test return as a coordinated and integrated supply chain. There is no reason that the cycle time in testing couldn't be cut from months to weeks to days if all the members of the testing chain got together and agreed to work as an integrated supply chain. Experience in business has shown that it will work, provided:

- They agree to cooperatively map the entire chain from beginning to end, and then start cooperatively working together to improve it.
- There is integration of planning and sharing of information among all parties in the chain.
- There is a common commitment to connect, and keep each stage in the chain sharply separated from the next.
- They agree on information to be supplied to each, on the systems and equipment to be used for transmitting information.
- They agree that all entities will do whatever it takes—different scheduling, more flexible work assignments, different roles and responsibilities to cut cycle time and reduce errors.

EXHIBIT 2 The Costs of Poor Quality (COPQ)

Traditional COPQ
("the tip of the iceberg")

- Overtime
- Penalty Fees
- Payroll
- Contractor Fees
- Materials
- Travel

Lost Opportunity COPQ
(less obvious, below the surface)

- Rework
- Downtime
- Inspection
- Moving things
- Waiting Time
- Idle Time
- Turnover
- Reviewing
- Sorting
- Invoice errors
- Coding errors
- Work duplication
- Lost documents
- Cleaning data

Regardless of whether an Integrated Supply Chain is created, what methodology can be used to get action among the participants, and actually reduce delays, errors, and costs? The first thoughts of action are likely to be traditional improvement methods: launch cost reduction drives, create more quality checkpoints, hire more people, improve processes, increase training, and launch a cost reduction drive. However, they won't work any more than they have in the past, and sometimes will even increase costs and delays. Why?

- Those methods deal mostly with symptoms, not root causes of the problem. They are

SIX SIGMA

Sigma level	Errors per million opportunities for error
1 Sigma	690,000
2 Sigma	308,537
3 Sigma	66,807
4 Sigma	6,210
5 Sigma	233
6 Sigma	3.4

only a temporary pain reliever and, inadvertently, they sometimes add more costs, delays, and errors.

- Most organizations reducing errors with these tools are in a "find and fix" mode, rather than "plan and prevent." The ad hoc improvements are usually "bolted on" to existing processes that often add complexity, rely on inspection, increase costs and delays, and don't work on prevention.
- Most of the cost reduction methodologies tackle the costs reported by the traditional accounting systems, which lump together value-added and non-value-added costs. Business calls this the "hidden factory." Delays, errors, and costs cannot be reduced much unless non-valued-added costs (waste) are clearly identified and reduced. The profile in Exhibit 2 shows the obvious costs captured by traditional accounting systems and the hidden 80 percent below the surface costs that are non-valued-added.

Some gains can be made using any of the traditional methods. However, they don't solve the underlying problems—they are often short-lived and, in many cases, end up producing more delays and costs without solving the underlying system root causes.

Six Sigma

Of all the successful methodologies that business has used over the years, the one that stands out most for its potential to improve the education testing cycle in an integrated supply chain—in the short and long run—is Six Sigma.

Six Sigma has captured the attention and backing of some of America's largest and best organizations, and has been called by some leading business executives "the most powerful breakthrough management tool ever devised." That rhetoric may be a little over the top, but the following results it has achieved are not:

- It has reduced error rates in some processes almost to zero.
- It has saved some organizations not just millions of dollars, but billions—and increased revenues by similar amounts.
- It has greatly diminished late deliveries, reworks, and inspection.
- It has cut cycle time from months to weeks to days to hours.

While Six Sigma has become a familiar methodology for business, migrating mature approaches into new sectors demands careful, deliberate tutoring. Six Sigma may be *the* best methodology for helping the integrated supply chain to meet both current and future challenges of increasing speed, reducing errors, and lowering costs—quickly.

"I don't even know what sigma is, so why do I need six of them?" Simply stated, Six Sigma is a method that uses data and statistical analysis to solve problems, make decisions, and get results.

Sigma is a statistical term, a measurement of standard deviation (variability) around a mean or a goal. Sigma calculates the number of defects likely to occur in a million opportunities for error. So if an organization is 2 Sigmas away from a goal, their number of errors per million opportunities for error is 308,537.

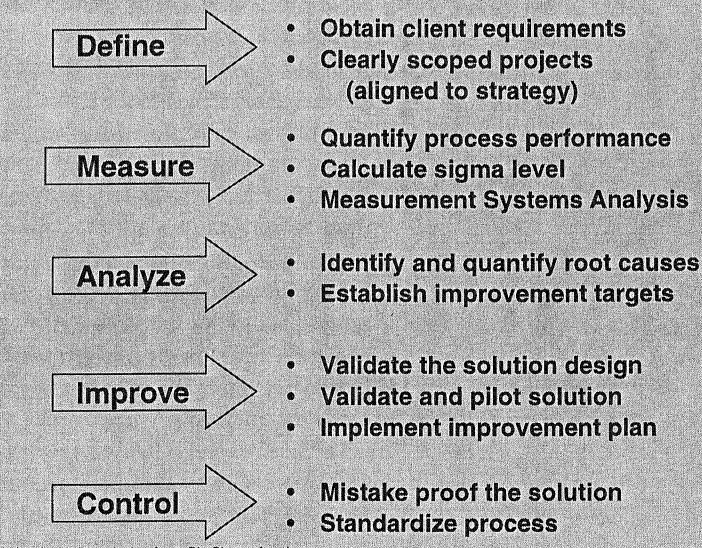
The potential gains (fruit) can be harvested at each sigma level, using various tools where:

- 1–2 Sigma merely harvests "ground fruit," obtaining minimum gains from things like cost reduction projects.
- 3 Sigma harvests the easily obtainable "low hanging" fruit through process improvement.
- 4 Sigma harvests the bulk of the fruit by using the DMAIC methodology (Determine-Measure-Analyze-Improve-Control).
- 5–6 Sigma harvests the sweetest fruit of all, with a combination of DFSS (Design for Six Sigma) and DMAIC, typically breakthrough projects with large payoffs.

While everyone would like to be at the 6 Sigma level, only a very, very few processes or corporations are operating at a 6 Sigma level today. 6 Sigma is usually a lofty long-term goal that acts as a driver for continuous improvement.

That underlying concept is not difficult to understand. However, the methodology, structure, and processes behind Six Sigma are not that simple or easy. The methodology is demanding, rigorous, and disciplined. Six Sigma requires extensive training, with actual projects intermixed

EXHIBIT 3 DMAIC Methodology



Used with Permission from Six Sigma Academy

during the training, and it requires an almost fanatical focus on implementation and results in a short period of time.

There are many myths and misconceptions about Six Sigma. Some think Six Sigma is just "warmed over" TQM, reengineering, or performance management. While Six Sigma may indeed use these tools, it does not preselect any tool. It applies whatever tool fits the problem and produces results. Many believe that Six Sigma is only for manufacturing organizations. Not so. It is also being used by banking, insurance, healthcare, retail, pharmaceuticals, defense, oil and gas, and *should be* used in education.

Still others see Six Sigma as just another fad that will soon fade. Maybe it will, but right now it is far from just an empty fad. There are documented large cost savings and huge revenue increases from over 100 corporations. If it's a fad, it is paying off big time in hard savings—not vapor or soft savings.

For example, GE Plastics conducted 3,600 Six Sigma projects in 2001 and slashed annualized costs by more than $300 million. In 2001, ITT Industries saved about $135 million from its value-based Six Sigma program with 1,018 projects. Dow Chemical has average savings of $250,000 per project and determined that Six Sigma could add $1.5 billion in cumulative earnings by 2003. In less than 18 months, Community Hospitals of Central California decreased their supply item costs by more than 19 percent.

While Six Sigma can be used as a stand-alone methodology, many organizations combine or blend Six Sigma with other methodologies, such as benchmarking, change management, lean, breakthrough strategies, knowledge management, and WorkOut.

Six Sigma's Origins and Track Record. Six Sigma got its start in 1979 at a Motorola management meeting. One manager stood up, and said, "The real problem at Motorola is that our quality stinks!" A number of engineers, notably Mikel Harry and Bill Smith, started working to improve Motorola's quality and from that work Six Sigma emerged.

Since then, it has spread to hundreds of firms like General Electric, Allied Signal, Dow Chemical, DuPont, American Express, Citibank, Bank of America, and Caterpillar. Most of these companies report huge gains in cost savings, increased revenues, reduced cycle time, and fewer errors.

Six Sigma—How It Works. If an organization wants to use Six Sigma, the American Productivity and Quality Center strongly recommends that they start with Six Sigma training from personnel inside or outside their organization. The training typically lasts four months. Each of the four months includes one week of training, followed by actual work on a selected, well-defined project. The training and the project work follow a methodology known as DMAIC (usually pronounced "Dee-may-ic"). See Exhibit 3 for a diagram outlining the five stages and steps:

1. *Define.* Determine what needs to improve.
2. *Measure.* Measure the current state against the desired state.
3. *Analyze.* Analyze the root cause of the gap.
4. *Improve.* Brainstorm solutions, select, and implement the best.
5. *Control.* Control and monitor the long-term sustainability.

A companion methodology to DMAIC is DFSS (Design for Six Sigma). The DFSS methodology is used when designing or redesigning a product or service. DMAIC and DFSS use some of the same methodology with DFSS focusing more on design on the front end, followed by many of the same stages as DMAIC. Regardless of whether DMAIC or DFSS is used, everything in Six Sigma becomes a "project," with a designated project team, project goals, timetables, roles and responsibilities, and measures.

Part of the methodology is the designation of people involved as Champions, Master Black Belts, Black Belts, and Green Belts. (A few organizations have White or Pink Belts!)

- Champions are senior level people who oversee projects at a high level, obtain necessary resources, and settle inter-organizational disputes or barriers.
- Master Black Belts advise Black Belts, and act as roving experts and advisers to help project teams.
- Black Belts are those who go through the full DMAIC training, and lead projects. After a number of successful projects, and more training, some are labeled Master Black Belts.
- Green Belts are those who take some or all of the DMAIC training, and are typically members of a team led by Black Belts.

Some managers and educators are sometimes put off by these martial arts terms. The designations are intended to clearly identify and signal Six Sigma competency and commitments to the methodology, similar to the use of such terms in the martial arts.

This short explanation of Six Sigma doesn't do it justice. It is far more rigorous, disciplined, and demanding than these words seem to imply. Ask anyone who has been a member of a Six Sigma project.

Will Six Sigma methodology work just as well in the education testing cycle? Yes.

Six Sigma in the K-12 testing cycle

Here is a very high-level description of how Six Sigma in an integrated supply chain would work. In the set-up phase, senior representatives from each member of the supply chain meet to form an integrated supply chain. They agree to coordinate and work together as though the entire testing process was one continuous process. Each entity in the chain designates a project team. The supply chain members then hold a joint meeting of all the project teams to design and map an overall process supply chain. They also create an overall governing project charter, goals, milestones, and norms. Next, the supply chain members form an overall Steering Committee composed of one or two members from each of the project teams. Each organization also adds one more senior member to the Steering Committee who becomes the Six Sigma "Champions" in their organizations.

THE REDUCTION OF CYCLE TIME, ERRORS, AND WASTE IN AN INTEGRATED SUPPLY CHAIN ARE THE PRIMARY BENEFITS.

In the training phase, Six Sigma DMAIC training for the Steering Committee and each of the project teams begins. The participants move through DMAIC methodology, stage by stage, with a project goal of completing individual and joint projects in four months. Training includes DFSS when projects need to design or redesign processes, products, or services.

The Steering Committee holds frequent meetings to monitor progress, asks entities to provide progress reports of their DMAIC training and project progress, and sets aggressive goals to drive change in areas of greatest need for improvement:

- *Delays.* In the first year, cut the testing cycle time so that accurately scored tests are returned to districts before the end of the current school year in which the test is given. In subsequent years, reduce that cycle to one month, to weeks, to days.
- *Errors.* Make the "defective" rate zero. Reduce the "defect" rate to 3 Sigma in two years. Aim for an eventual 6 Sigma defect rate of 3.4 errors per million opportunities for error by 2015.
- *Costs.* Reduce costs of testing in each of the entities in the supply chain by 25 percent in each of the next three years.
- *Overall Goals.* Cheaper *and* Faster *and* Better (not "or").

Benefits

Will the investment of time and money be worth it? The answer is absolutely "yes" from most every

business organization that has deliberately initiated Six Sigma. The reduction of cycle time, errors, and waste in an integrated supply chain are the primary benefits. Will this work in the education testing cycle? Why not? In addition to the primary benefits of Six Sigma, most business firms have found additional benefits that may someday apply to all members of the K–12 testing supply chain and their education customers:

- Increased customer satisfaction, leading to retention of customers, and obtaining new customers
- Enhanced reputation in the marketplace as a quality organization
- Increased revenue
- Employee satisfaction and pride at "doing it right the first time"
- The ability to use Six Sigma in all parts of their business as an improvement methodology
- Early return of investment—they don't have to wait years for gains. The "low-hanging fruit" is obtained quickly, which often pays back the investment, with much larger breakthrough gains as the organizations move to higher sigma levels

Expect objections

Leaders who attempt this integrated supply chain approach using Six Sigma can expect the usual resistant remarks:

- "Here comes another fad, another program of the month."
- "It's a business tool—it doesn't apply to education."
- "It's just a bunch of statistics. We're dealing with kids here."
- "All that stuff about Black and Green belts is silly."
- "We may have a few errors, but we eventually fix them."
- "We're too busy right now; maybe later."
- "The faults lie somewhere else—not in our organization."

Leaders should:

- Acknowledge the legitimacy of those feelings.
- Answer them the best they can.
- Read some of the Six Sigma books and articles.
- Talk to those who have used Six Sigma.
- *Continue with Six Sigma implementation!*

The leadership team should make it abundantly clear that they are going to implement Six Sigma—regardless of the doubts of others—and actually follow through with it with no reservations. This is why it's so very important to not only get senior management's commitment, but also their involvement.

Dan Burnham, Chairman and CEO, Raytheon, put it this way: "Six Sigma takes the passion and obsession of the CEO to make it happen ... it started with me and ends with me."

A wake-up call

The wake-up call for education testing has arrived. It can be heard from No Child Left Behind, dissatisfied customers, rising costs, falling profits, embarrassing errors, lawsuits, and penalties. Education today is where American companies were 20 years ago, when they faced a similar wake-up call from global competition and consumer dissatisfaction. The good news is that no one need respond by creating a new model. It already exists. Just adapt and use integrated supply chain and Six Sigma.

Historian Arnold Toynbee pointed out years ago that when nations became successful after using certain strategies, they became attached to those strategies. However, when a different sort of challenge came along and they used the same, once successful strategies, they failed. The same is true of the military, business organizations, and education.

New and expanded testing requirements can be viewed as a burden, a paper mountain, or an administrative nightmare, *or* as a new challenge and opportunity. The response to the new challenge should not be more of the same, but a *new* response—Six Sigma is a truly integrated system for better, faster, and cheaper results. ■

References:

Mikel J. Harry and Richard Schroeder, *Six Sigma* (New York: Doubleday, 2000).

Greg Brue, *Six Sigma for Managers* (New York: McGraw-Hill, 2002).

Dick Smith and Jerry Blakeslee, with Richard Koonce, *Strategic Six Sigma* (New York: John Wiley & Sons, 2002).

George Eckes, *The Six Sigma Revolution* (New York: John Wiley & Sons, Inc., 2001).