

ATP
**Innovations
in Testing**
Palm Springs, California
2007



www.InnovationsInTesting.org

Continuous Improvement of Testing Processes: Lessons Learned

Cheryl L. Wild, Ph.D.

Wild & Associates, Inc.

Rohit Ramaswamy, Ph.D.

Oriel, Inc.

David Anderson, Ph.D.

Educational Testing Service

Today's Presentation is a Preview

Improving Testing (eds. Wild & Ramaswamy) published by LEA will be coming out in fall 2007

- Chapter authors include: David O. Anderson, Roger Brauer, Charlie DePascale, Rick Dobbs, Lisa Ehrlich, David Foster, Sharon Goldsmith, Ronald Hambleton, Ning Han, Stuart Kahl, Joan Knapp, Peter Kronvall, Dennis Maynes, Rohit Ramaswamy, Marten Roorda, Michael Rosenfeld, Richard Smith, Judson H. Turner, Cheryl L. Wild, Lonnie Wright

Secretary of Education Margaret Spellings, 2006

- “Errors not only hurt those directly involved—they hurt the entire testing industry by undermining public confidence”

Quoted from the 2006 *ATP Newsletter* article “From the Desk of the Education Division Chair”

Problems Common to Most Testing Companies

- Demands for new technologies for test registration, test assembly and test administration
- Demands for more rapid test development and score reporting
- Demands for increased customization
- Haphazard and insular development of processes within a company

Continually Improving Testing

- Is more than psychometrics
- It involves using quality oriented process management tools along with psychometric principles to manage product and service quality

Quality in Testing

- In order to reach the goal of no defects
 - Identify and study your processes
 - Measure the defects
 - Make informed improvements
 - Monitor the results

Quality in Testing

- Overall goal is to meet the needs of the customer
- Quality can be defined as
QUALITY = NO ERRORS or DEFECTS

Quality in Testing

- Defect
 - when a business requirement is not met
- May relate to
 - Timeliness
 - Accuracy
 - Cost

Our Quality Model

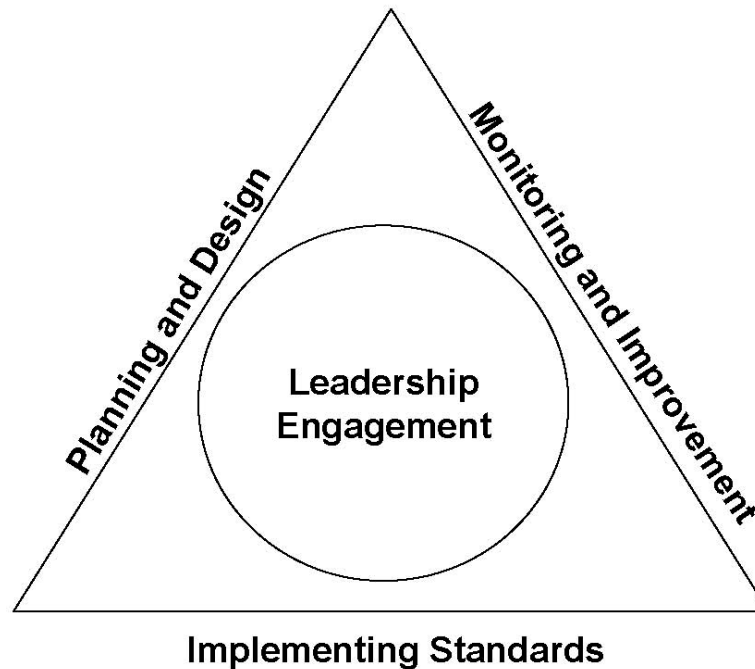


Figure 1: The Quality Triangle

From *Improving Testing (in press, 2007)*
Edited by Cheryl L. Wild & Rohit Ramaswamy

ATP
**Innovations
in Testing**
Palm Springs, California
2007



www.InnovationsInTesting.org

Continuous Improvement of Testing Processes: Lessons Learned on Standards

Cheryl L. Wild, Ph.D.
Wild & Associates, Inc.
Joan Knapp, Ph.D.

Knapp & Associates International, Inc.

Overview



Figure 1: The Quality Triangle



How do Standards Relate to Quality?

- Quality can be defined as
 QUALITY = NO ERRORS or DEFECTS
- Defect
 - when a business requirement is not met
- A standard is a business requirement

Standard Dichotomies

- Internal vs. external
- Process oriented vs. product oriented
- Guidelines vs. requirements

What are standards?

“Documents established by consensus and approved by a recognized body that provides for common and repeated use of rules, guidelines or characteristics for activities or their results aimed at achieving the optimum degree or order”

(ISO/IEC Guide 2 (1996), p. 10 under clause 3.2)

Three Groups of Standards in Use in the Testing Industry

- Testing Standards without third-party review (guidelines)
- Testing Standards with third-party review (usually called accreditation)
- Generic Quality and Business Standards with third-party review

Who Creates Testing Standards/Guidelines?

- American Educational Research Association
- American Psychological Association
- Association for Assessment in Counseling and Education
- Association of Test Publishers
- Council on Licensure, Enforcement, and Regulation
- European Federation of Psychologists' Associations
- IMS Global Learning Consortium
- International Test Commission
- Joint Committee on Testing Practices
- National Council of Measurement in Education
- Society for Industrial and Organizational Psychology

Sample of Testing Standards

- European Federation of Psychologists' Association Review Model for the Description and Evaluation of Psychological Tests
- IMS Question and Test Interoperability Overview: Version 2.1 Public Draft Specifications
- Standards for Educational and Psychological Testing

Testing Standards with Third-Party Review

- Specialty Related Accreditation Standards
 - American Board of Medical Specialties
 - American Board of Nursing Specialties
 - Association of Real Estate Licensure Law Officials
 - Conference for Food Protection
 - Council of Engineering and Scientific Specialty Boards
- International Standard ISO/IEC 17024 Conformity Assessment—General Requirements for Bodies Operating Certification of Persons
- National Commission for Certifying Agencies Standards for the Accreditation of Certification Programs
- Standards for BIACO Accreditation of Proprietary Testing Programs

Business Standards

- ANSI/ISO/ASQ Q9001-2000 American National Standard: Quality Management Systems – Requirements
- CMMI® Maturity Model Integration
- ISO/IEC 27011:2005 International Standard: Information Technology – Security Techniques– Information Security Management Systems – Requirements
- Suitable Trust Services (SysTrust™ and Web Trust™)

Best Practice Testing Organizations

- See internal audits and third-party review as “a cost of doing business”
- Are using business process standards as well as psychometric standards to improve work processes
- Use internal metrics/standards for self-assessment

Benefits of Accreditation

NOCA Survey Results Indicate:

- Benefits of accreditation include continued internal quality improvement, enhanced reputation of the program, and defense of the program's integrity in the event of a legal challenge.
- 50% felt accreditation important for compliance with third party organizations

Benefits of Third-Party Reviews: ISO 9000 and CMM

- Review of ISO 9000 and CMM certification found that “quality certification helps software firms to attain higher operational performance in the form of better product attributes and higher returns on quality”
- Comparison of ISO 9000 certified firms versus a control group of non-certified firms show that “firms experience significantly better performance after deciding to seek their first ISO 9000 certification than a control group of firms with similar performance prior to that decision”

Is the Environment Changing for Testing Organizations?

- In 2006 Senator La Valle of New York filed legislation for the creation of a board to oversee standardized testing
- An Education Sector report (2006) recommend an independent national testing oversight agency to independently audit state testing programs and the testing industry

The Future of Standards in the Testing Industry

- Combination of psychometric and process standards
- Standards developed jointly with vendors, sponsors and consumers
- Combination of self-assessment and periodic external audit

References

1. Corbett, C. J., Montes, M. J., Kirsh, D. A., & Alvarez-Gil, M. J. (2002). Does ISO 9000 certification pay? *ISO Management Systems*, (July - August), 31-40.
2. Isaac, G., Rajendran, C., & Anantharaman, R. N. (2004). Significance of quality certification: The case of the software industry in India. *Quality Management Journal*, 11(1), 8-32.
3. National Organization for Competency Assurance (2005, Summer/Fall). *The value of NCCA accreditation Survey*. Unpublished manuscript. Excerpts available in *NOCA News*, Washington, DC: Author.
4. Toch, T. (2006). *Margins of error: The education testing industry in the no child left behind era*. Retrieved February 11, 2006 from www.educationsector.org/usr_doc/Margins_of_Error.pdf
5. Wild, C. L. & Knapp, J. E. (In press). Standards in the testing industry. *Improving Testing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

ATP
**Innovations
in Testing**
Palm Springs, California
2007



www.InnovationsInTesting.org

Continuous Improvement of Testing Processes: Lessons Learned about Planning and Design

Rohit Ramaswamy

Oriel, Inc.

Planning and Design

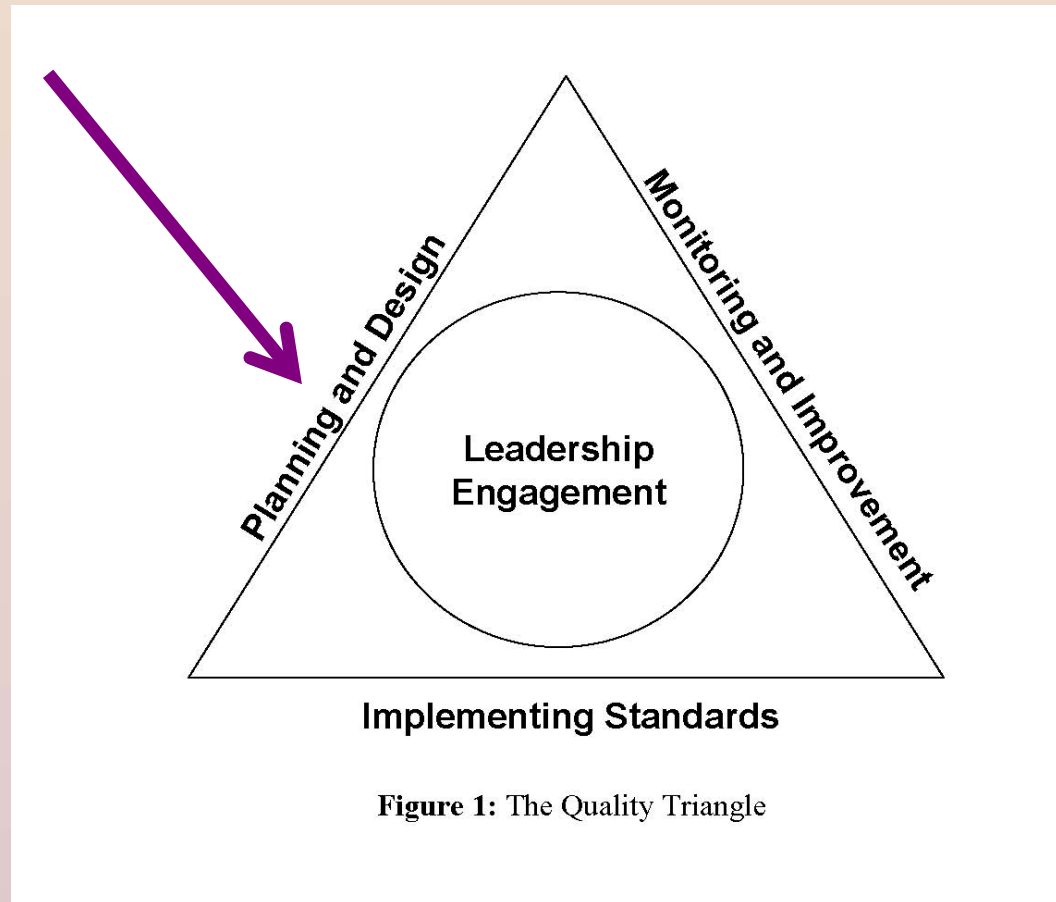


Figure 1: The Quality Triangle

How to Reduce Testing Errors ?

- Create processes that reduce the possibility of errors
- Enhance inspection and correction procedures
- Many testing organizations emphasize correction over prevention
- Overall, more effective to eliminate errors
- Important to consider trade-off between elimination and inspection

Error Definition - Review

- Defect or Errors
 - when a business requirement is not met
- May relate to
 - Timeliness
 - Accuracy
 - Cost

Balancing Design and Correction - Approach

- Identify errors that have the greatest impact on customers and the business
- Focus organizational and technological resources on eliminating these errors
- Develop inspection and correction systems for the less serious errors

Balancing Design and Correction – Steps

- Understand customer requirements
- Understand sources of errors
- Identify possible errors
- Assess customer and business risk from errors
- Develop methods to address errors

Step 1 - Customer Requirements

Two parts to Customer Requirements:

1. “Voice of the Customer” - Customer verbatim statements
2. “Customer Specifications” – Customer statements translated into measurable

Customer Requirements - Example

- Voice of the Customer:
 - “I want the diagrams in the test to be clear and easy to read”*
- Customer Specifications:
 - *color scheme*
 - *brightness*
 - *line thickness*
 - *screen resolution*
 - *pixel density*

Who are Customers ?

- End users (test takers)
- Test sponsors (state governments, school districts, or professional organizations)
- Regulatory bodies (accrediting bodies, government)
- Test administrators (principals and teachers, training departments of organizations),
- Senior management of the testing company
- Organizations that are responsible for the downstream processes (e.g., the scoring and reporting organizations are customers of the test development organization)
- Some combination of the above.

Collecting VOC data

- Reactive Techniques

- *customer complaints*
- *customer feedback after item or test reviews, calls or emails from potential customers into sales or customer service*
- *completed feedback or satisfaction forms*

- Proactive Techniques

- *free form interview*
- *focused elaboration*
- *most significant moments*
- *appreciative Inquiry*
- *contextual Inquiry*

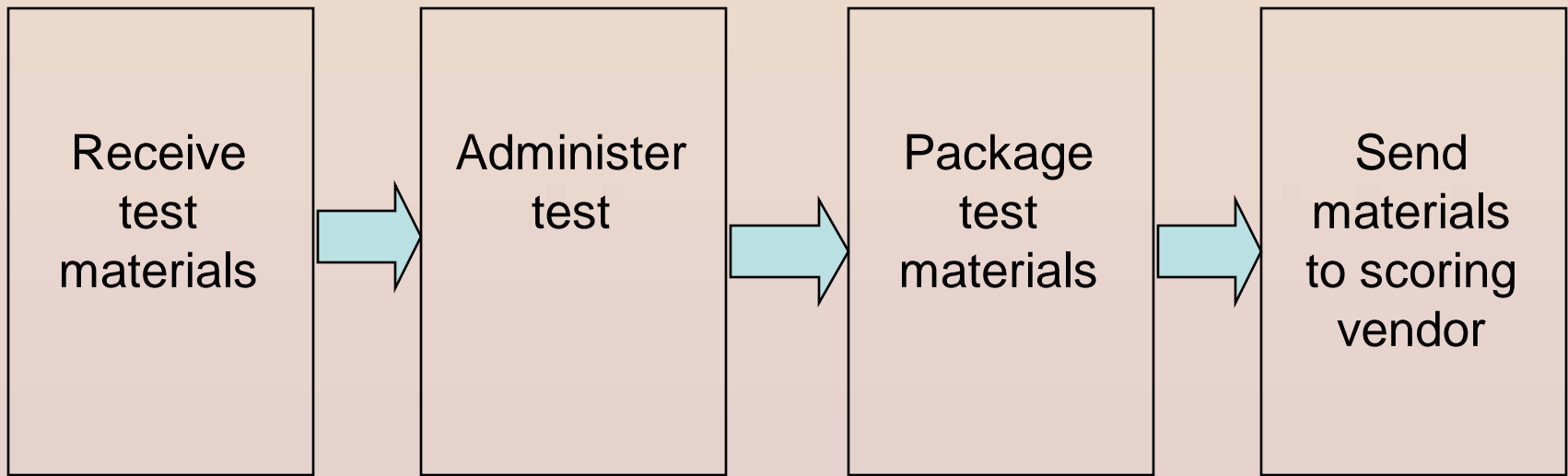
Step 2 – Sources of Errors

- Document key processes for the enterprise
- Include “Core” and “Enabling” processes
 - Core processes deal with the creation, administration and scoring of tests
 - Enabling processes are those such as sales, marketing, human resources or finance, which support the core activities of the organization.

Core and Enabling Process Examples

Process Name	Process Type
Author management	Enabling
Item development	Core
Test creation	Core
Test administration	Core
Test scoring	Core
Report generation and distribution	Core
Item bank administration	Enabling
Test security administration	Enabling
Sales	Enabling
Employee recruiting and support	Enabling

Process Documentation - Example



Test administration process

Step 3 - Identify Possible Error Sources

- Errors are made in process steps
- Review process maps to brainstorm error sources
- Tie error identification EXPLICITLY to processes

Examples of errors:

- Failure to schedule shipping
- Failure to deliver or failure of shipping company to pick up
- Delays in the shipping process
- Failure to log receipt of forms
- Materials damaged in transit

Step 4 - Assess Customer and Business Risk

- Identify the customer requirement affected by the error
- Assess the importance of the customer requirement
- Determine how likely the error is to occur
- Multiply the importance by the likelihood to get a composite score

Assessing Risk: Example

Error	Associated Process	Associated Reqmnt.	Reqmnt. Importance	Prob. Rating	Composite Risk Score
Failure to schedule shipping	Test Admin.	Return Timeliness	4.8	2	9.6
Failure to deliver or pick up	Test Admin.	Return Timeliness	4.8	3	14.4
Shipping delay	Test Admin.	Return Timeliness	4.8	4	19.2
Failure in logging receipt	Test Admin.	Return Timeliness	4.8	2	9.6

Other Risk Considerations

- Whether the error, however rare, will result in catastrophic failure of one or more processes with consequences on reputation;
- Whether the error needs to be mitigated as a legal or governmental requirement;
- Whether correction of the error, if it occurs, will take significant time, cost or resources to fix.

Step 5 – Address Errors

- Reduce the *opportunities* for errors;
- Control the *occurrence* of errors.
- If neither of these approaches is feasible then look for ways to *detect* errors.

Reducing Opportunities for Error

- Automating steps that are performed manually.
- Combining steps so that errors in hand-offs between steps are eliminated.
- Reducing decision points by streamlining processes so that mistakes in decision making do not occur.
- If there is enough volume, creating separate processes for different products to minimize having multiple paths in a single process (e.g. having a separate process for creating paper-and-pencil and computer based tests).
- Reducing the number of steps overall in *any* process, through simplification.

Reducing Opportunities for Error

- Standardizing processes and work activities and clearly documenting how work flows through the process.
- Providing checklists, templates and other job aids that facilitate the doing of work in a consistent and logical manner.
- Color coding exceptions, expedites and other variations to the normal process, so that these are clearly recognizable as they are being processed.
- Incorporating simple auto-check procedures immediately after the task has been completed so that errors can be locally corrected.
- Implementing auto-detect and auto-correct technology using logic filters wherever possible.
- Introducing just-in-time flow techniques so that work flow is balanced through the process.
- Establishing the optimal work speed for systematic on-time task completion and staffing the processes to support this work speed.
- Using “forcing functions” so that certain tasks cannot be started until prior tasks are correctly completed (e.g. correct completion of a test review step will result in the issue of a password that needs to be entered into the system to move to the next process step).
- Using auto population of standard information where possible.

Improving Error Detection

- Ensure that the errors can be detected as soon as possible, and especially before they become visible to the customer.
- Detect errors in as few steps as possible after they occur.
- Have no more than one or two major inspection points in a process.
- Engineer mechanisms for early error detection

Conclusions

- It is more cost effective and less risky to try to create processes where errors do not occur rather than developing extensive inspection procedures to detect and correct errors as they occur.
- An “error” or a “defect” corresponds to conditions where customer or business requirements are not met.
- Processes must be designed based on the “voice of the customer” translated into performance requirements.
- A systematic procedure for identifying areas of risk and vulnerability in processes helps to prioritize areas of focus for error prevention.

Good References

- Ramaswamy, Rohit (1996) *Design and Management of Service Processes*. Prentice-Hall
- Chowdhury, Subir. (2003) *Design for Six Sigma*. Dearborn Trade.

ATP
**Innovations
in Testing**
Palm Springs, California
2007



www.InnovationsInTesting.org

Continuous Improvement of Testing Processes: Six Sigma David O. Anderson, Ph.D. Educational Testing Service

Overview

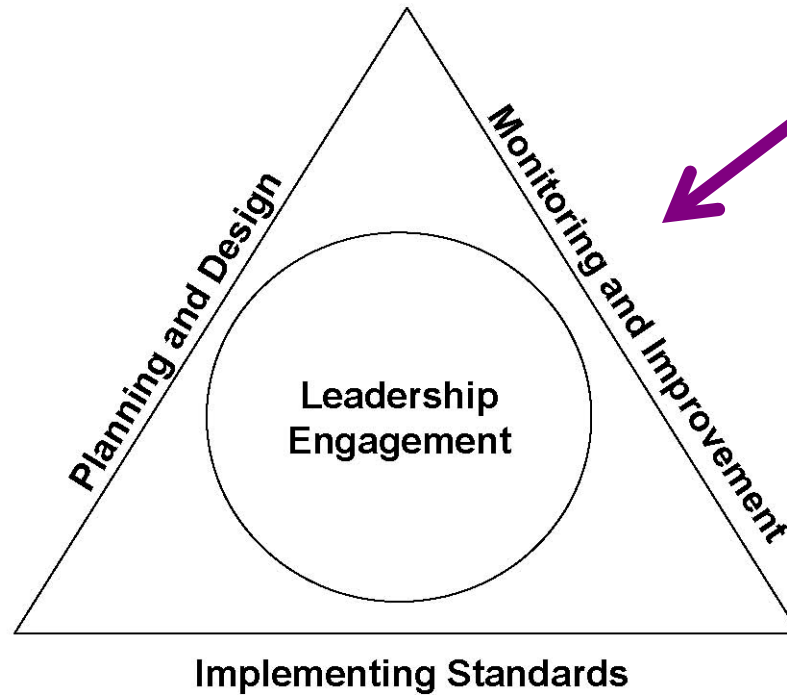


Figure 1: The Quality Triangle

Quality in Testing

- Defect
 - **when a business requirement is not met**
- May relate to
 - Timeliness
 - Accuracy
 - Cost

Brief Overview of ETS

- 2,500 employees around the world
- Wide variety of testing programs
- Variety of responsibilities
 - Test Development
 - Test Administration
 - Test Scoring
 - Test Reporting

Some Functions within ETS

- Test Development
- Statistical Analysis / Psychometrics
- Client Relations / Administration
- Operations
 - Call Center, Scanning, Resolutions, Shipping
- Publications
- Finance, Legal, R&D

Overview of Six Sigma

- Disciplined problem-solving approach to process management
- Data-driven approach to monitoring process outputs

Overview of Six Sigma

- Originated in manufacturing
 - Motorola and General Electric
- High volume, standardized processes
- Goal was to eliminate waste by achieving near-perfect results
- Has moved to the service industry

Key Concepts

- **Cost of Poor Quality**
- **Voice of the Customer**
- **Defects per Million Opportunities**

Cost of Poor Quality

- Beyond the cost of finding and fixing the defect, some indirect costs include:
 - Failure to meet customer expectations
 - Increased cycle time
 - Over-production
 - Extra costs for expedited delivery
 - Excessive delays

Voice of the Customer

- The customer defines quality
- Customers are many:
 - Test sponsors
 - Test administrators
 - Test-takers
 - Score recipients
 - Essay/CR readers
 - Legislators

Voice of the Customer

- Internal departments are customers, too.
- Even within a company, there are different expectations across departments
- Ill-defined handoffs between areas are the source of many delays and errors

Defects per Million Opportunities

- Six Sigma strives for reduce defects to 3.4 per million (DPMO)
- Aims for 99.99966% accuracy
 - 99% accuracy is only 3.8 Sigma
- From 5,000 misplaced answer sheets to 1-2 per year (per million)
- From 10,000 incorrect scores to 3-4 per year (per million)

A Laudable Goal

“All scores reported on-time”

But what about a/s in resolution, a/s from special forms, or a/s not even received?

It soon gets very complex.

Defects per Million Opportunities

- The only way to reduce defects is to
 - study your processes in detail,
 - quantify the results of each process,
 - monitor those results
 - make continuous improvements
- Six Sigma methodology makes this possible

The DMAIC Process

- **DEFINE**
- **MEASURE**
- **ANALYZE**
- **IMPROVE**
- **CONTROL**

The DMAIC Process

- **DEFINE** the problem
- Choose the ones with most cost-benefit if resolved
- Identity and focus on the processes involved in those problems

The DMAIC Process

- **MEASURE** the extent of the problem
- Calculate frequency of defects/errors and compare to total opportunities
- A problem might be the variability across test administrations
- Use this DPMO to measure progress

The DMAIC Process

- **ANALYZE** the data to find the cause of the problem
- Processes have inputs, outputs, and sub-processes
- Collect metrics at strategic points in the process
- Drill down to find the root causes

The DMAIC Process

- **IMPROVE** the process
- Generate and evaluate possible solutions
- Select solutions that result in greatest benefits

The DMAIC Process

- **CONTROL** the new process
- Take frequent measurements to ensure consistency and to track progress
- Actively monitor the metrics

Examples of 6S Projects

- **Inaccurate Forecasting of Testing Volumes**
- Bad Payments for Services Rendered
- Over-production of Test Bulletins
- Rejected Honoraria Payment Vouchers
- Excessive Use of Warehouse Storage
- **Answer Sheet Rejections**

Inaccurate Forecasting of Testing Volumes

- One program had forecast variances up to 10% from actual volume
 - \$4.5 million problem
 - Last-minute “walk-ins” not included
 - “Gut-feeling” increases or decreases
- Team used multiple regression approach
 - 8 years worth of data of reg. and actual
 - Daily counts for 63 days before admin

Inaccurate Forecasting of Testing Volumes

- Reduced variance to 5% 8 weeks prior to registration deadline
- Now a data-driven forecast

Answer Sheet Rejections

- In one program, 5% of a/s reject at scanning
- Found that 33% rejected due to incorrect input of test taker names
- Also a/s processed before registration
- By changing instructions on a/s and changing process flow, reduced labor costs by \$142,000 annually

Conclusion, Lessons Learned

- Not completely problem-free
- However, documented savings of over \$10M
- Ongoing schedule for training new Six Sigma candidates
- 6S tools are being used in variety of areas across ETS
- More decisions are data-driven

Good References

- www.iSixSigma.com
- www.apqc.org
- www.bptrends.com
- www.bpmenterprise.com
- *iSixSigma magazine*
- Chowdhury, Subir. (2001) *The Power of Six Sigma*. Dearborn Trade.

Contact Information

David O. Anderson, Ph.D.

doanderson@ets.org

Rohit Ramaswamy, Ph.D.

rohitr@worldnet.att.net

Cheryl L. Wild, Ph.D.

cheryl@wildandassociates.com

Questions?

Bad Payments for Services Rendered

- We require immediate payment for test registration or additional score reports
- 89% of bad payments were from credit-card declines
- Further analysis disclosed that only 5% of CC processing was real-time, the rest were batch-processed at night
- Services provided before payment verified

Bad Payments for Services Rendered

- Team focused on domestic sales for one testing program with most problems
- Implemented real-time CC processing
- Resulted in \$1M improved cash flow
- Once expanded to other programs, expect over \$2.4 annual savings

Over-production of Test Bulletins

- For two programs, 8:1 and 5:1 ratio of printed bulletins to projected registrants
- Caused reduced product margins; increased costs for printing, shipping, storage and disposal, inventory management
- 34% of unshipped bulletins destroyed in-house.

Over-production of Test Bulletins

- Among causes
 - Web-registration trends not considered
 - Advanced canvassing of school volumes discontinued
 - Regionally misdirected shipments
 - Unnecessary reprints
 - Initial distribution over-shipments

Over-production of Test Bulletins

- Multi-faceted phased-in approach
 - Better utilize regional volume projections
 - Aggressively promote web downloads
 - Revise bulletins to extend shelf-life
 - Implement just-in-time inventory methods
- One program reduced bulletin quantities by 75% and saved \$700,000 in print costs alone

Rejected Honoraria Payment Vouchers

- Honoraria for test administrators
- In 1 recent year, 76% were rejected
- Team found 68 points on voucher form causing defects; 2 of which caused 64% of rejects
- Poor voucher design and limitations of financial software were ID'd as root causes
- Redesign has saved \$60,000 to date

Excessive Use of Warehouse Storage

- In 2001, warehouse storage exceeded 99% of capacity
- One root cause was lack of a policy for automatic destruction of materials after extended period of inactivity.
- Spending over \$600,000 annually for materials inactive for 2 years or more

Excessive Use of Warehouse Storage

- Implemented new warehouse policy
- Reduced warehouse usage to 86%
- Eliminated need for off-site storage building, first time in 20 years
- Reduced off-site storage and transportation costs by \$185,000

Example Projects

Customer Service Area

- Variation in handling time by customer service representatives
- Large numbers of repeat calls
- Inefficient workflow for handling customer complaints
- Low up-sell conversion rate
- Poor and incomplete information collection
- Inaccurate call volume forecasting & staff scheduling
- Unacceptably high level of abandoned calls
- Slow response to email inquiries

Example Projects

Test Administration Area

- Inaccurate testing volume projections & test center scheduling
- Expensive variation in candidate bulletin content across testing programs
- Excessive cycle time to process requests for testing accommodations
- Excessive costs for expedited shipping to candidates and/or test centers
- Over-staffing of test centers
- Non-optimal use of test center space
- Late or incorrect delivery of materials to test centers
- Excessive variation in production cycle times (test booklets, study guides, bulletins, etc.)

Example Projects

Test Development, Scoring and Reporting Areas

- Large proportion of newly written test questions fail to meet psychometric specifications
- Inaccurate forecasting of required numbers of readers at CR/Essay readings
- Late delivery of score reports
- Inaccurate results on candidate score reports or state summary reports
- Excessive time resolving registration and answer sheet scanning problems
- Excessive cycle time for Standard Setting (Cut Score Setting) report production

Example Projects

Finance Area

- Variable profit margins from contracts
- Excessive costs for test administration services
- Unacceptable delays in paying honoraria
- High proportion of errors on honoraria payments
- Excessive number of bad checks received
- Excessive number of customer refund requests
- Excessive cycle times to handle purchase orders
- High volume of credit card declines
- Noncompliance to corporate travel policy

Example Projects

Other Areas

- Perceived poor use of administrative and secretarial support staff
- Excessive inventory costs and inefficient use of warehouse space
- Non-standard and inefficient proposal writing process
- Non-standard methods to cost out proposed new work
- Delayed removal and reuse of surplus property from work areas