FSTC REPORT: OCTOBER 31, 2005

Image Quality and Usability Assurance

FINANCIAL SERVICES TECHNOLOGY CONSORTIUM

Where business and technology leaders build solutions for the financial services industry

44 Wall Street
NY, NY 10015
(212) 461-7116
www.FSTC.org

CONTENTS

EXECUTIVE SUMMARY

Phase two of the Financial Services Technology Consortium (FSTC) Image Quality and Usability Assurance (IQ&U) project is now complete. For the first time ever, the financial services industry can now make informed choices to ensure that every check image, regardless of its capture point or method, meets the industry's minimum requirements.

The Project

FSTC's phase-two IQ&U project team of twelve financial services firms and seventeen solution providers applied the financial services industry's most experienced and knowledgeable talent to the problem of defining minimum requirements for check-image quality. Specifically, our objectives in phase two were to:

1. Define image usability and usability metrics
2. Determine the effectiveness of the phase-one image-quality defect measurements
3. Recommend thresholds at which an image-defect measurement might predict that an image has usability issues

FSTC accomplished all of these objectives.

Through rigorous testing and analysis, not only did the team make unprecedented strides in assessing how effectively phase one's image-quality defect metrics predicted image usability, the team set preliminary thresholds for image exchange and diagnostic monitoring.

Key Findings and Insights

Phase two of the project provided FSTC and the financial services industries with an array of new information and insights into the roles of check imaging and image usability.

First, the project provided groundbreaking scientific data about which image-defect metrics provide the most value. Based on our automated testing of about 3.5 million live items, human review of approximately 700,000 predominately black-and-white images, and detailed statistical analysis of the resulting data, FSTC determined the following for black-and-white images:

- Our Image Too Dark metric was the best predictor of image-usability issues.
- Our Image Too Light metric can be a very precise predictor of image-usability issues.
- Most of our other defect metrics have limited to no ability to predict which individual images will have usability issues. However, these metrics can be useful for monitoring system health, identifying system malfunctions, and ensuring overall image quality.

Second, the project team determined that check-image usability is a function of the usability of individual check fields. We identified the check fields that were most important for usability testing and then defined usability metrics for testing them. In addition, we developed usability guidelines that financial institutions can use to design their own usability-testing programs because usability testing will be key for financial institutions that wish to minimize the risk of unusable images.

Thanks to FSTC's extensive analyses, the financial industry now has scientific data that for the first time:

- Documents the usability issues of a large sample of check images
- Enables financial institutions to factor usability issues into their daily production
- Enables financial institutions to understand the true probability of creating and exchanging images with usability issues, which helps them design effective programs to test image quality and usability
- Identifies the relationships between image-quality defect metrics and usability issues

Our project uncovered strengths and weaknesses of image-quality defect testing that have industry-wide significance. In particular, FSTC found that:

- Defect testing alone will not capture most of the images with usability issues.
- Defect testing cannot detect all image-usability issue causes. Detecting some causes, such as faint text, requires testing beyond the image defects identified in phase one.
- Defect metrics cannot identify all images with significant usability issues and also provide reasonable suspect rates.

These findings and insights, as well as our new understanding of the roles and limits of image-quality defect testing, formed the foundation of our recommendations to the financial services industry.

Key Recommendations

Based on our research results, FSTC makes the following recommendations for the use of image-quality defect metrics:

- Fifteen of our original sixteen metrics (all except the metric Carbon Strip Detected) should be incorporated into a general image-quality assurance program.
- Two metrics, Image Too Dark and Image Too Light, should be applied to individual images in image-exchange environments.
- Suspect-review programs should be tailored to the predictive value of the individual metrics.
- The initial threshold values specified in this report should be adopted for the metrics that showed ability to predict usability issues. These threshold values represent what the project team found to be an optimal balance among suspect rates resulting from the metrics testing, the predictive ability of the metrics, and the purpose for which the metrics will be used.

FSTC also recommends that the financial industry take the following steps:

- Use the defect metrics as the foundation of image-quality defect testing
- Collect actual performance data using the recommended defect-metric thresholds; continue to discuss image quality and usability test performance with an industry view; and periodically (perhaps annually) adjust the recommended defect-metric thresholds based on shared experience
- Continue to analyze the defect-metric data collected to identify and validate:

  1. Metric values and combinations that effectively identify images with progressively poor usability scores

  2. Metric combinations that reduce suspect rates while increasing metric precision
- Incorporate the fifteen recommended image-quality defect metrics into the official registry of image-defect tests
- Incorporate appropriate revisions into exchange agreements and Electronic Check Clearing House Organization (ECCHO) rules
- Evaluate individual requirements for ensuring image usability based on the capture and escape-rate data collected in this study, and determine whether image-usability testing is an appropriate component of each financial institution's image-quality assurance program

ACKNOWLEDGMENTS

FSTC would like to thank its project members and project management team for their contributions to the completion of the IQ&U project. The following organizations and individuals contributed to the project:

Advanced Document Imaging LLC	Dr. Peter Anderson, Todd Johnsson, Dr. Brad Paxton
BancTec	Steve Goff, Bob Icard, Carl Reitmeier
Bank of America	Matt Calman, Mark Felse, Joe Shuster, Kenneth Trice, Daniel Welch
Bank of New York	Lou Arkenau
Canadian Payments Association	Jude LeClerc
Carreker	Gary Ernst, Harry Hankla, Terry Will
Citigroup	Brian McRae, Wayne Rothstein, Andy Thomas
Deluxe	Angela Hendershott, Thomas Snavely, Linnea Solem, Ralph Stolp
Diebold	Laura Drozda
ECCHO	Phyllis Meyerson
EDS	Walter Jackson
Epson America	Neil Edmonds, Paul Gay
Federal Reserve	Dexter Holt
Fiserv	Senthil Thiagarajan
IBM	Rod Moon, Ravi Prakash
JPMorgan Chase	John Acton, Norman Davis, John Fricke, Jackie Pagan
Mitek Systems	Jan Charvat
NCHA	Tom D'Aquisto, Glenn Wheeler
NCR	Ian Goodall, Stewart Kelland
NetBank	Tracy Clements
Orbograph	David Kliewer
Parascript	Dmitry Gershuny, Kaz Jaszczak
Silver Bullet	Ray Higgins
Unisys	David Concannon, Robert Klein
U.S. Bank	Tim Kent, Lisa Sabrowsky
VectorSGI	Bryan Kahlig
Viewpointe	Cindy Bird, Mike Brant, Chris Carter
Wachovia	Rod Formyduval, Andy Garner, Chris Klapheke
Wells Fargo	Al Hecht, John Kagey, Ann Kirk

FSTC would also like to thank the following organizations that, although not part of the project, assisted the industry by measuring checks to determine the distance from the edge of a check to meaningful data:

- Fidelity IFS
- Southwest Corporate Credit Union

FSTC would also like to thank the following individuals and organizations for particular or extraordinary assistance:

- Viewpointe LLC for the use of their lab infrastructure and their support throughout the testing process, and Fiserv for assistance in creating the needed index and image files
- BancTec, Parascript, Silver Bullet, and Unisys for providing software and support for use in the image-quality defect metrics data collection
- Emily Will and the Check 21 Advisory Committee of the Association of Forensic Document Examiners for assistance in understanding image-related check fraud requirements
- Dick Fraher and Joe Barressi of the Federal Reserve for help in clarifying issues around Check 21 image requirements for accurate representation
- Greg Murphy of Carreker and Michelle Walkley of EDS for assistance in creating IRDs from the accuracy deck developed during the project
- David Concannon of Unisys and Stewart Kelland of NCR for capturing images of IRDs created from the accuracy deck
- Ralph Stolp and Deluxe for creating an edge measurement tool
- Stewart Kelland of NCR for working with ADI to incorporate the usability score into the analysis
- About thirty different individuals at ten financial institutions for their assistance in capturing edge-defect measurements

The FSTC project management team for this project was:

John Fricke	JPMorgan Chase	Project Director
Frank Jaffe	MorSecure	Project Manager
Rebecca Wetzel	Interisle Consulting Group	Lead Author Project Co-Manager
Keri Toolan	MorSecure	Project Administrator
Rudy Gabrielson	MorSecure	Edge-Defect Measurement Project Manager
Mike Salzman	Inlite Research	Accuracy Deck Lead

PROJECT BACKGROUND AND HISTORY

In March of 2004, FSTC launched the IQ&U initiative to help the industry ensure that any check image, regardless of its capture point, meets the industry's minimal requirements. FSTC's ultimate aim was to define image-quality standards that the entire financial services industry will adopt.

Phase one of the initiative focused on understanding the effects of poor quality and defining check-image quality defect metrics. FSTC successfully completed this initial phase in July of 2004. FSTC completed phase two, which began in November of 2004, in October of 2005. Phase two focused on validating the image-quality defect metrics defined in phase one, establishing thresholds to identify images with potential usability problems, and defining usability and usability metrics. Fifteen financial services firms and twelve solution providers participated in phase one, and twelve financial services firms and seventeen solution providers took part in phase two.

FSTC first recognized the potential need for check-image quality standards during the Paperless Automated Check Exchange and Settlement (PACES) project, launched in late 1997 as a collaborative effort led by FSTC with the Image Archive Forum, ECCHO, the Clearing House Association of the Southwest (formerly CHAS, now the National Clearinghouse Association Regional Exchange or NCHA), the New York City Clearinghouse (now The Clearing House), banks, vendors, and the Federal Reserve System.

FSTC members created the IQ&U initiative in response to the imminent mandates of the Check Clearing for the 21st Century Act (Check 21), which was implemented in October of 2004. With the advent of Check 21, the U.S. banking industry is quickly approaching adoption of check-image exchange and check-image capture at centralized as well as merchant, teller, and ATM locations. Financial institutions face two types of risk due to defective or "untransactable" check images: liability risk and operational risk.

Banks incur liability risk because they must ensure that an image is an accurate representation of the original check, as required by Check 21 and image-exchange rules, including Clearing House rules and Federal Reserve Bank regulations and operating circulars. Financial institutions face operational risk when low-quality images require extra handling and hamper banking functions or the quality of services delivered to customers.

As image exchange becomes commonplace, financial institutions must ensure that image-quality assessment and assurance capabilities protect them and their customers from risk. Because we expect that physical checks will be destroyed soon after the conclusion of a payment transaction rather than processed through the payment system, image quality becomes paramount.

In the absence of interoperable definitions for image quality and usability, different image-quality assessment solutions currently assign diverse scores to the same image, making it hard for financial institutions to judge whether an image meets common quality and usability expectations. The industry is already experiencing situations in which lack of common scoring methodology is leading to rework, returns, and/or extra suspect review.

As FSTC embarked upon this project, there were no widely accepted industry-level definitions for what makes an image acceptable for processing and payment, nor was there a common language to describe image defects. At the conclusion of phase one, the project team had begun to remedy this situation by developing core requirements for image quality and usability and for defining a set of image-quality metrics and their scales.

In phase two, FSTC used "real world" information to assess whether the image-quality metrics predicted check usability and to set thresholds the industry can use to filter out poor-quality images. This work will lay a critical foundation for industry-wide, interoperable image-quality assurance to reduce risk and increase operational efficiency in image exchange.

PROJECT GOALS AND OBJECTIVES

The overarching objective of IQ&U phase two was to establish minimum requirements for image quality and usability for the financial services industry. With this objective in mind, project participants identified the following five goals at the outset of phase two:

- Validate and calibrate the image-quality defect metrics identified in phase one by conducting experiments with actual check images
- Determine which defects are important for automated image-quality assurance
- Establish thresholds for each defect metric by analyzing the experimental data
- Evaluate combinations of defect measurements that might represent image-quality "signatures," in order to strike an optimal balance between escapes and false positives
- Develop specific, field-level image-usability metrics that accurately determine the usability of an image for a specific purpose

OVERVIEW OF IMAGE-QUALITY DEFECT METRICS

During phase one of the IQ&U project, the project team identified sixteen image-quality defect metrics that could affect the usability of an image for its required business purposes. The project team updated the image-defect metric definitions during phase two of the project, and these updates are contained in the separate "Image Quality Defect Metrics" document. During phase two, the project team examined the ability of these metrics to predict image usability and derived thresholds for those defect metrics that predicted image-usability issues well or could prove useful for monitoring system health to prevent usability problems.

The project team designed each of the image-quality defect metrics to measure a common defect condition quickly so that the metric can be applied in a low- or high-speed environment. All of the metrics address the image as a whole and require no knowledge of the semantic information on the image or the location of specific fields of information.

Figure 1 summarizes the image-quality defect metrics and the standard measurement scale for each metric. The Image Quality Defect Definitions document contains full details for all of the metrics.

Figure 1. Summary of Image-Defect Metrics and Units of Measure

Metric/ Defect	Measure	Image Types
1. Undersize Image	Image size in tenths of inches	Both
2. Folded or Torn Document Corners	Tenths of inches	Both
3. Folded or Torn Document Edges	Tenths of inches	Both
4. Document Framing Error	Extra scan area in tenths of inches	Both
5. Excessive Document Skew	Angle in tenths of degrees	Both
6. Oversize Image	Image size in inches	Both
7. Piggyback Documents	Flag	Both
8. Image Too Light	Pixel percentage	Black & White
	Brightness and contrast	Grayscale
9. Image Too Dark	Pixel percentage	Black & White
	Brightness and contrast	Grayscale
10. Horizontal Streaks in Image	Pixel percentage in scan line	Both
11. Below Minimum Compressed Image Size	Compressed image size in bytes	Both
12. Above Maximum Compressed Image Size	Compressed image size in bytes	Both
13. Excessive "Spot Noise" in Image	Average count of noise spots	Black & White
14. Front-Rear Image Dimension Mismatch	Image size difference	Both
15. Carbon Strip Detected	Flag	Both
16. Image Out of Focus	Pixel gradient	Grayscale

During phase two, the project team tested all sixteen metrics on front and back check images.

ACCURACY ASSURANCE PROCESS

A key activity at the outset of phase two of the IQ&U project was to verify the accuracy and consistency of vendor-supplied software designed to measure the image defects identified in phase one of the project. The goal of accuracy testing was to ensure agreement, or "interoperability," of results across vendors.

To facilitate this interoperability, the project team designed accuracy tests for each metric. The team subjected the test designs to repeated review and revision until it was confident that the tests would achieve the project's goals. Once the test designs were approved, the team consolidated them into an overall testing approach, the centerpiece of which was the creation of a deck of calibration images containing multiple examples of all sixteen image defects.

Figure 2. **Accuracy Test Designs**

	Calibration-Deck Images Measured by:		
Defect Name	Participating Vendor Software	General-Purpose Image Analysis Software	Manual Measurement
Undersize Image, Oversize Image	✓		✓
Folded/Torn Document Corners, Folded/Torn Document Edges	✓		✓
Document Framing Error	✓		✓
Excessive Document Skew	✓	✓	
Piggyback Document	✓	✓	
Image Too Light, Image Too Dark	✓	✓	
Horizontal Streaks Present in Image, Carbon Strip Detected	✓	✓	✓
Below Minimum Compressed Image Size, Above Maximum Compressed Image Size	✓		✓
Excessive "Spot Noise" in Image[1]	✓		
Front-Rear Image Dimension Mismatch	✓		
Image "Out of Focus"	✓		

Team members contributed and/or created calibration images that reflected a range of defects within a defect type. For example, the team created images with various degrees of document skew in both negative and positive orientations, and it tore and/or folded sample paper checks before imaging to show a range of torn edge and corner conditions. Members included images with multiple defects in the deck to explore the effect of compound defects on test results. None of the calibration-deck images contained private or proprietary information.

[1] Artificially generated images were used for calibrating this test.

The next task was to determine "truth" values for sample defects in the calibration deck. To remove as many variables as possible from the truth-testing process, the team clarified ambiguous defect-metric definitions and set consistent test parameters. The aim of these efforts was to ensure consistent test implementation and therefore consistent, interoperable results across vendors.

Five vendors performed automated testing of the calibration deck, and the results of the initial round of testing were shared (anonymously) with the testing vendors as well as the project management team. At least two vendors performed each test in this and subsequent testing rounds. Where results among vendors disagreed significantly, the team manually measured the images to determine the correct values or used general-purpose image analysis software to determine truth values (e.g., for the brightness and contrast components of the Image Too Light or Image Too Dark defects). Additionally, vendors made some independent measurements as a control measure to ensure proper calibration. When differences among vendors' results could be attributed to ambiguities in image-defect definitions, the team revised these definitions for clarity.

FSTC encouraged the vendors to use the information from the initial testing round to calibrate and improve their software implementations to align test results with truth values within a specified margin of error.

FSTC performed a second round of testing involving four vendors to assess the consistency of results, and the team made additional independent measurements to validate truth values. Again, FSTC shared the results (with vendors' identities withheld) with testing vendors and the project management team. The results of the second round of accuracy testing showed marked improvement over the first testing round, however, the project team felt that it could achieve an even higher level of interoperability. To do this, the team further clarified metric definitions, updated the metric-definitions document, and revised implementations to bring results into closer alignment. After these actions were completed, vendors participated in a final round of testing.

The final round of testing by the same four vendors resulted in excellent interoperability of results, with one exception. Some vendors identified horizontal framing errors as streaks but others did not. The project team determined that this and other potential interactions among metric defects did not need to be addressed to meet the project objectives, but the team agreed that the issue should be revisited at the end of the project to determine if further metric-definition clarification was warranted. After testing concluded, the project team agreed that because framing errors and streaks occurred infrequently, there was no need to address potential interaction between them. The team also concluded that it was not necessary to address interactions among defects within the metric definitions.

It is important to note that perfect agreement of results is not feasible, because some check-image attributes are not wholly deterministic. Thus, different algorithms are likely to produce slightly different results. This was most noticeable in measurement of skew.

By the end of the project, a total of eight vendors tested against the calibration deck to optimize the interoperability of their test results. See appendix E for a list of the information reported by vendors during the calibration testing. FSTC believes the large number of vendors involved in the initial testing was a great benefit to the industry.

EDGE-DEFECT TESTING

To help determine thresholds for edge defects, the project team measured a statistically significant sample of production checks to determine how close to a check edge meaningful data is located. Using this data, the project team created a model to predict the probability that a tear or fold will impinge on meaningful data. Live image testing validated this model.

The test required measurement of actual production checks. To accomplish this and ensure confidentiality, ten financial institutions measured their own checks. Bank operations staff made all measurements at their own sites using checks from the daily production work stream.

Manual testing found that data was located near check edges on a high percentage of checks. Unlike image-defect measurements, which are made from the perceived edge of the check and may include areas created by "stubble," edge-defect measurements are made from the true edge of the check. FSTC could not conclude which fields were most likely to be affected by tears or folds, because measurements were not made to the same information fields on each check.

Appendix F includes the results for the manual edge-defect testing. These results, together with information about the frequency and severity of edge-defect occurrences in the test sample, provide rich data for understanding the potential sizes of edge defects that could affect data in preprinted payment-related fields as well as information added to the check through subsequent handling.

Analysis of the automated edge-defect metric test results indicated that only very large defects (one inch or greater) generally impacted any of the eight fields included in the testing. We are not sure if this is due to the potential pre-screening for our sample, or our methodology, or both. See the discussion of detailed results for more information.

LIVE IMAGE TEST PROCESS

The IQ&U project's primary phase-two activity was a live item test to obtain data needed to determine if the image-defect metrics defined in phase one correlated to usability issues in real images. The team analyzed whether the defect metrics were useful predictors of usability issues by assessing the metrics' correlation to image usability.

A Viewpointe-owned secure electronic-archive facility stored the live images. Viewpointe provides check-image archive, retrieval, and exchange services for financial institutions, and the project team extracted approximately 3.5 million images for testing from the normal workflow of eight Viewpointe financial institution members during May 2004[2]. Under subcontract to Viewpointe and the FSTC project, Advanced Document Imaging LLC (ADI) conducted the image testing and results analysis for the project.

FSTC performed all analysis with legal approval from participating financial institutions under tightly controlled and monitored privacy and security.

FSTC and ADI designed the live image test process to sort images into the following categories:

- Images that manual review identified as having usability issues
- Images that defect measurements indicated may have usability issues, and manual review validated the presence of such issues
- Images that defect measurements incorrectly indicated may have usability issues
- Images with usability issues that were not identified by the defect metrics

FSTC and ADI also designed the live image test process to provide a statistically significant sample of images that adequately represented a spectrum of usability issues and image defects.

Definitions

The following are definitions for terms used in the live image testing analysis:

Suspect: An image identified by a defect measurement as possibly having usability issues. A suspect is identified when its measurement value exceeds a threshold.

Suspect Rate: Suspects as a percentage of total sample size.

Capture Rate: The percentage of suspects that have usability issues (from all causes).

False Positive: An image incorrectly identified as a suspect.

Escape: An image with usability issues that is not identified by a defect measurement.

Precision: The percentage of suspects that have usability issues.

[2] The test sample contained black-and-white images from eight banks and grayscale front images from only one bank.

Item Attributes

The bulk of the items tested had black-and-white fronts and grayscale backs. To collect data on black-and-white check backs, the original sample was supplemented to include approximately 100,000 items with black-and-white backs and fronts from one bank. Because black-and-white images are used for exchange, the analysis focused on those, but ADI also collected and analyzed data for grayscale images as well. The black-and-white rear images analyzed did not have grayscale counterparts for direct cross comparison. Black-and-white images in the sample were 200 and 240 dpi (dots per inch), and the grayscale images were either 80 or 100 dpi.

It is important to note that these items were live items originally retrieved from Veiwpointe's archive. As such, some of these items may have previously been subject to some forms of image quality assurance testing. Readers should bear in mind that when testing images "right off the camera", they may find somewhat different rates of occurrence and precision for the defect metrics.

Automated Image-Defect Metric Testing

Four participating vendors used their automated image-quality measurement engines to test the live images for the image-quality defects defined in phase one. Multiple vendors tested each defect to ensure the most robust measurement results possible. The vendors captured test results for each defect metric in an XML format defined for this purpose, and ADI averaged the results from the test engines that yielded the most accurate and consistent output during earlier accuracy testing (see the "Accuracy Assurance Process" section above) to provide a single measurement for each defect-metric element. ADI used these averaged values in its usability prediction analysis. In the case of skew, ADI used measurements from participating vendors with the most accurate truth testing results for various degrees of skew.

Image Prescreening

To investigate the relationship between image-quality defects and usability, it was critical to identify items most likely to exhibit usability issues. To accomplish this, ADI selected and employed usability analysis software to prescreen live images for those most likely to have usability issues (i.e., illegible data) in any of the fields to be subjected to manual review.

In addition to prescreening for usability, some of the financial institutions that made live images available to the project provided filters to remove noncheck images such as deposit slips.

Manual Review Process for Image Fronts

ADI performed two types of manual review: "rapid-fire" manual review and methodical manual review. See appendix G for a graphic of the review workflow. The methods are described below.

Rapid-Fire Manual Review: Prescreening software flagged approximately 550,000 image fronts as suspects, and the automated defect-metric testing identified approximately 120,000 fronts identified as "outliers" (80,000 of which were not identified as suspects by prescreening software). ADI subjected these image fronts to rapid-fire manual review. By examining images from these two sources, ADI hoped to accumulate a statistically appropriate sample of images with usability issues. Additionally, to estimate numbers of escapes (i.e., the number of images with usability issues that were not identified as suspects by the usability test software or by their status as defect-metric outliers), ADI reviewers also performed rapid-fire testing on approximately 100,000 randomly chosen nonsuspect images. Figure 3 shows a sample view of the screen that the reviewers used to perform rapid-fire manual reviews.

Figure 3. Rapid-Fire Review Data Collection Screen

During rapid-fire review, human reviewers quickly examined each image to place it into one of three categories: "all fields clearly legible," "review further" (at least one field NOT clearly legible), or "not a standard check." Reviewers performed a subsequent methodical manual review of images tagged for further review, and specially trained reviewers later performed rapid-fire review on items that were not standard checks.

ADI filtered out items that were not standard checks from the methodical review population for productivity reasons, not because some of these items would not be treated as "checks" within the banking system. FSTC project management independently reviewed approximately 2,000 of these nonstandard items, which included:

- Adding-machine tapes
- Debits
- Cash tickets
- Withdrawal slips
- Deposit slips
- "Sorry" documents
- United States bonds
- Women, Infants, and Children (WIC) checks
- Travelers checks
- Money orders
- Regular checks with or without correction strips

Anecdotal data from the rapid-fire review of nonstandard items indicated that faint text was a frequent usability issue for money order images, however, information on these items was not included in the manual methodical review data.

Note that the sample contained no image replacement document (IRD) images.

Methodical Manual Review: ADI subjected approximately 55,000 suspect images to methodical manual review. The primary goal of manual review was to identify and capture detailed information about images with usability issues. ADI used the information to investigate relationships between defect-metric test results and usability, and to provide the basis for setting usability-related threshold values for the quality-defect metrics. See appendix H for a view of the screen that the reviewers used to enter their observations.

During methodical review, reviewers examined the following eight fields on the check image for usability:

- Maker
- Check number
- Date
- Payee
- Convenience amount
- Legal amount
- Signature
- Magnetic ink character recognition (MICR) line

For each of these eight fields, ADI asked reviewers to determine if the data was legible and/or present, using the rules in figure 4.

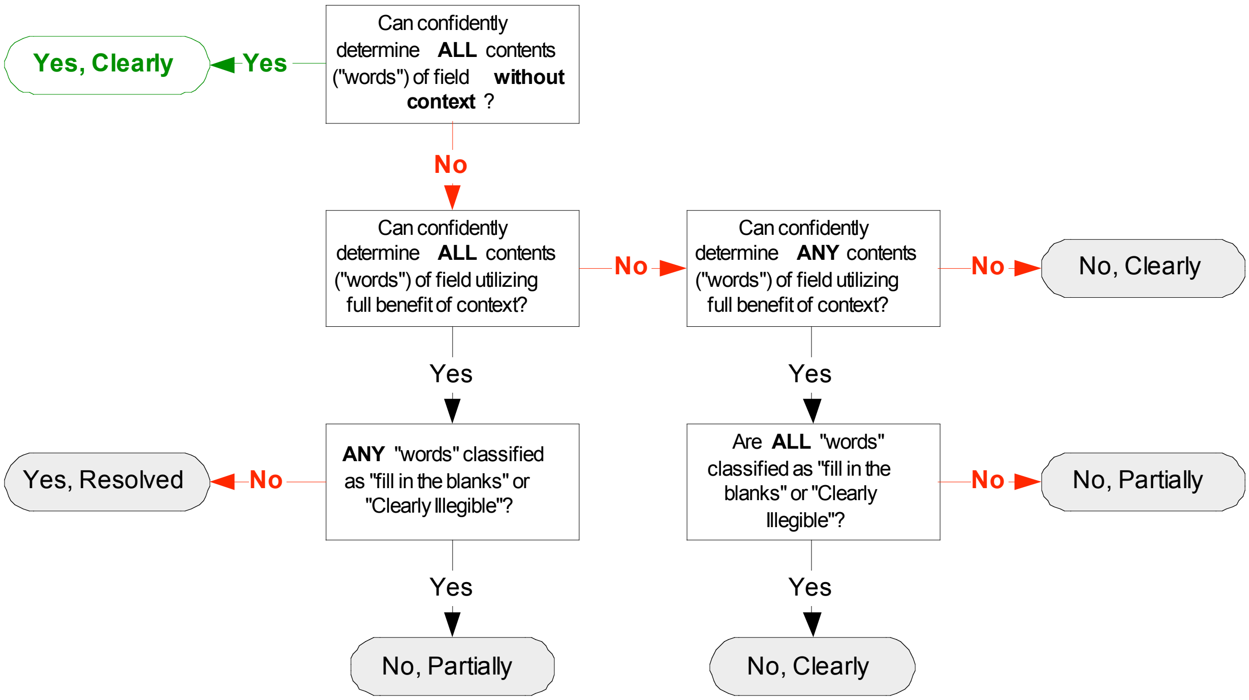
In addition to capturing the above information, methodical reviewers also noted whether an image contained two or more check images, indicating a piggyback defect, and they noted whether checks were made out by hand or machine printed.

Figure 4. Usability Rules for Methodical Manual Review

Result	Legibility	Presence (Signature Field)
Yes, Clearly	Can clearly and completely determine information in field (no ambiguity, no utilization of context).	Can determine that field information is completely present.
Yes, Resolved	Can completely determine information in field (low ambiguity, pretty sure/fairly certain, utilization of context).	N/A
Maybe	Ambiguous–a guess.	Ambiguous–a guess.
No, Partially	Can clearly determine or resolve part of information in field. Cannot resolve other information in field.	Can determine that part of field information is present and part is missing.
No, Clearly	Cannot determine any information in field, even with the utilization of context.	If completely obscured, presence is "no, clearly," as we cannot identify either as a "no field" or an "empty field".
Empty Field	Can determine that field is present, but it has no information in it.	Can determine that field is present, but it has no information in it.
No Field	Cannot find field.	Cannot find field.

ADI trained reviewers to use the following legibility logic to resolve information on check images. In some cases, ADI allowed reviewers to use context to help with the resolution process. For example, reviewers were allowed to look at the convenience amount field when resolving the legal amount field, and vice versa.

Figure 5. Legibility Logic for Methodical Manual Review



Manual Review Process for Image Backs

For check-image backs, three fields of interest were identified: the payee endorsement, the name of the bank of first deposit, and the bank-of-first-deposit number from the endorsement stamp. ADI selected approximately 5,000 images (one half of them black-and-white, the other half grayscale) for manual methodical review. ADI chose approximately 4,000 images for review based on their status as metric-defect outliers, and chose approximately 1,000 more randomly. See appendix I for a sample screen shot.

Analysis Process

At the suggestion of a project team member, ADI developed a scoring system to measure “degrees of usability problems.” This scoring system differentiates among images with various degrees of compromise, enabling analysis of usability definitions.

ADI then developed a classification system using quality-defect measurement data to determine which defects were effective predictors of image usability and which were not. For defect metrics that were predictors of usability issues, ADI recommended thresholds based on performance as defined by the precision, escape rate, and suspect rate criteria.

METRICS-TESTING FINDINGS

The data from the live image testing process proved very informative and is an asset that the industry can draw on to answer future as well as current questions about image usability. In phase two, the team was tasked with determining the usefulness of the image-quality defect metrics defined in phase one and setting thresholds for those metrics that predicted usability issues. This section describes how these goals were met and reports on the results of the analysis.

Human Image-Usability Evaluation Results: ADI's reviewers methodically examined over 60,000 images that were identified as suspects based on automated prescreening, identification of automated defect-metric test "outliers," and random review of images believed to have no problems. Figure 6 shows the outcome of the methodical manual review of defect suspects from all of these sources.

Figure 6. Front Manual-Review Results

	Maker Legible	Payee Legible	Date Legible	Check # Legible	CAR Legible	LAR Legible	Signature Presence	MICR Legible
Rapid Fire	358,485 All fields legible in rapid-fire review							
Yes, Clearly	45,808	41,304	52,421	59,413	56,486	38,971	56,463	56,101
Yes, Resolved	2,171	5,769	2,898	1,114	2,317	10,691	0	1,099
Maybe	13	25	15	12	2	3	9	3
No, Partially	13,578	10,150	5,409	1,365	2,852	7,221	3,607	3,820
No, Clearly	1,365	5,253	1,617	849	1,353	2,714	716	2,237
Empty Field	297	808	929	425	290	1,861	2,464	48
No Field	78	1	21	132	10	1,849	51	2

Upon close visual examination, most of the eight fields reviewed on the suspect images were clearly legible. The payee field was most often the source of "no, clearly" manual review results.

Image-Scoring Methodology: ADI's image-scoring system assigned a numerical score to combinations of "no, clearly" and "no, partially" for all eight fields as shown in figure 7. A field that was "no, clearly" was assigned a value of 3, and a field that was "no, partially" was assigned a value of 2. Therefore, if one field in an image was designated "no, partially," the image received a numerical score of 2, and an image with one field classified as "no, clearly" received a numerical score of 3. All numerical combinations of "no, clearly" and "no, partially" are included in the scoring system–up to and including the situation in which all eight fields received a "no, clearly" value of 3, for a score of 24 (8 x 3).

Figure 7. Image-Usability Scoring System

Usability Score	Description
0	All fields "yes, clearly"
2	1 field "no, partially"
3	1 field "no, clearly"
4	2 fields "no, partially"
5	1 field "no, partially"; 1 field "no, clearly"
6	3 fields "no, partially" or 2 fields "no, clearly"
24	8 fields "no, clearly"

The project team considered giving additional weight in the scoring system to fields of greater importance, but determined that because the defect metrics apply to an image as a whole, the relative importance of individual fields should not be taken into account.

For its analysis of the ability of image-defect metrics to predict usability, ADI used a score of 5 or greater (at least one field "no, clearly" and one field "no, partially" up to and including a score of 24--all fields "no, clearly") as the "set point" at which an image was considered to have a possible usability problem. An image with this condition has as least two fields that cannot be reliably read. This is the minimum point at which a normal check image could have both dollar-amount fields unusable.

ADI based its image scoring system on "no, clearly" and "no, partially" scores (described above) for the eight fields reviewed. The project team excluded the "no field" and "empty field" categories from the correlation analysis because it determined that a field absent in the image was likely to be absent on the original document as well, which would reflect item usability not image usability. The exception to this was if all three of the convenience amount, legal amount, and payee fields were all labeled either "empty field" or "no field." The team also included images labeled as piggybacks.

Image-Scoring Results: As shown in figure 8, test results indicate that approximately 2.5 percent of images had at least one field that was only partially legible, but only a small percentage of images had usability issues with many fields. This is a projection across the entire sample based on the results observed from the detailed reviews outlined in the tables above.

Figure 8. Image-Usability Scores–Projected to Entire Sample

Usability Score	Count	Percentage	Usability Score Description	Minimum Affected Fields
Total Volume	3,027,128			
0	2,953,217	97.556%	All fields "yes, clearly," "yes, resolved," or "not present"	0
2 or more	73,911	2.442%	1 field "no, partially"	1
3 or more	31,789	1.050%	1 field "no, clearly"	1
4 or more	21,110	0.697%	2 fields "no, partially"	2
5 or more	13,171	0.435%	1 field "no, partially"; 1 field "no, clearly"	2
6 or more	9,259	0.306%	etc.	2
7 or more	6,494	0.215%		3
8 or more	5,142	0.170%		3
9 or more	3,588	0.119%		3
10 or more	2,658	0.088%		4
11 or more	1,963	0.065%		4
12 or more	1,415	0.047%		4
13 or more	820	0.027%		5
14 or more	599	0.020%		5
15 or more	409	0.014%		5
16 or more	266	0.009%		6
17 or more	197	0.007%		6
18 or more	116	0.004%		6
19 or more	74	0.002%		7
20 or more	38	0.001%		7
21 or more	23	0.001%		7
22 or more	7	0.000%		8
23 or more	1	0.000%		8
24	none	0.000%	8 fields "no, clearly"	8
Piggybacks	41	0.001%	Actually observed 11	
No Key Data	198	0.007%	No Amount or Payee	3

Just under one-half percent of images in the sample population had two or more of eight fields with questionable usability (a score of 5 or greater), and 0.014 percent showed issues with five or more fields (a score of 15). Note that in figure 8 the scores include field legibility but not field absence. The number of items associated with each usability score is not derived from physically examining every image, rather it is a projection from statistical sampling projected onto the total sample size (similar to methods used by opinion pollsters to project opinions from a statistically valid sample onto an entire population).

Defect Tests as Predictors of Usability: Entering the project, FSTC believed that defect metrics would not be 100 percent successful in predicting which images human reviewers would identify as "usability challenged," however, FSTC expected them to prove useful in predicting usability issues.

The project team analyzed whether the defect metrics were meaningful predictors of usability issues based on the data collected. To facilitate an understanding among the individual metrics, ADI developed a ranking system. Comparatively effective predictive ability resulted in a high "ranking" assigned by the classification system, and low predictive ability resulted in correspondingly low rankings (see figures 9 through 12). Where defect metrics predicted usability issues, ADI determined thresholds by identifying the threshold values most closely associated with the optimal relationship among precision, capture rate, and suspect rate.

The project team was pleased to find that some of the defect metrics were strong predictors of usability issues. Image Too Dark (percent black pixels) and Front-Rear Dimension Mismatch (height) most strongly predicted usability for black-and-white front images in our sample, as shown in figure 9. When combined, these defects captured over one fifth of the images with five or more "usability challenged" fields.

It is important to note that although the test data caused Undersize Image (width and height) and Front-Rear Dimension Mismatch (height) to rank strongly, two anomalies in the test sample made these defect metrics relatively prominent: over 2,000 images with clipped MICR lines and some images with incorrect resolution in the tagged image file format (TIFF) tags. The project team adjusted its recommendations to account for these sampling issues as well as other identified measurement issues.

By far the most effective individual metric element for black-and-white images is the percentage of black pixels. This measurement applies to two defect metrics: Image Too Light and Image Too Dark. Image Too Dark occurs much more frequently in our test sample, and it proved a better predictor than other defect metrics on an item-by-item basis.

In cases where defect metrics predicted usability issues, ADI determined thresholds by identifying values most closely associated with the optimal relationship among precision, capture rate, and suspect rate.

Please note that all thresholds shown in the following tables represent recommended starting points for the industry based on the sample data analyzed and the usability scores used for optimization. Over time as we gain more experience, and we have images from more sources, the industry should fine-tune these starting points to provide the best ongoing operations and possibly reflect more subtle interactions among the metrics.

Figure 9. **Strong Usability Predictors for Black-and-White Fronts in Test Sample**

Metric / Element	Threshold	Precision	Suspect Rate	Capture Rate	Ranking (Higher is Better)
Image Too Dark (Tenths of a % Black Pixels)	> 300	64%	0.10%	15%	47
Undersize Image (Tenths of An Inch)	Height 20-22, Width > 57	59%	0.06%	9%	22
Front-Rear Dimension Mismatch (Height in Tenths of An Inch)	> 3	40%	0.12%	11%	15

Image Too Light showed limited predictive ability as figure 10 shows. However, at a very low value it was extremely precise.

Figure 10. **Moderate Usability Predictor for Black-and-White Fronts in Test Sample**

Metric / Element	Threshold	Precision	Suspect Rate	Capture Rate	Ranking (Higher is Better)
Image Too Light (Tenths of a % Black Pixels)	< 21	72%	0.002%	0.34%	0.915

A number of quality-defect metrics (see figure 11) showed minimal predictive ability for black-and-white front images as well as grayscale front images. Although some of the quality-defect metrics that showed minimal predictive ability, such as skew, may be a minor usability annoyance to a human, they could pose a more serious impediment to automated processes.

Note that in figure 11 Undersize Image (Height) has a ranking that indicates its predictive ability is significant, but it appears in this table because its contribution is already recognized in figure 9 in the context of Undersize Image (Width and Height Combined). Oversize Image (Width and Height Combined) also shows potential, but ADI considers it unlikely to improve overall performance.

For the grayscale metrics in figure 11, none indicate with any confidence that they can be relied upon to predict usability issues.

Even though the metrics in the table below do not show strong predictive capabilities, most of them are still potentially very valuable. They can be used as part of an overall image-quality assurance program, and when the occurrence rate of images worse than the threshold increases, the industry should take preventive action. The industry can also use the metrics to examine trends and spikes in large sample sizes and to understand and manage image quality better.

The preliminary thresholds recommended below are based on findings of two or more fields with usability issues (usability scores of 5 or worse). Each financial institution must assess its own risk to determine whether this “set point” is appropriate. The project database is available to assist member financial institutions that wish to calculate their own thresholds based on customized “set points.”

Figure 11. Lesser Usability Predictors for Black-and-White and Grayscale Fronts[3]

Metric / Element	Threshold	Precision	Suspect Rate	Capture Rate	Ranking (Higher is Better)
Undersize Image (Height) (Tenths of An Inch)	< 23	31.00%	0.1400%	10.30%	8.887
Oversize Image (Height) (Tenths of An Inch)	> 45	17.83%	0.0040%	0.17%	0.035
Undersize Image (Width) (Tenths of An Inch)	< 57	9.93%	0.2120%	4.74%	0.466
Oversize Image (Width) (Tenths of An Inch)	> 90	22.29%	0.0160%	0.80%	0.267
Oversize Image (Width and Height Combined in Tenths of An Inch)	Height 40-49 Width 87-99	18.00%	0.1800%	7.50%	2.347
Folds and Tears (Height and Width) (Tenths of An Inch)					
Bottom Right Corner	Width >=5 & Height >=25	16.66%	0.0002%	0.01%	0.002
Top Right Corner	Width >=20 & Height >=20	21.42%	0.0009%	0.04%	0.012
Bottom Left Corner	Width >=25 & Height >=10	34.69%	0.0030%	0.25%	0.189
Top Left Corner	Width >=45 & Height >=25	40.00%	0.0005%	0.04%	0.039
Top Edge	Width >=10 & Height >=5	3.78%	0.0060%	0.05%	0.017
Bottom Edge	Width >=5 & Height >=5	8.14%	0.0040%	0.08%	0.003
Left Edge	Width >=10 & Height >=10	29.34%	0.0110%	0.76%	0.428
Right Edge	Width >=10 & Height >=20	88.23%	0.0011%	0.22%	0.787
Skew Angle (Degrees, Lower/Negative)	> -30	1.32%	0.0100%	0.03%	0.00002
Skew Angle (Degrees, Higher/Positive)	> 30	6.03%	0.0040%	0.05%	0.001

[3] ADI re-examined grayscale images as part of the project's data validation process. As a result of this, ADI rescored about 38 grayscale images.

Metric / Element	Threshold	Precision	Suspect Rate	Capture Rate	Ranking (Higher is Better)
Number of Horizontal Streaks	> 3	7.62%	0.0030%	0.06%	0.002
Largest Horizontal Streak Height (B&W) (ITenths of An Inch)	> 24	7.22%	0.0060%	0.09%	0.003
Compression Size Above (B&W) (Tenths of An Inch)	> 68000	9.20%	0.0450%	0.93%	0.0566
Compression Size Below (B&W) (Bytes)	< 600	28.57%	0.0002%	0.01%	0.0052
Spot Noise (B&W Only) (Average Count of Noise Spots)	> 575	3.99%	0.0140%	0.12%	0.0012
Front-Rear Dimension Mismatch (Width in Tenths of An Inch)	> 4	15.90%	0.0160%	0.55%	0.094
Image Too Light[4] (Grayscale) (Percent Average Brightness in Tenths)	Brightness 610-650, Contrast 230-259	3.49%	0.6087%	11.61%	0.411
Image Too Dark (Grayscale)	< 500	0.62%	2.6400%	0.90%	0.000
Compression Size Below (Grayscale) (Bytes)	< 15000	4.90%	1.0700%	28.57%	3.089
Compression Size Above (Grayscale) (Bytes)	> 37000	0.01%	11.3400%	0.90%	0.000
Out of Focus (Grayscale) (Pixel Gradient)	< 51	1.51%	3.1500%	25.90%	0.694

Note that the table above includes Undersize Image Height and Width. ADI placed these in the table because they used a combination of these two metrics as the primary predictor of unusable images. Further analysis of the data leads FSTC to believe that this combination was due to anomalies of the sample being tested. As a result, we did not combine these metrics in the "Metrics-Testing Recommendations, Observations, and Conclusions" section.

FSTC also reminds the reader that the grayscale findings above were based on front images provided by one bank, predominately at 80 dpi. The Compression Size Above and Compression Size Below metrics, and potentially the others, may require different thresholds at different resolutions.

[4] Note that even though this threshold has a correspondingly high suspect rate, the team selected it because it has relatively high precision associated with it.

Figure 12. Nonpredictors for Black-and-White and Grayscale Fronts

Metric / Element	Threshold	Precision	Suspect Rate	Capture Rate	Ranking (Higher is Better)
Framing Error					
Bottom Height	NA	0.00%	0.060%	0.00%	0
Left Width	NA	0.00%	0.060%	0.00%	0
Top Height	NA	0.00%	0.050%	0.00%	0
Right Width	NA	0.00%	0.040%	0.00%	0
Piggyback	NA	0.00%	0.080%	0.00%	0
Horizontal Streaks (Grayscale)					
Number of Streaks	NA[5]	0.07%	2.3200%	0.90%	0
Largest Streak Height	NA[6]	0.68%	0.2400%	0.90%	0

METRICS-TESTING RECOMMENDATIONS, OBSERVATIONS, AND CONCLUSIONS

The project team members and ADI analyzed the test data and applied their industry and scientific knowledge to make the following recommendations. This section also summarizes the team's insight into image-usability issues and ideas about the role of image-quality testing in banking operations.

Summary of Metrics-Use Recommendations

Image Too Dark and Image Too Light Are Best Predictors of Image-Usability Issues: One image-quality defect metric, Image Too Dark, proved to be the most effective predictor of image-usability issues. Image Too Dark as well as Image Too Light are measured by determining the percentage of black pixels in an image. Although Image Too Light was not as useful as Image Too Dark in identifying images that contained usability issues, it was very precise at a low threshold setting.

FSTC believes that these two metrics are sufficiently precise to merit being incorporated into an exchange environment as part of image-quality standards for individual images.

5 There were insufficient images with horizontal streaks in the sample to determine a statistically valid threshold.
6 Same as above.

Image-Quality Metrics Are Useful for Multiple Purposes: In phase two the project team evaluated the sixteen image-quality defect metrics defined in phase one. At the project outset we planned to evaluate the metrics solely on their usefulness for exchange, but our test results and discussions led us to examine the metrics' utility for other purposes. FSTC now believes that all of the image-quality defect metrics, except Carbon Strip Detected, are useful for general banking operations because they provide valuable insight into source document issues and the health of the image-capture process. Given this conclusion, we recommend that financial institutions incorporate image-quality defect metric testing into programs designed to ensure high-quality images.

Today many banks review image-quality defect suspects regardless of which image-quality defect metric identifies the image as a suspect. Based on our findings, FSTC believes that this is not the most cost-effective practice, because most image defects showed little value in predicting whether an individual image will have significant usability issues. Therefore, FSTC recommends that financial institutions use image-defect measurements (excluding Carbon Strip Detected) as system performance indicators, and gain understanding of the application of these metrics to different capture devices. The industry should use these indicators to identify symptoms of possible system performance degradation, such as:

- Increases in rates of occurrence of defect measures outside operational thresholds
- Spikes in defect occurrences

Financial institutions should monitor defect metric performance by capture device (or device type) and adjust threshold and review policies accordingly. Financial institutions may also wish to conduct routine random reviews of suspects generated by the recommended defect metrics to highlight any issues or inconsistencies that should be addressed to ensure high-quality images.

Threshold Values Should Reflect Testing Purpose: FSTC recommends that financial institutions use different threshold values for exchange than for general internal monitoring. We make this recommendation because the workload for suspect review in an exchange environment must not interfere with clearing deadlines, and images that exceed exchange thresholds may not be of "good quality"; if so, they should require action on the part of the truncating bank. This action may include:

- Reviewing the image (either manually or using advanced, automated usability testing) to identify legitimate usability problems
- Keeping the paper check until the image is "accepted" as useful by the paying bank
- Keeping or obtaining a better image
- Clearing paper items rather than images

Manual Review Results Should Prevail in Exchange: The project team recommends that if an image that exceeds an exchange threshold is found acceptable after manual review or usability testing, a truncating bank should be allowed to override the image-defect threshold and indicate that an image is acceptable for exchange.

Additional Thresholds Should Be Assessed: FSTC observes that additional thresholds could be established, above which an image is ineligible for exchange and a paper item must be cleared. For example, such thresholds might be established if it would be impossible to obtain a usable IRD from an image that exceeds the thresholds. This is similar to current Federal Reserve Bank practices. FSTC did not attempt to establish such thresholds, and financial institutions should carefully consider the operational issues of establishing such practices.

Defect-Metric Testing Cannot Capture All Unusable Images: In a final observation, we note that despite testing and action in the case of exceeded thresholds for exchange, financial institutions should be aware that some items with image-usability issues will escape detection and enter the workflow.

Observed Value of Image-Quality Defect Metrics

Although FSTC does not recommend that all metrics be included in exchange definitions for image quality, our analysis amply demonstrates that the metrics, except Carbon Strip Detected, are potentially useful components of an overall image-quality assurance process. For example, the test sample contained a batch of approximately 2,000 check images with clipped MICR lines. The Front-Rear Dimension Mismatch metric effectively identified these problem images. Without those items (whose occurrence should be infrequent) in the sample, the metric would not have shown significant value in predicting usability issues.

In another case, both the Front-Rear Dimension Mismatch and Undersize Image metrics flagged items with incorrect resolutions in the image headers.

These findings support retaining metrics with apparently marginal predictive value, because more extensive experience may uncover new ways in which the metric test results can be used to help ensure image quality.

Need for Usability Testing

The project team concluded from the live image-testing data analysis that image-quality defect testing alone will not completely protect a financial institution from risks posed by poor-quality images, nor does it ensure that images will be of sufficient quality to support all downstream processes, because testing cannot be depended upon to capture all images with usability issues. Additionally, image-quality defect testing cannot solely ensure the usability of a specific check field.

The team, however, also concluded that image-quality defect metrics and testing constitute a necessary step toward the following financial industry goals:

- Identifying images that may have usability issues caused by image-quality defects
- Enhancing the effective performance of image-capture systems by serving as a diagnostic tool
- Enabling more effective automated character recognition (based on industry expertise)

It is impossible to know in advance whether a usability issue, even if it does not appear "severe," will cause an operational problem, return, or loss downstream. Also, even if quality is sufficient to meet a truncating bank's needs, it may not meet the needs of all others in the image path. For example, industry experience has demonstrated that even slight imperfections can degrade the performance of automated processes like automated character recognition.

From the analysis to date it is not possible to determine fully what rate of truly bad images will escape the image-defect tests and FSTC's currently recommended thresholds. However, FSTC believes that image-defect testing using the current metrics will capture no more than 20 percent of all checks with usability issues (usability score >= 5), and probably no more than 60 percent of check images with severe (usability score >= 15) usability issues. In addition, FSTC does not believe the current defect metrics will effectively identify items in which text is "too light" to survive conversion to a black-and-white image.

Because it is impossible to determine if any given image will result in a problem, a financial institution must determine whether it can tolerate the risk from these "usability escapes" or whether it should introduce other protective measures, such as usability testing or storage of grayscale images (which appear less susceptible to "faint text" problems).

Sample Issues and Measurement Reliability Observations

Because the majority of grayscale front images were 80 dpi, from one bank, and because we were unable to obtain images in both renditions from the same paper items in the study, FSTC cannot draw any statistically meaningful conclusions about grayscale versus black-and-white image legibility. Further, 80 dpi grayscale images are not the common operating point within the industry at this time. As a result of both of these factors, the preliminary metric thresholds for grayscale need further experience to validate our recommendations. In addition, our sample size was limited for black-and-white backs, and as a result FSTC recommends those thresholds be used as an industry starting point.

Framing errors occurred infrequently in our test sample. Also, by design of our test, a framing error could not directly lead to a usability issue at the field level. As a result of both of these factors, framing errors are found to have no predictive value as a defect metric. However, although the Framing Error metric is not a predictor of image usability in a non-IRD environment, it may prove useful as part of the IRD printing process. This is because a framing error can cause the image printed on the IRD to be smaller than it would otherwise be, possibly causing IRD usability issues. IRDs were not included in the test sample.

Some image-quality defect metrics are more difficult to implement than others, making them prone to measurement errors. Although FSTC enabled vendors to test the accuracy of their implementations against a calibration deck of images, the project goals did not include determining measurement accuracy during live testing. We detected measurement errors in the calculation of negative skew, as a set of images was incorrectly identified as having negative skew of nearly 20 degrees. Over time, FSTC expects implementations to improve, reducing measurement errors, but in the meantime, FSTC adjusted the recommended negative skew threshold to remove the inaccurate measurements from the sample.

As part of the data collected during the methodical manual-review process, ADI reviewers noted whether an imaged check was written by hand or printed by machine. Analysis of this data indicated the following trends:

- Payer information was more legible on machine-printed than handwritten checks
- Amount information was more legible on handwritten than on machine-printed checks
- MICR lines were more legible on machine-printed than on handwritten checks

Although no root causal analysis was done, these results point to background interference with payer and MICR lines as the probable cause. The stringent requirements for a clear background in the convenience amount field placed on personal checks seem to provide substantial legibility benefits to the check amount field. Reviewers also indicated that poor handwriting was not a major contributor to the legibility results for handwritten checks.

Anecdotal information from reviewers and industry experience indicate that faint text is a frequent contributor to image-usability issues in black-and-white images. This condition does not appear to be detected by the existing image-defect metrics; rather, it appears that it can only be detected by field-level usability analysis. The collective experience of the project team has shown that source-document characteristics such as ink color, "noisy" backgrounds, and low contrast also adversely affect item usability after black-and-white image conversion.

One other important factor to recognize is that the sample analyzed in this study was not truly random. It represented a sampling from eight financial institutions, and although representative of their workloads and processes at that time, it cannot be assumed to be representative of the mix of future images that may be created with the adoption of distributed capture and new technologies. As a result, financial institutions need to track the performance of these recommendations over time and adjust the thresholds based on changes that occur in their images.

Root Cause Validation

Root cause analysis was specifically outside the scope of the project. However, given that we may not have an opportunity to review the images again in such a controlled setting, ADI voluntarily performed a very quick, superficial, visual root cause analysis for the 420 images in the dataset that had the "worst" usability scores (>=15). Figure 13 summarizes their findings:

Figure 13. Manual Root Cause Validation Analysis

Cause	Count
Image Mostly Too Light	37
Image Partially Too Light	118
Field Input Too Light	51
Image Partially Too Dark	114
Image Mostly Too Dark	51
Image Blotchy	14
Other	35

Image Too Light and Image Too Dark directly target the items shown in bold in the table above and represent 320 of the 420 images with severe usability issues.

This analysis supports the purely statistical analysis performed with the raw data results from the metrics. In addition, analysis of the collected data indicates that the recommended thresholds for exchange for these two metrics would identify 180 of the 320 (56 percent) images with Too Light/Too Dark causes. Using just these metrics leads to an estimated suspect rate under .025 percent in our sample.

FSTC cautions the reader, however, that for efficiency reasons the methodical manual-review process excluded money orders, which are believed to be particularly prone to faint printing. As a result, FSTC believes that the rate of occurrence of Field Input Too Light, or "faint text" as it is sometimes called, is likely to be higher than captured in this study.

The Data Set as An Industry Asset

The FSTC IQ&U phase-two project has created a database of image-quality information that is a very valuable asset for the industry. This is the most rigorous study of check-image quality defects ever undertaken. It includes field-level usability data for almost 65,000 check images and manual assessments of nearly 700,000 items. FSTC believes that the findings put forward in this report are a fraction of what can be gleaned from further analysis of the information, and to help meet the needs of the industry FSTC plans to make the data set available to members who participated in phase two of the project. Among other things, members can use this information to calculate the efficacy of revised defect-metric thresholds.

In addition, the project created an accuracy deck of images that had metric values "truthed." The industry can use this deck of images to improve image-quality defect measurement implementations and to detect incorrect implementations of interoperable defect metrics. This tool will help ensure that test results are consistent across test engines.

Recommended Thresholds

FSTC makes the following recommendations based on its live image test data. These recommendations are based on the ADI analysis but were adjusted by the experts on the project team to ensure an appropriate balance between suspect rates and effectiveness at capturing images with usability issues.

Image Fronts–Recommendations for Exchange: If pre-exchange suspect review is required between exchange partners, FSTC recommends the following minimum settings for image fronts. Should these thresholds be exceeded, FSTC recommends that the truncating financial institution take action.

Figure 14. Recommended Thresholds for Exchange–Black-and-White Fronts

Metric/Feature	Threshold	Unit of Measure
Image Too Dark (Percent Black Pixels)	> 390	Tenths of a percent
Image Too Light (Percent Black Pixels)	< 21	Tenths of a percent

Image Fronts–Recommendations for General Use: FSTC recommends that financial institutions use the following metric thresholds (in order of priority) as part of an image-quality control program for calibration and performance tuning, and apply them in exchange consistent with a financial institution's risk mitigation requirements and business practice agreements. Individual outlier results for defect-metric testing should raise suspicions; however, each financial institution must determine its own risk tolerance and acceptable review workload to decide whether to review suspects or store better images for isolated occurrences. Financial institutions should routinely investigate frequent occurrences above the thresholds to identify possible machine, system, or source document issues.

Figure 15. Recommended Thresholds[7] for General Use

Metric/Feature	Threshold	Unit of Measure
Undersize Image (Height)	< 22	Tenths of an inch
Oversize Image (Height)	> 45	Tenths of an inch
Undersize Image (Width)	< 57	Tenths of an inch
Oversize Image (Width)	> 90	Tenths of an inch

[7] ADI adjusted threshold recommendations for Image Height, Image Width, Compression Size Above, Front-Rear Dimension Mismatch, and Negative Skew based on the project team's view of acceptable suspect levels and/or to account for anomalies present in this particular test sample.

Metric/Feature	Threshold	Unit of Measure
Image Too Dark (Percent Black Pixels)	> 300	Tenths of a percent
Image Too Light (Percent Black Pixels)	< 21	Tenths of a percent
Front-Rear Dimension Mismatch (Width)	> 4	Tenths of an inch
Front-Rear Dimension Mismatch (Height)	> 5	Tenths of an inch
Compression Size Above (BW Only)	> 90000	Bytes
Compression Size Below (BW Only)	< 600	Bytes
Largest Streak Height	> 24	Pixel % in scan line
Number of Streaks	> 3	Streak count
Spot Noise (BW Only)	> 575	Count of spot noise
Skew Angle (Lower/Negative)	< -30	Tenths of a degree
Skew Angle (Higher/Positive)	> 30	Tenths of a degree

If financial institutions create black-and-white renditions from grayscale images, then FSTC recommends that the grayscale images not be tested. Rather, FSTC recommends that financial institutions test the black-and-white image created from the grayscale rendition.

Edge defects (Torn and Folded Corners and Edges) were infrequent in our sample, and required very large defect sizes to reasonably indicate a usability issue. At this time, FSTC does not recommend a specific threshold for edge defects.

Framing errors were infrequent in our test sample and showed no predictive ability in identifying images with usability issues. FSTC is unable to recommend a preliminary threshold setting for framing errors.

Figure 16. Recommended Thresholds for General Use–Grayscale Fronts

Metric/Feature	Threshold	Unit of Measure
Image Too Light	610-649 230-259	% Average brightness in tenths % Average contrast in tenths
Image Too Dark	< 550	% average brightness in tenths
Compression Size Below Minimum	< 15000	Bytes
Compression Size Above Minimum	> 45000	Bytes
Out of Focus	< 46	Pixel gradient

As mentioned before, the grayscale recommendations are based on the images available in the test sample. For grayscale fronts these were from one bank and were predominantly 80 dpi images. Banks using these recommendations should closely monitor system performance and adjust the recommended grayscale thresholds as appropriate based on experience in their own environments.

Image Backs–Recommendations: FSTC recommends the following metrics as a starting point for image exchange and for general use. Because the sample size was not as large and the analysis was not as rigorous, FSTC recommends no specific action on a per-item basis should a threshold be exceeded.

Figure 17. Recommended Thresholds–Black-and-White Backs

Metric/Feature	Threshold	Unit of Measure
Image Too Dark	> 21	Tenths of a percent
Undersize Image (Height)	< 22	Tenths of an inch

Image Too Light, when applied to image backs, presents a dilemma. Although the metric is potentially useful to identify backs that contain no data, in cases where an image is captured in a distributed environment and is not endorsed, this condition could occur legitimately and therefore should not be identified as an image defect. In addition, our data sample did not contain sufficient very light rear images to determine if even zero percent and 0.1 percent black pixels would truly represent a usability problem. For this reason, FSTC recommends that the industry collect more data and review more samples of image backs that are very light before determining if a threshold should be established for this metric in an exchange environment.

FSTC recommends that financial institutions consider employing two additional metrics for testing of grayscale backs. As mentioned above for image fronts, if financial institutions do not plan to retain grayscale back images, FSTC recommends that they not be tested. Rather, FSTC recommends that financial institutions test the black-and-white images created from the grayscale renditions.

Figure 18. Recommended Thresholds–Grayscale Backs

Metric/Feature	Threshold	Unit of Measure
Grayscale Image Too Dark	< 300	% Average brightness in tenths
Out of Focus	< 35	Pixel gradient

Piggybacks: FSTC recommends that the piggyback image-defect metric be treated as a special case. Piggyback detection exclusively from images does not yet appear reliable, although a piggybacked image represents severe usability issues. FSTC recommends, therefore, that financial institutions perform piggyback detection using hardware sensors until vendors can improve their image-based piggyback detection. Based on live image-test results, current implementations of this metric do not appear to be sufficiently reliable to base specific actions on.

Recommended Actions for Exceeded Thresholds

Financial institutions have a number of options when an image-quality defect metric threshold is exceeded. For metrics used in exchange, FSTC recommends that action be taken when thresholds are exceeded, and response to other metrics is best determined by risk management policies, system capabilities, and suspect-review tolerance. Depending on the circumstances, financial institutions can choose to:

- Take no palliative action
- Manually review the image to judge if it is usable
- Retain the paper check longer
- Clear the item as paper
- Rescan the check
- Retain a "better" image (e.g., a financial institution can automatically capture and retain a grayscale image if a defect threshold or a dollar-amount threshold is exceeded)

For metrics not specifically recommended for image exchange, financial institutions may not choose to act on individual occurrences because the "suspect-to-true-defect rate" would be too high, which may slow processing. Nonetheless, financial institutions should continue to monitor for these defects, and if rates of occurrence gradually rise, they should investigate the cause. A sudden spike in occurrences may warrant review of some or all of the suspicious images to determine if a system problem has created a batch of unusable images. In all cases, financial institutions may wish to randomly sample and review images outside recommended thresholds to fine-tune the thresholds and minimize the possibility that images with usability issues enter the image-exchange process.

For metrics recommended in image exchange, FSTC anticipates that image-exchange business practice agreements and clearinghouse rules will specify actions required when thresholds are exceeded. FSTC believes that the suspect rate on these items is manageable, and the image-defect metric and recommended thresholds are good enough predictors of usability issues. Therefore FSTC anticipates that a truncating bank will be able to take appropriate action on these items before including them in exchange.

Conclusions

Readers should consider the defect-metric thresholds put forward in this report as preliminary indicators of points at which image-quality defects can indicate usability issues. More work remains to refine the thresholds, apply them to individual types of capture devices, and to examine combinations of metric elements that may be stronger usability predictors or reduce the number of suspects captured. Over time, as financial institutions collect live data using these metrics, this refinement will occur and financial institutions should adjust these threshold recommendations accordingly.

Even given the untested nature of these recommendations in production environments, FSTC is confident that the industry will obtain great benefit from these recommendations and this project. Specifically, we now have scientific data and analysis for black-and-white images that:

- Distinguishes the relative importance of the metrics
- Enables us to understand approximately what percentage of usability problems we are successfully capturing with image-quality defect measurements
- Gives us tools to reduce suspect-review rates substantially

FSTC believes that the study results provide an excellent outcome for the financial services industry. For the first time, the industry has the scientific data to understand which image-defect metrics provide the most value and to apply these metrics most cost-effectively to ensure image quality. In addition, the industry now has sufficient information to understand the limits of image-quality defect testing effectiveness. Financial institutions can now make informed decisions about testing implementation as well as the risks associated with their testing programs.

USABILITY DEFINITIONS AND METRICS

Two key project objectives were to reach agreement on the terminology and definition of usability for check images and to develop metrics to identify images with usability problems. After much discussion and analysis of the fields required for the various use levels defined in phase one (see "Considerations in Designing a Usability-Testing Process" below for use-level definitions), the project team agreed to define usability at the level of individual fields. Check-image users can then assess individual fields to determine if the image will be usable for their purposes.

FSTC identified three key questions for determining field usability:

1 Is the field itself present on the check?

2 Is the data for the field present on the check and in the check image?

3 Is the data legible?

1. Field Not Present: FSTC noted that not all fields are present on all types of checks. For example, some checks do not contain a legal amount field where the amount of the check is spelled out in words. This distinction is important because if a field is not present on the original source document, it is not an image-usability issue and should, therefore, not be included in an image-usability metric.

2. Data Not Present: A field with no information in it can pose an image-usability issue but often does not because check issuers routinely omit information accidentally or intentionally. For example, a check issuer may leave the issue date blank. This condition will not make the check unpayable, however, banking practices and regulations spell out the appropriate handling of cases where information is missing from the face of the document.

3. Legibility: Legibility reflects the ability to use the information on the face of the check or in the image for business purposes. FSTC defines legibility as "distinctness that makes perception easy.[8]" As with the presence of data, legibility issues may be caused by the source document or introduced during the imaging process.

FSTC concluded that, although desirable, there is currently no technology available to distinguish all imaging-caused usability issues reliably from those caused by incomplete source documents when only black-and-white images are available. As a result, FSTC's field-legibility metrics do not attempt to attribute causes to legibility issues.

In concluding its analysis, FSTC determined that the appropriate metric to develop was a metric based on the absence or legibility of information on the check. Because (as illustrated in appendix C) there is such a variety of data contained within checks, FSTC recommends the following key data fields, which implementers and financial institutions should focus on when developing usability assessment tools:

- The MICR line and MICR fields
- Amount–both convenience and legal
- Payer information
- Payee information
- Signature(s)

[8] Source: WordNet ® 2.0, © 2003 Princeton University

The usability metrics themselves are available separately. It is important to note that FSTC does not expect image-quality defect metrics to identify at least one known image-capture problem: faint text. The two primary causes of faint text are:

- Worn printer ribbons or low toner
- Low contrast between backgrounds and data added to the check, particularly where the ink used is a color similar to the check background

In conducting the image-quality defect study, FSTC did not manually review images of money orders for productivity reasons. Money orders fall into a specific class of documents whose faint image text is frequently caused by worn printer ribbons.

FSTC believes that should financial institutions overcome the field-location challenges described below, the usability metrics can identify items with faint-text-based usability issues.

FSTC's development of image-usability metrics follows the same approach as the development of image-quality defect testing metrics. Through its image-quality project team, FSTC has brought together the best collection of expertise from the technology and banking communities to work cooperatively to define interoperable metrics that can determine the usability of a specific field of data within a check image.

For image-quality defect metrics, FSTC embarked on a four-step process to:

1. Define a set of metrics based on the best collective expertise
2. Work with implementers to ensure consistent results
3. Scientifically evaluate metrics against a large sample of live images to assess effectiveness and usefulness
4. Encourage Accredited Standards Committee (ASC) X9 to include certain metrics in a standard registry of image-quality defect tests

FSTC is now near completion of the first step of a similar industry process for image-usability metrics. The second step, achieving interoperability of field-level image-usability metrics, is expected to be more challenging than the same step for image-quality defect metrics. The image-quality defect metrics were strictly quantitative, using well-defined units of measure. The image-usability metrics are based on legibility, which is not definable in fully quantitative terms. There is no standard methodology for assessing legibility in the industry, other than through human tests that primarily assess the legibility of various fonts.

To address the consistency-of-implementation challenge, FSTC developed a set of definitions for gradations of legibility. These gradations are based on a human interpretation of legibility. FSTC derived its recommended five-point field-level scale from definitions initially developed by Unisys and Parascript and validated through testing. The scale, described below, includes the absence or complete illegibility of a field and three gradations of legibility.

Figure 19. **Usability Truth Gradations**

Usable	Legible	Unambiguously able to interpret the intended meaning or characteristics of the data (without context).
	Mostly Legible	Can reasonably (reliably) interpret the intended meaning or characteristics of the data (in context) with some ambiguity.
	Mostly Illegible	Can only partially (low reliability) interpret the intended meaning or characteristics of the data with significant ambiguity.
Unusable	Illegible	Unable to interpret the intended meaning or characteristics of the data.
	No Data or Field Present	No data is found in the field.

Although slightly different than the approach used in the image-defect metric testing performed in this phase, FSTC believes that the industry can, should it so desire, leverage the data already collected on the images reviewed in this project to construct a test for usability vendors to calibrate their implementations.

The difficulty in correctly locating fields of interest within checks will make image-usability testing more complex. This is relatively simple on personal checks because those have very standard layouts. Unfortunately, however, commercial checks and other "check" types such as money orders have substantial variety in their layouts.

To address the field-location challenges, FSTC has encouraged ASC X9 to develop more specific standard requirements for field location, font sizes, and backgrounds for commercial checks.

In considering all the issues identified in usability testing, FSTC project members had many in-depth discussions about how to design a usability-testing program that provided the best balance between cost effectiveness and identification of images with usability issues without identifying too many false positives. The following section of this report summarizes these discussions.

CHECK-IMAGE USABILITY TESTING GUIDELINES

This section provides guidelines that financial institutions can use to design their own usability-testing programs. It seeks to provide practical guidance for balancing the tradeoffs associated with usability testing.

What Is Usability Testing?

Image-usability testing helps determine if tested information is present and legible in the check image. It is performed on selected check fields, and an overall image-usability assessment is derived from the field-level information. Image-usability testing does not validate or authenticate the information, for example, it is not used to determine if a convenience amount and legal amount are equal. Legibility in the context of check-image usability testing refers to the ability to interpret the intended meaning or characteristics of the data (without context other than the type of data appropriate to a field).

It is important to note that today's image-usability testing technology cannot cleanly distinguish image usability from source-document usability for all types of images. For example, an issue-date field is unusable if it is not present in a check image, regardless of whether the field was lost during image capture or was absent from the source document. By necessity therefore, usability testing will identify situations in which data is not present on the source document as well as cases in which data was lost during imaging.

Role of Usability Testing in Risk Management

Unusable and hard-to-use check images increase financial institutions' risk of loss--therefore, usability testing can serve as a helpful component of a risk management program for item processing. Usability testing is one step toward "certainty" or mitigation of risk because it determines the presence and legibility of information contained in a check image.

An image has usability under two conditions:

1 Presence: Information is present.

2 Legibility: Information is discernable to some degree.

Additional steps toward increasing certainty are not part of usability testing but are part of item processing. These include determining information appropriateness, verifiability, and provability as described below.

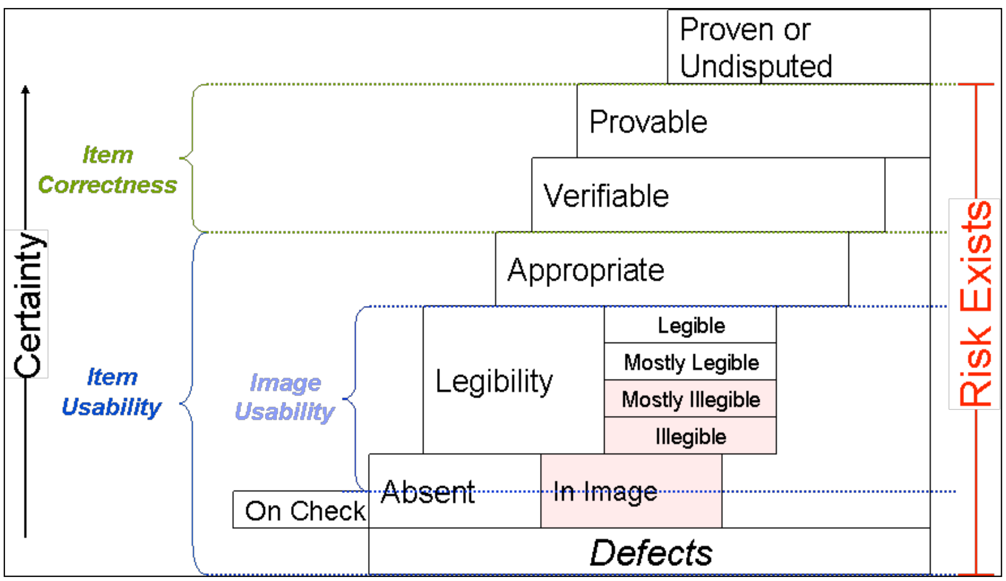
- **Appropriate:** Information is of the appropriate type. For text fields this means numeric characters in the convenience amount and alphanumeric characters for other fields.
- **Verifiable:** Information can be verified with the transaction record. For example, a maker signature can be matched with a signature card, and an amount field can be verified with a transaction record.
- **Provable:** Information can be verified using information outside the check and transaction record. This step might only be completed with the maker's participation.
- **Proven or Undisputed**: For a check transaction to be proven, it must be known beyond dispute to be accurate and authentic. This occurs through concurrence of the maker and beneficiary, or when the period during which a transaction may be challenged has expired.

Figure 20 illustrates how image-defect and usability testing contribute to the eventual certainty of a transaction. As shown, the steps toward certainty build upon one another, with gross image-defect testing the first step in weeding out potentially unusable images.

Usability testing determines the presence of the most important information on the check and assesses the legibility of that information. Beyond the usability testing described in this document are tests for application uses such as testing to determine if a convenience amount field contains numeric characters, or testing to determine if the MICR line in the image matches the transaction record. As certainty of information in a check image increases, risk associated with that image decreases.

Usability testing can be an explicit, separate activity, or it can be implicit through the use of the image. If an image is successfully used to read the convenience amount field through an automated or manual process, then the convenience amount field is proved to be usable in the image, mitigating the need for separate usability testing.

Figure 20: Steps to Certainty in Risk Mitigation



NOTE: When a payment becomes final and claim periods have expired, risk is completely mitigated.

Absent, partially legible, or illegible information not only increases risk, it can also impede bank operations by reducing productivity, increasing errors, and increasing the number of returned items. For this reason, usability testing may be a useful tool to ensure that images will meet processing needs.

Usability Suspects

As described above, image-usability testing will identify suspects that have one or more fields with a possible usability issue (e.g., poor legibility), regardless of what caused the issue. Historically, most checks with usability issues pay without problem, and customers do not dispute the item or the image quality. Unfortunately, however, it is impossible to determine for any given item if a field-usability problem will cause an operational issue, write-off, customer dispute, or service request.

Because it is impossible to predict if a transaction will result in a dispute, many institutions are concerned that usability testing identifies too many suspects that will not pose actual usability problems. A number of factors contribute to this concern, including:

- Testing at the field level
- Difficulty locating fields within the check
- Absence of data on the source document
- Limitations in accurately gauging legibility
- Interfering marks and image artifacts
- Not knowing future uses

FSTC has taken these factors into account in developing recommended guidelines for usability testing and options for handling checks with usability concerns.

Considerations in Designing a Usability-Testing Process

There is no agreed-upon, universal description of what constitutes an unusable check image or how many fields must have usability issues before an item should be considered unusable. In the case of paper items, financial institutions have based usability on a simple question: can they clear or pay the item? In most cases, financial institutions will likely apply the same criteria to check images, even with the additional risks surrounding image capture and check truncation.

Image use is an important consideration when designing appropriate usability-testing processes. FSTC identified four use levels, each with increasing minimum information needs:

- Level I: collection, exchange, and posting
- Level II: exceptions and returns
- Level III: fraud detection and loss prevention
- Level IV: customer usage

Determining legibility is a key aspect of usability testing. Legibility for text generally refers to the ability of a person to make out the information, and usability-testing systems need to match their legibility determination to the results that a "reasonable" person would give.

Generally, FSTC feels that absence of one or several fields is less likely to be an indication of an image-usability problem than illegible or mostly illegible fields, as long as other fields from the same step in the transaction life cycle are present. We add this caveat because often the maker of a check will not fill out all fields. Financial institutions will still want to assess transaction risk using other factors, in which case absent fields may contribute differently to the risk assessment, depending on issues such as the transaction amount.

To support Level I use, it is vital that either the convenience or legal amount (or both) on the check image is mostly legible or legible, and that the check image contains a legible MICR line. A complete list of information on a check, mapped to use levels, is available in appendix C.

For Level II use, all payment-related information on the check image front and back should be mostly legible or legible. If some of the fields are mostly illegible, the image may still be acceptable if there are alternate methods to obtain that information (such as an account lookup).

Level III use requires slightly more from a check image, including payment-related information on the front and back check image that is mostly legible or legible, security features on the original check that survive black-and-white imaging, noninformational payment-related data that are sufficiently legible to allow identification and comparison of fonts and other attributes, and mostly legible or legible authentication data.

Finally, to support Level IV use, all information on the front and back check image should be mostly legible or legible, including items such as processing stamps.

When a truncating financial institution captures an individual image, it often does not know at what level the image will be used. A financial institution must, therefore, examine risk associated with image usability for possible use levels, while also bearing in mind the likelihood that the image will never be viewed at all.

Assessing Usability Risk

In the context of the usability levels described above, FSTC recommends a risk-based approach to determining which images to test and what tests to employ. In this approach, factors such as account history, transaction value, capture process, and defect and usability measurements are weighted, and these weighted factors are examined in combination to assess risk.

- **Account history:** If an account is new or has a history of issues or questionable activity, it is likely to have more associated risk.
- **Transaction value:** Generally a high-dollar-value transaction carries a higher risk level than a low-value transaction.
- **Capture process:** Capture platforms and sources without known and trusted controls, such as those on a commercial merchant's premises, may have more associated risk than platforms under strict control at a financial institution.
- **Defect and usability measurements:** If an image has already been subjected to image defect and usability testing (e.g., by a trusted exchange partner) and has received acceptable scores, then associated risk is likely to be lower than if the image was not tested.

FSTC also suggests that financial institutions tailor their usability-testing programs to reflect the image-use patterns and usability track records associated with the following scenarios:

- Images captured through single-pass and multipass high-volume systems
- Images obtained from bank-controlled distributed-capture systems
- Images obtained from uncontrolled distributed-capture systems
- Images obtained from ATMs and other unmanned devices
- In-cleared images
- Images used to create IRDs
- Images that have already been successfully used (because successful-use experience is a strong indicator of image usability)

A bank providing images for IRD creation (or creating its own IRDs) must provide indemnification from losses if the IRD is not an accurate representation of all information on the front and the back of the check at the time of imaging. Because IRD creation generally results in a less usable item than the source document (particularly for checks larger than wallet size), usable IRDs require high-quality and very usable images. For this reason, although IRDs are common, check truncators may wish to subject images to more in-depth usability testing.

What to Do with Usability Suspects

A number of actions can be taken once usability suspects are identified. For a truncating bank, these actions can include:

- Continue to process the item
- Perform manual review
- Recapture the image
- Save a better image (e.g., a grayscale image) for a period of time or permanently
- Save the source document for a longer period
- Clear the paper check

Appropriate action may vary depending on the environment in which the capture is performed. For example, in a centralized environment, clearing the item as paper may be an option because the paper check may be readily available, whereas for timely transaction processing at a remote ATM or branch location, it may be necessary to clear an image.

A bank receiving images from outside sources generally has the following options:

- Continue to process the item
- Perform manual review
- Request a better image
- Refuse the item

FSTC recommends that each financial institution select the action(s) best reflecting its risk, cost, and customer satisfaction considerations. Because it is best not to slow payment, and because returns and exception processing can be costly, FSTC recommends that items not be returned for image-quality reasons without first using automated testing and/or manual review to determine if they are truly unusable.

Candidate Fields for Usability Testing

Fields of interest vary depending on use level, but for general risk management and bank operations efficiency, the FSTC project team identified the information elements (and corresponding fields) in the following table as most important for financial institutions to consider testing for usability in cases of at-risk transactions.

Figure 21. Candidate Fields for Usability Testing

Test For	Information Elements	Fields to Evaluate	Use Levels
Legibility	Amount	Convenience amount Legal amount	I → IV
Legibility	MICR line	MICR line	I → IV
Legibility	Beneficiary	Payee name(s)	II → IV
Legibility	Maker	Maker name and address	II → IV
Legibility (for Comparison)	Signature	Signature(s)	III → IV

The project team considered fields on the back of the check, but after analysis of field usage and location challenges, and the expectation that these fields will become less frequent over time, the project team determined that there was insufficient value to recommend usability testing of the back of the check.

USE OF IMAGES FOR FRAUD DETECTION

As check-image exchange becomes prevalent, financial institutions will increasingly rely on check images to detect fraud. The project team deemed it important, therefore, to survey financial institutions' fraud departments and the Check 21 Advisory Committee of the Association of Forensic Document Examiners (AFDE) to understand check-image requirements for fraud detection. This section summarizes the results of our survey.

Fraud departments work early in the check-clearing process to detect fraud and prevent losses, and forensic examiners are called upon to help identify fraud perpetrators after a claim has been made. Because the two groups consequently have different check-image requirements, survey responses from the two groups are described separately.

Current Practices

All of the fraud departments surveyed use automated screening programs to identify fraud suspects. These programs screen digital information accompanying the check image for anomalies or suspicious patterns, such as unexpected check number ranges or inconsistent account activity. Only items flagged during automated review or items questioned by customers are examined manually. No institution surveyed relies on manual review of check images for initial fraud screening.

The industry is currently enhancing fraud-detection capabilities to include systems that automatically analyze check images to identify fraud suspects. Financial institutions can add this new image-screening capability to the current practice of analyzing only the digital information accompanying the image.

Once suspect check images are identified, human reviewers examine handwriting, check style and layout, field placement, font characteristics, logo placement, writing style, and other visual attributes of a suspect image. Although it is not possible to examine an image for such physical attributes as paper texture and coloration, the fraud experts surveyed indicated that this is offset by faster, easier comparisons with known good transactions drawn against the same account and with known fraudulent items.

Counterfeits, Forgeries, and Alterations

The fraud departments we surveyed did not deem it important to use images to specify fraud type. The only operational implications of determining fraud type are that a forged[9] or counterfeit[10] check is subject to the normal midnight return deadline, making rapid detection critical, whereas an altered[11] check or a check with a forged endorsement is not subject to the midnight deadline, giving a bank up to three years (depending on state and circumstances) to dispute the item.

---

[9] According to the *ABA Deposit Account Fraud Survey Report - 2004*, forgery is defined as "Losses due to any check or other negotiable instrument bearing a forged, unauthorized, or improper endorsement of the payee(s)/maker's signature, includes both on-us and transit items."

[10] According to the *ABA Deposit Account Fraud Survey Report - 2004*, counterfeit is defined as "Losses due to any printed, photocopied, or other reproduction of a check or other negotiable instrument not authorized or issued by the bank or the holder of the account the item is drawn against, includes both on-us and transit items."

[11] According to the *ABA Deposit Account Fraud Survey Report - 2004*, alteration is defined as: "Losses due to any check or other negotiable instrument bearing any mechanical or chemical alteration of the payee(s) name(s) and/or amount, includes both on-us and transit items."

Unlike fraud-department staff, forensic examiners are called upon to identify fraud type, and they shared their experiences using images to perform their examinations. Forensic examiners are often able to detect forgeries by using images because maker signatures frequently identify forged checks, and signatures tend to survive imaging.

Counterfeit checks can be more challenging to detect using images because many counterfeit-detection methods rely on physical examination of check-stock characteristics, watermarks, microprint, color differences, features seen under ultraviolet light, and other check attributes that may not survive imaging. Some counterfeit-detection methods do rely on image-survivable attributes including printing techniques, font use, field placement, field content, and even misspellings. These attributes survive in black-and-white images, but if available, forensic examiners find that high-resolution grayscale and color images provide even better results.

One fraud area in which image use can actually provide an advantage over original checks is deposit fraud. With image-based systems, it is possible to identify checks drawn on the same account, even if that account is not "on-us." This enables possible document comparison, even in the case of a deposited transit item.

The Check 21 Advisory Committee of the AFDE concluded that quality scanning can yield useful forensic information to prove legal responsibility for a loss. The committee also noted, however, that reliance on images precludes physically based avenues of investigation to prove that check elements have been "raised," "washed," or "obliterated" and to detect physical abrasions due to erasure, differences in ink detectable using infrared or ultraviolet light, or physical indentations from previously written checks.

Check Elements Important in Detecting Fraud

When fraud is suspected, experts routinely examine the following fields for evidence of fraud:

- Maker
- Payee
- Address
- Amount (legal and convenience)
- Check serial number
- Routing number
- Signature
- Date
- Memo line (although less frequently than the above)

The financial institutions we consulted indicated that maker information, payee line, address, legal and convenience amounts, check serial number, and routing number constitute the most frequently altered information fields on a check. The Check 21 Advisory Committee of the AFDE noted that forensic examiners are most often called upon to investigate maker and endorsement signatures, amounts, and dates.

Figure 22. Check Fields and Fraud

Check Field	Most Frequently Altered	Routinely Examined by Fraud Departments	Routinely Examined by Forensic Examiners
Maker	✓	✓	
Payee	✓	✓	
Address	✓		
Amount (Legal and Convenience)	✓	✓	✓
Serial Number	✓	✓	
Routing Number	✓	✓	
Signature		✓ (maker)	✓ (maker & endorsement)
Date		✓	✓
Memo Line		✓	

Check-design elements most frequently scrutinized in fraud detection include:

- Font type and consistency
- Logos
- Field placement (especially for maker name, logo, and fractional transit number)
- Character alignment
- Character spacing
- Check stock (although less frequently than the items above)

Fraudulent items may not be printed with the same quality standards as authentic checks, resulting in a loss of sharpness in some check features, such as a logo. Financial institutions indicated that the ability to distinguish the sharpness of such features in an image can help identify fraudulent items.

Financial institutions indicated little reliance on check backgrounds or hard-to-copy imagery such as high-resolution borders, portraiture, or holograms to detect fraud.

Financial institutions noted that items with correction strips and those in carriers are more likely to be fraudulent than other items, but the legitimate volume of these is so substantial that it is infeasible to review them all for fraud.

Fraud-Detection Recommendations and Conclusions

The financial institutions surveyed indicated that to detect fraud and prevent loss, high-quality black-and-white images are sufficient to support their operations.

To support fraud-perpetrator identification, forensic examiners recommended that "the clarity and resolution of an image allow the examiner to see striations in ballpoint ink, that there be minimal pixilation of the writing line, that writing-pressure variations be visible, and that toner did not fill in the writing line so that disconnected writing looked connected." The examiners also recommended that scanners operate at a resolution that does not introduce interpolation into the image, and that scanners store images in a "lossless" format.

In conclusion, our survey indicated that given high-quality images, financial institutions' fraud-and-loss department staff and forensic document examiners are able to fulfill their functions in an imaged environment. As a result of their input, the project team assigned a high priority to testing the image usability of important fraud-detection fields.

PROPOSED NEXT STEPS

FSTC and its project members recognize that the industry has much yet to learn about image-quality defects and their relationship to usability, therefore the project team recommends that activities and discussions about image quality and usability continue in order to advance industry understanding and effectiveness. FSTC's Check Standing Committee (Check SCOM) plans to establish regular conference calls to facilitate this.

These discussions should enable sharing of information and experiences regarding image-quality defect testing and image usability. Potential discussion topics may include:

- Whether to request that Viewpointe preserve the image-quality lab and images used in this study
- Rates of occurrence of image-quality defects
- Specific occurrences of usability issues and whether they were (or could have been) identified by image-quality defect testing
- Implementation strategies for image-quality defect testing
- Refinement of image-usability metrics
- Implementation strategies for image-usability testing
- Refinement of operational models for handling image-quality exceptions
- Extension of the usability scoring concept from the image-defect test to a general image-usability scoring method that the industry could apply to any image

In November 2005, ADI and FSTC will finish reanalyzing the edge-defect results. We will update the FSTC website with threshold recommendations for those metrics when they are available.

FSTC project members whose images were used in the study can work with Viewpointe to provide more in-depth manual review of the images flagged for usability issues in the study. This will enable us to associate a usability score more closely with a financial institution's decision to clear or accept an image with usability issues.

As part of the IQ&U phase-two project, FSTC developed two important assets with ongoing value to the industry:

- An accuracy deck
- A database of results from image-defect metric testing and methodical image review

FSTC will develop a process by which image-quality vendors can test their implementations against the accuracy deck. This testing will facilitate improved interoperability of the defect metrics within the vendor community. Financial institutions may wish to encourage their vendors to test their software against the accuracy deck.

FSTC and its members will also continue to analyze the data obtained through the image-quality metrics testing. This data may be used to:

- Refine thresholds for metrics based on different usability scores or reset threshold recommendations based on a general check-level usability score, should the industry adopt one
- Evaluate metric combinations to reduce suspect rates
- Further evaluate the Image Too Light metric in the context of check backs to understand better if any viable metric threshold can be established for check backs

In addition to the immediately available opportunities to tune the recommendations made in this report to specific operational environments, FSTC recommends that financial institutions collect and share their own metric measurement results so that threshold values can be refined based on production rates of occurrence and usability impacts observed through suspect-review processes. After gaining sufficient experience (which the team estimates will take at least six months), financial institutions can determine whether the thresholds recommended in this project require any changes.

Once the industry has gained additional experience with the current image-quality defect metrics, FSTC recommends that FSTC members determine if further study is justified. This additional study could take several forms, including:

- A similar four-step effort to ensure interoperable usability metrics and implementations.
- A full analysis of the root causes of image-usability issues.
- A comparative analysis of grayscale and black-and-white images of the same "problematic" source documents to validate the assumption that maintaining a grayscale image will protect financial institutions from image usability issues caused by faint text and low-contrast printing

Should the industry wish to pursue either of the first two directions, it can leverage the images that have already been scored in this study.

FSTC and ECCHO will continue their close working relationship to:

- Update ECCHO rules (as appropriate) to reflect the evolution of industry terminology and definitions for legibility
- Update ECCHO rules to provide specific guidance for financial institutions regarding the warranty that an image is of "good quality"

FSTC and its members will submit fifteen of the sixteen image-quality defect metrics to ASC X9 for inclusion in the newly created image-quality test registry. The Carbon Strip Detected metric will not be submitted to the ASC X9 registry.

In addition, when ready, FSTC and its members will submit the image-usability metrics developed during this phase of the project to ASC X9.

Finally, FSTC and its members will continue working on issues that affect the quality and usability of check images. Together, through a formal or informal process, the industry will work to improve continuously the quality of images.

APPENDICES

APPENDIX A: GLOSSARY

The FSTC IQ&U project team determined, consistent with the project objectives, that common definitions were important for key terms being used in the project and in the industry. Toward that end, the project team reviewed available definitions, and where no suitable definitions were identified, established and agreed upon new definitions.

Definitions from Phase One

The following items represent the definitions for critical terms agreed to by the project team during phase one of the initiative. The team did not change the definitions during phase two.

(Check) Image

Definition: A digital representation of all or part of a physical item, including any associated parameters required to interpret the digital representation. The digital representation is created by sensing light reflected from the item[12].

Commentary: An image is a digital representation of an underlying (front and/or back) source document. A Check Image is a digital representation of a source document where that document was a check. An image taken using special equipment may contain information which is not visible to the human eye.

Image Faithfulness

Definition: The accuracy and completeness of a digital representation of the information and graphic details contained within the source document.

Commentary: Image Faithfulness is a narrower term than Image Quality, as faithfulness relates to the degree of preserved detail of both informational and noninformational faithful image to satisfy most business requirements.

Different image technologies have different intrinsic degrees of faithfulness. A proper black and white image of a standards compliant check will not contain background details, for example. Generally, the faithfulness or fidelity of an image will improve with an increase in the bit depth (number of colors or shades of gray) and resolution (measured in dots per inch). A very high resolution color image could achieve the highest degrees of faithfulness to the source document, up to being virtually indistinguishable from the original.

(Check) Image Defect Assessment

Definition: An analysis of an image of a source document based upon an established defect list. Defects will be assumed to be present when metrics exceed industry threshold values.

Commentary: Image defect assessment is measured using a set of image defect metrics. These metrics may measure characteristics of the source document (e.g. skew), or characteristics of the image itself (e.g. too few or too many pixels). Image defect assessment must be able to be performed reasonably accurately without reference to the source document.

[12] Source: ASC X9B TG-100

(Check) Image Defect Metrics

Definition: The set of measures used to quantify the likelihood that a digital check image has conditions that would render the information contained within the source document unusable in the image.

***Commentary*:** Image Defect Metrics are the measurements defined by the FSTC IQ&U Project to permit the determination and description of the condition of a check image. These metrics provide a mechanism for describing the condition of the document/image as a whole with the intent of determining the probability of an image being good enough to satisfy the four required business uses ((A) collections, exchange, paying, (B) exceptions and returns, (C) fraud detection and loss prevention, and (D) customer uses).

Many of the metrics identify conditions which are related to a potential defect in the image of the source document. For each defined Image Defect Metric, a material defect exists if the measurement indicates a high probability that a loss of information has occurred between the source document and the image. The specific values for these measurements that may render the check information unusable will be determined in a subsequent project.

Image Quality

Definition: The totality of characteristics of an image that bear on its ability to satisfy stated or implied needs.

Commentary: Image quality is defined as the totality of characteristics that bear on the ability of an image to meet the "needs", or to satisfy the business uses for which the image is necessary. Image quality may be described by a set of metrics to identify image defects, the presence and legibility of information, and capture characteristics. The work of the FSTC Phase 1 Image Quality and Usability Project has focused on specifying metrics to quantify the presence and extent of defects in an image, particularly where those defects may impact the ability of the image to satisfy the business needs (usability).

For the purposes of satisfying the requirements under the Check 21 Act, the information on the check is what is important. A high quality image will, by virtue of its quality, provide an accurate and sufficient representation of the information on the original source document.

This definition was derived from a definition of quality originally included in ISO 8402 "Quality management and quality assurance – Vocabulary" (now superseded by ISO 9000:2000). The FSTC project team has adapted this definition to be specific to images.

Image Usability Assessment

Definition: An analysis of an image of a source document to determine the likely usability of selected information contained within the image.

***Commentary*:** Image usability assessment specific metrics have not yet been assigned. These metrics measure the legibility of specific fields of information on the check, and may be a result of a specific assessment process, or may be a byproduct of other processes, such as automated recognition.

Image Usability Metrics

Definition: The set of measures used to determine the likelihood and extent to which the information in an image is usable if it was present on the source document.

Commentary: Image Usability Metrics are the measurements to permit the determination and description of legibility of key information in a check image. Usability metrics may also define alternate aspects of legibility, such as might be required to support an automated character recognition process, or reading of a barcode from a document. Usability Metrics are to be defined to (a) identify the absence of sufficient data in an image when the data was likely present on the source document, and (b) to assess the likelihood that important information in an image is sufficiently legible to meet one or more of the four defined business uses ((A) collections, exchange, paying, (B) exceptions and returns, (C) loss prevention, and (D) customer uses).

Definitions Revised in Phase Two

The following definition was changed during phase two:

Legibility

New Definition: Distinctness that makes perception easy.[13]

Phase-One Definition (Obsolete): The ability of a human viewer to decipher the information in a digital representation of a source document.

Old Commentary: Although the definition of legibility refers to the ability of a human viewer to decipher the information from a check image, other degrees of legibility may exist, such as necessary for automated character recognition. During phase one of the IQ&U project, FSTC did not identify any specific measurement that identifies the legibility of text without using an automated recognition process and obtaining a confidence rating.

New Commentary*:* There were two major reasons the project team revised the original definition:

To allow the same term to be used for text and nontext fields

To implicitly support application of a scale, for example, degrees of distinctness

FSTC believes that the new definition of legibility applies to both textual and graphic elements as well as to a scale that distinguishes among degrees of legibility. Both of these became important as the project team began formulating an in-depth definition of usability and how it might be measured.

[13] Source: WordNet ® 2.0, © 2003 Princeton University

Definitions from Phase Two

Capture Rate

Definition: The percentage of suspects that have usability issues (from all causes).

Commentary: The capture rate represents the percentage of suspects that are identified by a single metric or combination of metrics and have usability scores equal to or worse than the target usability score or point at which an image is agreed to be unusable. It is equal to 1 – escape rate.

Escape

Definition: An image with usability issues that is not identified by a defect measurement.

Commentary: It is important to note that escapes may occur for reasons entirely unrelated to any image-quality defect.

Escape Rate

Definition: Escapes as a percentage of the total sample size.

False Positive

Definition: An image incorrectly identified as a suspect.

Commentary: False positives are an important factor in evaluating the effectiveness of individual metrics. In this FSTC study, we evaluate a false positive on the basis of whether the image that exceeded a given metric threshold actually had a usability score worse than the usability target level.

False-Positive Rate

Definition: The proportion of suspects generated by a specific metric that were found to not actually exceed the target usability score.

Precision

Definition: The percentage of suspects generated by a metric that have usability issues.

Commentary: This represents the percentage of images that exceeded a metric threshold and were found to have usability scores worse than the target score, with the calculation limited to just images identified by that metric. Metrics that show higher precision are those that accurately identify images with usability issues above the target score without generating a large number of false positives.

Suspect

Definition: An image identified by a defect measurement as possibly having usability issues. A suspect is identified when its measurement value exceeds a threshold.

Commentary: Identification of suspects is dependent on the threshold values that are established and the target usability score. By varying the target usability score, metric thresholds, and metrics selected, institutions can manage the number of suspects.

For many image-quality defect metrics, there is a high rate of suspects and relatively low precision. For these metrics, financial institutions should monitor suspect rates and may wish to randomly sample a percentage of suspects to monitor for image-quality problems, which may affect the usability of images.

Suspect Rate

Definition: Suspects as a percentage of total sample size.

Commentary: It is important to note that just because an image is classified as a suspect does not mean that a financial institution must take specific action to investigate that suspect and determine if the image contains a true usability issue or represents a false positive. Instead, for many metrics, financial institutions should monitor the suspect rate to look for spikes as well as trends over time.

(Image-Defect Metric) Threshold

Description: The point of measurement that marks the boundary between images that are presumed to not contain a defect and images that contain a defect that may affect the usability of the image.

Commentary: Thresholds are fundamental points of measurement above (and below) which a metric value is believed to be potentially indicative of an image-quality defect that may affect the usability of an image.

For image-quality defect metrics with low precision, FSTC considers a threshold as the breakpoint when monitoring trends in order to identify potential problems with image-capture systems or source documents. This monitoring should be for both trends (increasing numbers of images exceeding thresholds over time) and spikes (sudden bursts of images exceeding thresholds in a short period of time).

Establishing a threshold is dependent on balancing the interaction among the precision, suspect rate, and target usability score. If a financial institution establishes suspects at extreme ends of image-defect metric measurements, then the escape rate may be unacceptable. If a financial institution establishes thresholds too far from the limits of image-defect metric measurements, then the suspect rate and precision of the metric may be unacceptable. We also note that thresholds can be set within a range, for example, where image-quality defect metrics are more successful at identifying images with usability scores above the target level within a range of measurement values as opposed to at the extremes.

Thresholds are ideally set at a tipping point where the best balance among suspects, precision, and escapes can be obtained for a given usability level. In addition, when a threshold is set, it is desirable if the precision of measurement increases with increasing usability scores, for example, the metric finds higher percentages of images with worse usability.

Usability Score

Definition: A score to distinguish among images with various levels of usability issues.

Commentary: FSTC developed a score to grade the degree of usability issues within an image and enable analysis of image metrics to determine the effectiveness of the image-quality defect metrics in identifying images with usability issues, especially significant usability issues.

The usability score is derived from the field-level usability information collected within the manual review process, and although generally reasonable, is specific to this analysis.

APPENDIX B: ADI STATEMENT OF VALIDITY OF METHODOLOGY

Background on the Experimental Design

This project represented some significant challenges driven by the expectation of a very low (1:10,000) error signal (estimated unusable check images) as compared to the sample population. The other challenge was that we could not manually evaluate each of the 3.4 million items in the image sample. In a nutshell, the challenge can be described as finding a “needle in a haystack” without operational capacity to review every image.

A standard approach would be to sample randomly within this population. The shortfall for this application would be the concern of not finding enough “unusable check images” with which to conduct the analysis of utilizing quality metrics to predict usability and assess the effectiveness of those quality metrics.

Our data acquisition approach was to include a combination of targeted and random sampling. The targeted samples were identified through the use of check-usability assessment software, which identified 500,000 images, and a “bottom feed” from the FSTC quality metrics, which identified 120,000 images. We then randomly sampled 100,000 images within the “good” population. ADI then manually reviewed 100 percent of the images within these sample sets.

ADI conducted the manual review with the intent to establish presence and legibility data on a check image down to the field level. First, the team reviewed images using a rapid-fire process in which images were quickly rated as “All Fields Legible” or “Not All Fields Legible.” Reviewers filtered images that did not pass this test to a methodical review process where they reviewed eight fields and graded for presence and legibility. From this data, legibility for each field can be deduced for the eight fields of every check image manually reviewed. Note that the reviewers identified and removed nonchecks from the test through this process.

The importance of the field-level scoring cannot be understated, as usability is defined in terms of a combination of field presence and legibility. This enables the labeling of any scored image as usable or not based on any definition of usability within the context of the eight fields. This discussion of the appropriate definition of unusable is still going on today. For the purposes of this analysis, the team defined unusability as any check image that had at least two fields of compromised legibility, with at least one of those fields clearly not legible (note that we conducted “tests of reasonableness” at increasingly worse unusability levels with performance trends as desired and expected). To enable evaluation and study of check images at various degrees of unusability, the team created a scoring system in which two points were given for any field rated “No, Partially Legible” and three points for any field rated “No, Clearly Not Legible.” A check-image score could be a value of 0 (good image) and 2-24, with a 24 being a check in which all eight fields were rated “Clearly Not Legible.” We refer to this internally as the NG score, with higher numbers representing worse usability scores. This has been a significant enabler of all the discussions and debates of appropriate definitions of un-usability, in that specific analysis at any of these levels is possible within the data set.

Upon receipt of the quality metric data and having completed our own usability data acquisition, we started the process of understanding the relationship between the quality metric data and usability data. First, we used visual techniques between usability labels and quality metric data with scatter plots, probing at metric versus NG label, metric versus NG score (as this was initially specified as NG >= 2 by the FSTC, and later landing on NG >= 5 to conduct the analysis), and metric versus metric. This effort was not trivial and certainly provided the feel and insight with which to conduct the specific analysis.

We then evaluated each metric (and some combinations) in regard to its ability to detect unusable check images, with the judgment criterion of precision, suspect rate, and escape rate. This was communicated via a Precision-Escape chart and a specific Confusion Matrix to the selected operating point. Operating points (tradeoff among precision, suspect rate, and escape rate) were selected manually applying the following context and logic.

Context:

FSTC communicated the following target overall values to ADI:

- Precision > 10%,
- Escape rate < 75%,
- Suspect rate < 2%, target 0.2%

Logic:

There was an immediate indication that Black Pixel Over was very effective as a predictor of nonusability. We established the Black Pixel Over threshold based on what we thought was a reasonable balance between the precision, escape, and suspect rates. We saw we could do much better than the target numbers. The choice then was to get a good "shovelful" of these images while maintaining relatively good precision, because this was the metric that was going to drive the overall captures. In any case, the precision/escape curve could allow financial institutions to select an operating point anywhere on the curve, depending on their needs.

Once we established this, we evaluated all other metrics within THIS context. While more shovelfuls were not in the offing (except for the Mismatch/Image Height due to the MICR Line issue, now discounted), we were then looking to add useful "tweezerfuls." This meant evaluating metrics with the desire to add value overall, even if small. This is why we began looking at the very tails of the sample distributions to see if anything was there (for Black Pixel Under, there was). This approach would drive a very low suspect rate (as desired). In the cases where there was no correlation using the tails, we looked closer to the median until something demonstrated a correlation, even if it had a very low precision (and we listed this as the threshold). We did not expect these thresholds to stay "in." Within this context, precision rates even in the 10-20% range would not be attractive because they would drag down the overall performance. Financial institutions would be better off reducing the precision slightly on Black Pixel Over, to get more captures, comparatively speaking.

It should also be noted here that our efforts emphasized qualification rather than disqualification when evaluating the metrics. This gets back to the "finding a needle in a haystack" mentality. When we found evidence of a metric being useful, we dug further to qualify and quantify the relationships affected by that metric. If a metric did not prove useful after basic analysis, we tended to not explore it further.

For these evaluations to represent "realistic" performance, we needed to create a test sample that represented all 3 million images (3.4 million minus the nonstandard images). We created two test sets using the following formulas:

1 For the larger set: (Presort Bad + FSTC Bottom Feed) + Presort Good x 29

2 For the smaller set: (Presort Bad + FSTC Bottom Feed) / 29 + Presort Good

The idea was to count the targeted samples once and the random samples multiple times for the larger set. For the smaller set, it was to count the random samples once but only a random fraction of the targeted samples.

Introduction Regarding Data Quality

In order to study the quality of the data and results reported herein, we will look at three primary types of errors:

1 Sampling error

2 "Truth" or labeling error

3 Classifier-performance error

The first, sampling error, is easily handled with standard sampling statistics. The second, "truth" error, is basically a measurement error, which is controlled by performing certain tasks twice and then adjudicating the differences to determine the "truth." By using the "truth" to score one or both of the data sets, an estimate of measurement error may be made. If the corrected "truth" is used as the data, then the measurement error is correspondingly reduced. The classifier-performance error, the third type of error, is described by a classic technique called a "confusion matrix," which is used extensively in this analysis.

Sampling Error

Basics

When one has a large population in which each element of the population is either a "success" or a "failure," then by suitably measuring each element and labeling each outcome as a success or failure, one can calculate the failure rate as the number of failures divided by the total population size. As often happens, one cannot afford to measure every element of a large population, so one resorts to "sampling" to make an estimate of the failure rate. When this is done, then there is an error associated with the resultant estimate of failure rate, usually called sampling error. This error may be made acceptably small by choosing a sufficiently large number of samples. Basic sampling theory (See Dillman and/or Meyer references) describes what the sampling error is for sampling a population to estimate a failure rate.

For large populations, the sampling equation in Dillman reduces to:

$$\sigma^2 = pq/n$$

where

σ = the standard deviation of the error due to sampling

p = the probability of a successful outcome

q = 1 - p = the probability of a failure

n = the number of samples taken from the population.

Note that we take two sigma values to represent approximately 95% confidence in the above formula, and also note that the formula is indifferent as to whether one assigns a "success" to be a "p" or a "q," as the product "pq" is used.

Example–A Gallup Poll

In order to understand elementary sampling theory, a simple example is helpful. In June 2005, a Gallup Poll reported that 56% of Americans did not think it was worth going to war in Iraq. The randomly selected sample of the U.S. population was 1,002 adults. The surveyors claimed that, with 95% confidence, the sampling error was +/- three percentage points. In this simple example, we have p = 0.56, q = 0.44, and n = 1,002. Using the above formula, we get σ = 0.0157, and using two sigmas to correspond to 95% confidence, we get 2σ = 0.0314, which corresponds closely to the claimed three percentage points of sampling error. This means that the quoted 56% could actually be as high as 59% or as low as 53%, with a confidence level of 95%.

Our Initial Estimates of Sample Size

When ADI, LLC first began to consider the problem of assessing the "quality" of check images in a large archive in early 2004, we were asked to study the archive to determine what the "failure rate" was in terms of image quality. At that time, the total size of the archive was in the tens of billions, so clearly sampling was called for. Our starting assumption was that the suspect rate was 0.2% and that the failure rate was 1% of the suspect rate, or q = 0.00002. (This is a very low failure rate for almost any system, because it implies that the success rate is p = 0.99998. As an example, for the United States Census 2000 data-capture system, which ADI helped design and test, we had about a 99.8% success rate for automated optical mark recognition, or OMR. OMR is used to determine a respondent's answer to a multiple-choice "check-box" question, which we consider to be a simpler problem than getting all the fields on a check to be clearly legible.) Another conservative assumption was that the resultant standard error should be less than 1/10 of the failure rate being estimated, that is, σ = 0.000002. Using the above sampling equation solved for sample size n gives

$$n = 0.99998 \text{ x } 0.00002/(0.000002 \text{ x } 0.000002)$$

$$= 4{,}999{,}900$$

So this is where the original estimate of a sample of five million images of checks came from. Viewpointe then went into its archives and pulled about five million check images from ten banks, weighted by volume. Before ADI had viewing access through our secure VPN link to Viewpointe, it was determined that about one million of the pulled images were "nonchecks," and so they were removed from the sample. This left about four million check images in the sample. Finally, two of the ten banks "bowed out" of the FSTC process so that we wound up with about 3.6 million check images to study. Fortunately, this is still a great deal of images, and if we use n = 3,600,000, with p and q as above and solve for the standard deviation of the mean failure rate, we get about 12% of the assumed failure rate as opposed to 10%, which still allows for very good estimates.

If the previous assumptions were correct, we would expect only 100 NG images out of five million, or one out of 50,000. By measurement of usability, we got about 14,000 images that were labeled NG [>=5] out of three million, which is one out of 214 (about 0.5%). One could see results like those assumed if one wished to work at the "badness" scale of NG [>=19] or below.

Sampling Error in this Study

Here, we use a slight modification of the above formula to account for populations that are not very large relative to the sample size, written as:

$$\sigma^2 = pq[1/N_s - 1/N_p]$$

where N_s = sample size and N_p = population size.

Note that for large population size, the formula reverts to the one used earlier, as it should. Also note that if $N_s = N_p$, then there is no sampling error (because we sampled them all).

We may indicate sampling error in this study by considering, for example, the number of check images labeled NG [>=5] on the "usability" scale, which was identified as 13,410 (this includes NG>=5 + Piggybacks + No Key Data). This number is not really accurate to five places, of course, but how good is it? To understand this, we consider two main subpopulations: "Presort Rated Goods" and "Presort Rated Bads." The latter group breaks into two subgroups from the rapid-fire 1 review: "perfect" and "not perfect."

Presort Rated Goods

We had a population of N_p = 2,774,869 check images that our presort process rated "good." We sampled N_s = 93,700 of them to see how many were "not good" at the usability scale of NG [>=5]. We found 117 images rated NG [>=5], so that q = 117/93,700 = 0.00125, or 0.125%. The resultant two-sigma sampling error in estimating q is then 0.00023.

Presort Rated Bads (Not Perfect)

We had a population of N_p = 252,259 check images that our presort process rated "bad." All these were put through rapid-fire 1 review, wherein 198,171 were labeled "perfect" and 55,058 were labeled "not perfect." (of these an additional 970 images were found to be non-checks). All 55,058 "not perfect" images were run through methodical review, and we found 9,795 check images rated NG [>=5]. Because we measured all of the "not perfect" images with the methodical review, there is zero sampling error for that group.

Presort Rated Bads (Perfect)

Of the 198,171 images labeled "perfect," we (re)sampled 25,672 and found 40 images rated at NG [>=5], so N_p = 198,171; N_s = 25,672; and q = 40/25,672 = 0.00156. The resultant two-sigma sampling error in estimating q is then 0.00046. (Note that this error is a measurement error, also called "truth" or labeling error. We discuss this more in the next section. Here we are focusing on the sampling error).

Roll-Up of Sampling Error for NG [>=5 + Piggybacks and No Key Data]

Combining the results from the previous three paragraphs, we can state the following, with 95% confidence:

- Presorted Goods
  - ✓ 3,468+/- 638
- Presorted Bads "Not perfect"
  - ✓ 9,945 +/- 0
- Presorted Bads "Perfect"
  - ✓ 0 +/- 91
- Total: 13,413 +/- 729 (0.44% +/- 0.024%)

This means that our estimate of the number of images which are rated NG [>=5] in the entire population of 3,027,128 images is 13,413, plus or minus 729, with 95% confidence, considering sampling error. In percentage terms, it means that the percent of NG [>=5+ Piggybacks and No Key Data] images was 0.44%, plus or minus 0.024%. This amount of sampling error is 729/13,413 = 0.0537, or about 5.4%.

"Truth" (or Labeling) Error

We showed above that some images in the "Presort Rated Bads" group were labeled "perfect" after methodical review. We resampled 25,672 of the 198,171 images and found 40 of them rated NG [>=5]. This is a labeling error and is quite small compared to sampling error. This labeling error is 0.156%, which is about 35 times smaller than the 5.5% sampling error.

We obtained another data point on "truth" error when we concentrated on 15,042 fields labeled "No, Clearly." Because these were very important data to get right in this analysis, we made two independent passes at rating these labels, and found 1,063 field discrepancies (7%). This meant each pass had about a 3.5% error, assuming independence of the errors. This result is very reasonable and is compatible with our experience with human keying of write-in fields in applications such as the U. S. Census. We then looked at all the discrepancies and reconciled the errors; this process we assume also had a 3.5% error associated with it. Therefore, the resultant truth error in this case would be about 3.5% x 7% = 0.25%, over twenty times smaller than sampling error.

As "No, Partially" contributes to our definitions of NG >= 5, we need to consider the labeling error in this inherently ambiguous rating. Although we can measure error at the field level, the important thing to consider is how it drives a check image to be rated NG or not. Consider the following table:

Figure B-1. "No, Partially" Truth Error Analysis

How Many Fields Judged "No, Partially" That Had Two Reviews?	32,928		
How Many "No, Partially" Discrepancies in Images That Had Two Passes Through Rapid-Fire 1/Methodical?	9,696		
% of "No, Partially" Fields Discrepancy	29.4%		
Assumed Error in a Single Pass	14.7%		
NG>= 5 Case		To Move To NG < 5, How Many "No, Partially" Labels Need to be Wrong?	Labeling Error (Count x Error^Need to be Wrong)
How Many NG=5 Images?	2,313	1	341
How Many NG=6 Images w/3 Fields "No, Partially"?	1,959	1	288
How Many NG=7 Images w/2 Fields "No, Partially"?	1,056	2	23
How Many NG=8 Images w/4 Fields "No, Partially"?	776	2	17
How Many NG=10 Images w/5 Fields "No, Partially"?	189	3	1
How Many NG=11 w/4 Fields "No, Partially"?	147	4	0
How Many NG =12 w/6 Fields "No, Partially"?	43	4	0
Sum			670
NG>=5 in RF1 Methodical			9,795

Also to note that as the analysis is done for increasing NG levels, thus with a greater reliance on "No, Clearly" (versus "No, Partially") to drive NG, the labeling error decreases.

Updating the results from the sampling error to include labeling error, we get:

- Presorted Goods
  - ✓ 3,468 +/- 638
- Presorted Bads "Not perfect"
  - ✓ 9,945 +/- 670
- Presorted Bads "Perfect"
  - ✓ 0 +/- 91
- Total: 13,413 +/- 1,399 (0.48% +/- 0.05%)

This means that our estimate of the number of images which are rated NG [>=5 + Piggybacks and No Key Data] in the entire population of 3,027,128 images is 13,413, plus or minus 1,399, with 95% confidence. In percentage terms, it means that the percent of NG [>=5 + Piggybacks and No Key Data] images was 0.44%, plus or minus 0.0462%. This amount of total error is 1,399/13,413 = 0.1043, or about 10.4%.

In ordinary language, one would say that about 13,000 of the three million images in the population were rated NG [>=5 + Piggybacks and No Key Data]

Classifier-Performance Error

We used a classic technique called a "confusion matrix" to determine how well our classifier (nonusability predictor) measured the relationship between the quality-metric data and the usability data. Refer to the confusion matrix charts in the separate ADI final briefing report to FSTC for clear definitions of the elements of the confusion matrix.

Below is the original confusion matrix for the Image Too Dark metric.

			a	b		
			c	d		
			Actual			
			OK	NG		
	Predicted	OK	105787	393	106180	(a+b)
		NG	40	71	111	(c+d)
			105827	464	106291	
			(a+c)	(b+d)		(a+b+c+d)
		Precision =	63.96%	d/(c+d)		
		Suspect Rate =	0.10%	(c+d)/(a+b+c+d)		
		Escape Rate =	84.70%	b/(b+d)		
		False Positive Suspect Rate =	36.04%	c/(c+d)		
		Accuracy =	99.6%	(a+d)/(a+b+c+d)		
		True Positive Rate =	15.3%	d/(b+d)		
		False Positive Rate =	0.0%	c/(a+c)		
		True Negative Rate =	100.0%	a/(a+c)		

Now applying the error in a most extreme way, 10.4% increase in total actual NG (b+d) AND a 10.4% reduction in captured NG (d), which then also drives a 14% increase in escapes, the results are as follows:

			a	b		
			c	d		
			Actual			
			OK	NG		
	Predicted	OK	105739	448	106187	(a+b)
		NG	40	64	104	(c+d)
			105779	512	106291	
			(a+c)	(b+d)		(a+b+c+d)
		Precision =	61.54%	d/(c+d)		
		Suspect Rate =	0.10%	(c+d)/(a+b+c+d)		
		Escape Rate =	87.50%	b/(b+d)		
		False Positive Suspect Rate =	38.46%	c/(c+d)		
		Accuracy =	99.5%	(a+d)/(a+b+c+d)		
		True Positive Rate =	12.5%	d/(b+d)		
		False Positive Rate =	0.0%	c/(a+c)		
		True Negative Rate =	100.0%	a/(a+c)		

Precision has reduced by 2.4 %
Escape rate has increased by 2.8%

This demonstrates the robustness of the result despite an extreme application of the NG count error (increasing total actual NG to the maximum and decreasing captured NG to the minimum).

Now, let's look at what has been called a "tweezerful" going through the same exercise.

Below is the confusion matrix for the Image Too Light metric.

			a	b		
			c	d		
			Actual			
			OK	NG		
	Predicted	OK	3067852	13650	3081502	(a+b)
		NG	18	46	64	(c+d)
			3067870	13696	3081566	
			(a+c)	(b+d)		(a+b+c+d)
Precision =			71.88%	d/(c+d)		
Suspect Rate =			0.002%	(c+d)/(a+b+c+d)		
Escape Rate =			99.66%	b/(b+d)		
False Positive Suspect Rate =			28.13%	c/(c+d)		
Accuracy =			99.6%	(a+d)/(a+b+c+d)		
True Positive Rate =			0.3%	d/(b+d)		
False Positive Rate =			0.0%	c/(a+c)		
True Negative Rate =			100.0%	a/(a+c)		

By applying the same 10.4% increase in total actual NG (b+d) AND a 10.4% reduction in captured NG (d), we get:

			a	b		
			c	d		
			Actual			
			OK	NG		
	Predicted	OK	3066428	15079	3081507	(a+b)
		NG	18	41	59	(c+d)
			3066446	15120	3081566	
			(a+c)	(b+d)		(a+b+c+d)
Precision =			69.49%	d/(c+d)		
Suspect Rate =			0.002%	(c+d)/(a+b+c+d)		
Escape Rate =			99.73%	b/(b+d)		
False Positive Suspect Rate =			30.51%	c/(c+d)		
Accuracy =			99.5%	(a+d)/(a+b+c+d)		
True Positive Rate =			0.3%	d/(b+d)		
False Positive Rate =			0.0%	c/(a+c)		
True Negative Rate =			100.0%	a/(a+c)		

Precision has reduced by 2.4%
Escape rate has increased by 0.07%

Again, this demonstrates the robustness of the result despite an extreme application of the NG count error (increasing total actual NG to the maximum and decreasing captured NG to the minimum).

These two examples demonstrate how the error estimate on NG count can propagate. We've applied the error in a worst-case fashion and shown the precision and escape rate not to be significantly sensitive. Note that as the performance decreases and the count of captured NG images approaches zero, validity of results can become problematic (for example, if precision is 50% based on predicting one NG out of two suspects correctly, then a change of one in the correct prediction of NG will drive significant percentage difference).

The methodology utilized for this project is sound and valid. In fact, it also enables unlimited relevant analysis possibilities considering the dynamics of what still is the core discussion about "what defines an unusable image." There should also be comfort in the fact that the methodology and the results have withstood scrutiny throughout the project by FSTC's industry experts.

References:

Dillman, Don A. *Mail and Internet Surveys*. New York: John Wiley & Sons, Inc., 2000.

Meyer, Stuart L. *Data Analysis for Scientists and Engineers*. New York: John Wiley & Sons, Inc., 1975.

APPENDIX C:
ATTRIBUTES AND INFORMATION ON A CHECK

The project team analyzed the contents of a check in detail to assess the importance of check attributes in an image-exchange environment. A key element of this analysis was to discriminate between "informational" and "noninformational" check attributes.

The team deemed this distinction important. Section XIX(A)(6) of ECCHO clearinghouse rules defines an image as "An accurate representation of the front and back of the Related Physical Check." Further, a presenting bank warrants[14] that the image meets both the accurate representation requirement and the image-quality requirements documented in the same section XIX.

In addition, the Check 21 legislation requires a reconverting bank to warrant that an IRD produced from an image "accurately represents all of the information on the front and back of the original check as of the time the original check was truncated,"[15] requiring that any bank that captures an image and truncates the item needs to capture an image suitable for creation of IRDs.

The project team feels that the distinction drawn in Check 21 of "all of the information" on the check is an important one. It allows for certain check design elements, such as backgrounds, to not be present in an image and not result in a breach of warranty. This reduced the number of possible check elements that could lead to a breach of warranty claim.

The team also analyzed each check attribute to determine its typical frequency and use. Further, the team determined if an attribute that is present on the source document but absent in an image is likely to increase the risk of a claim for a loss.

The analysis enabled the project team to recommend attributes within an imaged check that are the most promising candidates for usability examination. This subset of check attributes is described in the "Check-Image Usability Testing Guidelines" section.

The following explains and documents the results of the project team's analysis. Note that the analysis applies only to imaged checks, not to checks converted to ACH or other transaction vehicles.

[14] ECCHO Rules § XIX (M) & (Q)
[15] Check Clearing for the 21st Century Act–Public Law 108-100; 12 USC 5003 §4(b)(1)

APPENDIX D: THE CRITICAL ISSUE — "ACCURATELY REPRESENTS ALL OF THE INFORMATION"

There has been much discussion within the industry regarding what exactly constitutes an accurate representation of all information. The Federal Reserve has attempted to address this confusion, in part, with clarifications and commentary on the Check 21 regulations. This guidance may be found at:

http://www.federalreserve.gov/boarddocs/press/bcreg/2004/20040726/attachment.pdf

The key language is repeated below, with highlighting added by FSTC, for the convenience of the reader.

Page 22

Other commenters expressed concern about whether the accuracy requirement for legal equivalence would be met if the drawer or a bank applied payment information to the check using an ink color or ink type that would not survive the image capturing process. The commentary to the final rule clarifies that ***payment information always must be accurately represented*** *on a substitute check because that information is an essential element of a negotiable instrument.* ***If a substitute check failed the legal equivalence requirement because of ink choice or some other feature, such as check color or a decorative image, the reconverting bank would be responsible for associated liabilities.***

Pages 94-95 – Number 3

3. To be the legal equivalent of the original check, a substitute check ***must accurately represent all the information on the front and back of the check*** *as of the time the original check was truncated.* ***An accurate representation of information that was illegible on the original check would satisfy this requirement. The payment instructions*** *placed on the check by, or as authorized by, the drawer, such as the amount of the check, the payee, and the drawer's signature, must be accurately represented, because that information is an essential element of a negotiable instrument.* ***Other information that must be accurately represented*** *includes (1) the information identifying the drawer and the paying bank that is preprinted on the check, including the MICR line; and (2) other information placed on the check prior to the time an image of the check is captured, such as any required identification written on the front of the check and any indorsements applied to the back of the check.* ***A substitute check need not capture other characteristics of the check, such as watermarks, microprinting, or other physical security features that cannot survive the imaging process or decorative images, in order to meet the accuracy requirement.*** *Conversely, some security features that are latent on the original check might become visible as a result of the check imaging process. For example, the original check might have a faint representation of the word "void" that will appear more clearly on a photocopied or electronic image of the check. Provided the inclusion of the clearer version of the word on the image used to create a substitute check did not obscure the required information listed above, a substitute check that contained such information could be the legal equivalent of an original check under § 229.51(a). However, if a person suffered a loss due to receipt of such a substitute check instead of the original check, that person could have an indemnity claim under § 229.53 and, in the case of a consumer, an expedited recredit claim under § 229.54.*

To better understand the guidance that the Federal Reserve is providing, the project team summarized the guidance into the following five categories of information. The project team then incorporated these categories into the detailed field-level analysis of the contents of a check.

Figure D-1. Federal Reserve Guidance Summary

Category	Must Preserve	Federal Reserve Description
F1	Yes	Payment Instructions – amount, payee, drawer's signature.
F2	Yes	Other Information (1) – drawer and paying bank info, MICR line.
F3	Yes	Other Information (2) – other information placed on the check prior to the time an image of the check is captured. Specifically including any identification and endorsements.
F4	No	Other Characteristics (1) – watermarks, microprinting, other physical security features that cannot survive the imaging process.
F5	No	Other Characteristics (2) – decorative images.

One question not answered by the above is whether an accurate representation may have information added to it when compared with the original check. Per informal discussions with the Federal Reserve, information may be added to a check image and the image may still be an accurate representation as long as the information added does not obscure or alter information present prior to the new information being added.

Description of Columns

The table below does not specifically address either IRDs or the potential for information overlays into a check image.

Figure D-2. Column Descriptions

<table>
<tr><th>Column</th><th>Description</th></tr>
<tr><td>Frequency</td><td>An assessment of how often, by percentage of checks, an attribute of design or a data field appears on checks. This will help to prioritize the importance of attributes. We assessed frequency using the following scale:
<table>
<tr><th>Frequency</th><th>Estimated % of Checks</th></tr>
<tr><td>Very High</td><td>> 90</td></tr>
<tr><td>High</td><td>75 - 90</td></tr>
<tr><td>Medium</td><td>25 - 75</td></tr>
<tr><td>Low</td><td>< 25</td></tr>
</table></td></tr>
<tr><td>Check Design Attribute</td><td>Various aspects of a check, including background patterns, size, color, borders, embedded security features, and optional text</td></tr>
<tr><td>Expect to Survive</td><td>Data that is present and printed on the check in such a manner that the</td></tr>
</table>

Column	Description
BW Imaging (Black-and-White)	printing exceeds the requirements established in ANSI X9.7 for image survivability, even if the data is not printed in an "area of interest" defined by that standard. An ✖ contained in the column "Expect to Survive BW Imaging" indicates that this attribute is generally not expected to survive in a black-and-white image.
Required to be Negotiable	Fields that are required to make a check a legal negotiable instrument as defined in UCC §3-104. It is important to note that a check may be paid, and correctly so, without all of these attributes being present.
Used by FI Pay Process	Fields that are frequently used or required for payment and handling by banking operations and/or Regulation CC or an appropriate government circular. FSTC defined three broad categories where banking functions require payment related data. These were: collections, clearing, and posting; exceptions and returns; and fraud detection and loss prevention.
Needed to Prove Payment	Information on or added to the check which, if present, can be used by a customer to prove payment to a beneficiary, or for a beneficiary to properly identify how it processed a payment that it received. Note: This column documents a working set of assumptions. It has not been vetted with legal opinions. In many cases, there is no settled case law to guide the analysis.
Breach of RB/TB Warranty (Reconverting Bank/ Truncating Bank)	Indicates whether absence of a field in an image (when present on the source) may be considered a technical breach of warranty, without regard to whether there is a loss. Note: This column documents a working set of assumptions. It does not represent validated legal opinions.
May Affect Claim	Fields that are deemed most likely to introduce liability for a truncating bank in the event of a customer claim of loss. In order for the truncating bank to accrue a liability the: (a) information in the field must be present on the check, (b) information must be lost in the imaging process, (c) customer must experience a loss, and (d) customer must file a claim. These claims are exclusive to determining whether an item was properly payable. This does not include claims between a maker and a beneficiary over quality of goods and services, lack of delivery, or other non-payment-related issues.
X9.37 Usage	Fields that are "mandatory" or "conditional" per the DSTU X9.37-2003 specification. This analysis includes record types 25, 26, 28, 31, and 33.
When Added to Check	This column indicates at what point in the life of a paper check the attribute is most frequently added. See the section below, which details

Column	Description
	the information life cycle of a check.
Federal Reserve Category	A categorization of the check attribute to determine whether the Federal Reserve's guidance indicates that the attribute must remain intact in the image in order for the image to "accurately represent all of the information on the front and back of the original check as of the time the original check was truncated."
Use Level I	Collections/Exchange/Posting--Fields that are required for basic collections, exchange, and posting.
Use Level II	Exceptions/Returns--Fields that are required for exceptions and returns. This includes all payment-related information on the front and back of the check.
Use Level III	Fraud Detection/Loss Prevention--Fields that are required for fraud detection and loss prevention. This includes all payment-related information on the front and back of the check, as well as security features designed to survive black-and-white imaging, authentication data, and noninformational characteristics of payment-related data used for identification and comparison of fonts and other attributes.
Use Level IV	Customer Uses—Field analysis for all customer uses. This includes all information on the front and back of the check, including hand-added data not typically used by financial institutions.

Certain check attributes may appear on both the front and the back of the check, such as safety backgrounds. In these cases, we list these items in the table one time, for the front of the check only.

Legend for Use Levels

Financial institutions should view all legends in the context of information contained on the original check. In defining image usability, the primary concern is if information is present on the physical check but "lost" or damaged beyond use in the imaging process. It is this loss of information that gives rise to potential liability. If the information is lost from the check image, then the truncating bank will be unable to provide proof of the original contents. It is, however, difficult to distinguish a usability problem caused by the condition of the source document from a usability problem unique to the image. Where possible, we assume that the data was present and legible on the source document.

Figure D-3. Use Level Legend

Code	Description
M	Mandatory--Field is critical to business processes. Absence or illegibility may make the image unusable for this process. Financial institutions are interested in being warned of all cases of missing or illegible data for a mandatory field.
C	Conditional--Information that is important to support the business process and is frequently present. The information may be obtained from more than one source, on

Code	Description
	or off the check. See the "Conditional Fields in the Check Image" section of this document. Absence or illegibility of the information may not make a check unusable. Financial institutions are interested in having control over notification of potential absence or legibility problems with the individual fields of information. For X9.37 analysis, the following codes are used: CP – Include if present unless agree to omit CA – Include if available unless otherwise agree to omit CO – Omit unless otherwise agree to include
O	Optional—Information or design attributes may be useful to the business process. The information may be obtained from more than one source, on or off the check, and is often not present in many checks. Absence of information will not typically affect the usability of a check. Illegibility may make the check unusable. Financial institutions will not generally be interested in testing for the absence of this information. They may be interested in illegibility (particularly if there is a way to identify it as due to the imaging process), and for selected accounts, should systems support it, they may be interested in both absence and illegibility.
N	Nonessential—The field or design attribute is typically not used in a banking process at this level. Absence of the field or loss of detail is unlikely to cause a loss or result in liability. Financial institutions will generally not be interested in testing for the presence or legibility of these attributes of a check.
B	Blank—The importance of the field or design attribute at this level has not been specifically determined. Generally the field is expected to be nonessential. Absence of the field or loss of detail is unlikely to cause a loss or result in liability. Financial institutions will generally not be interested in testing for the presence or legibility of these attributes of a check.

It is important to note that images will be captured only one time. Therefore, the party capturing the image is responsible for ensuring that the image can support all subsequent uses.

Conditional Fields in the Check Image

A number of fields on a check are "conditional" in that if the primary field is not present, one or more backup fields may be used to supply the data necessary to complete the transaction. The table below identifies related fields. Generally, if one of these fields is present, then a lack of legibility in the other field will not represent an image flaw that would result in an image being considered unusable.

Figure D-4. Conditional Fields

Front /Back	Primary Field	Front /Back	Related Fields on Check
F	Courtesy amount	F	Legal amount
F	MICR account number	F	Maker name/ address or account verification number
F	Aux on-us or on-us field of MICR line	F	Printed check serial number
F	Payee name	B	Payee name in endorsement
F	MICR routing and transit number	F	Fractional transit number Issuing bank name/address
B	Bank of first deposit RT	B	Bank of first deposit name

In the case of the payee name, a bank conducting research on an item may be able to use information from the associated deposit transaction.

Front /Back	Field	Front /Back	Related "Off-Check" Fields
F	Payee name	N/A	Depositor name or account number from deposit ticket or ATM transaction
F	Courtesy amount Legal amount	N/A	Amount field of ANS X9.37

Check Life Cycle: When information is added to the check

The table below outlines the order in which information and design attributes are added to paper checks and to images of those paper checks. Note, however, that some fields may be added at different times in the life cycle, depending on the business process. For example, on consumer checks the MICR line is added at personalization, but it is added to many business checks at issuance.

Figure D-5. Check Life Cycle Codes

Code	Description
S1	Manufacture--Creation of paper stock, but before printing.
S2	Personalization--Individual bank and customer information is added to the check. This information can be added at a check printer or when the check is issued by a business. It is generally preprinted data that does not change from check to check (except the serial number).
S3	Issuance--The check is completed with transaction information.
S4	Acceptance--The payee accepts the check.
S5	Deposit--The payee (or the party to whom they negotiated the check) deposits it.
S6	Clearing--The bank of first deposit places the check in the interbank clearing system for eventual payment by the account-owning institution.
S7	Payment--The account-owning institution accepts and pays the check if it determines the check is valid and the funds are available.
S8	Return--The account-owning institution returns the check to the collecting bank or the bank of first deposit if the check is determined to not be valid or funds are unavailable.

These codes are used in the analysis below to identify the point in the check life cycle at which certain pieces of information are most commonly added. With widespread adoption of image exchange, information addition to the check itself will stop in S4 or S5, and electronic records will maintain all information required to support subsequent processes. These electronic records are outside the scope of this analysis.

Figure D-6. Fields on a Check

Front of Check Attributes	Frequency	Check Design Attribute	Expect to Survive BW Imaging	Required to be Negotiable	Used by FI Pay Process	Needed to Prove Payment	Breach of RB/TB Warranty	May Affect Claim	X9.37 Usage	When Added to Check	Federal Reserve Category	Use Level I	Use Level II	Use Level III	Use Level IV
Check (serial) number (upper right)	VH	✓	✓		✓	✓	✓			S2 or S3	F3	O	C	C	
Label--Issue date	VH	✓	✓				✓			S2	F3		N		
Issue date	VH		✓	✓	✓	✓	✓	✓		S3	F1		C		
Payer name	VH	✓	✓		✓	✓	✓			S2	F2	O	C	C	
Payer address	VH	✓	✓		✓		✓			S2	F2		O	C	
Fractional transit number	VH	✓	✓		✓		✓			S2	F3	O	C	C	
Label--pay to the order of	VH	✓	✓	✓			✓	✓		S2	F1		M		
Payee name or names	VH		✓	✓	✓	✓	✓	✓	CA	S3	F1		C	C	
Other payee descriptive data	H		✓		✓	✓	✓			S3	F3		O	O	
Payment expiration information (e.g., void after 90 days)	L	✓	✓				✓	✓		S2	F3		O		
Account verification number	L	✓	✓		✓		✓			S2	F3		C		
Label - $ (symbol) or dollars	VH	✓	✓				✓			S2	F3		N		
Amount			✓						M		F1				
Convenience amount	VH		✓	✓	✓	✓	✓	✓		S3	F1	C	C	C	

Front of Check Attributes	Frequency	Check Design Attribute	Expect to Survive BW Imaging	Required to be Negotiable	Used by FI Pay Process	Needed to Prove Payment	Breach of RB/TB Warranty	May Affect Claim	X9.37 Usage	When Added to Check	Federal Reserve Category	Use Level I	Use Level II	Use Level III	Use Level IV
Legal amount	VH		✓	✓	✓	✓	✓	✓		S3	F1	C	C	C	
MICR Line			✓					✓		S2 or S3	F2				
Amount	VH		✓		✓		✓	✓		S5		C	C		
On-Us Field		✓	✓		✓		✓	✓	CP	S2 or S3		C	C		
Bank account number	VH	✓	✓		✓		✓	✓		S2 or S3		C	C		
Serial number (consumer)	VH	✓	✓		✓		✓			S2			C	C	
Processing/ trans code field	M	✓	✓		✓		✓			S2 or S3			C		
Routing and transit number (ABA)	VH	✓	✓		✓		✓	✓	M	S2 or S3		M	M	C	
External processing code (Pos. 44)	L	✓	✓		✓		✓		CP	S2 or S3		C	C		
Auxiliary On-Us Field			✓						CP						
Serial number (commercial)	VH	✓	✓		✓		✓	✓		S3		C	C	C	
Other data	H	✓	✓		✓		✓	✓		S3		C	C		
Signature	VH		✓	✓	✓		✓	✓		S3	F1		C	C	
Label--signature requirements (e.g. 2 sigs req. over $x,xxx)	L	✓	✓		✓		✓	✓		S2	F3		O		
Second signature	L		✓		✓		✓	✓		S3	F1		C	C	

Front of Check Attributes	Frequency	Check Design Attribute	Expect to Survive BW Imaging	Required to be Negotiable	Used by FI Pay Process	Needed to Prove Payment	Breach of RB/TB Warranty	May Affect Claim	X9.37 Usage	When Added to Check	Federal Reserve Category	Use Level I	Use Level II	Use Level III	Use Level IV
Bank (or bank branch) name	VH	✓	✓		✓		✓			S2	F2		C		
Bank address	M	✓	✓		✓		✓			S2	F2		N		
Label--memo line	H	✓					✓			S2	F3		N		
Memo line contents	L		✓				✓			S3	F3		O	O	
Automated security symbol (e.g., bar code, secure seal)	L		✓		✓		✓	✓		S2 or S3	F3		C	C	
Return item reason	VH		✓		✓		✓		M	S8	F3		C	C	
Registered check stock sequence number	L	✓	✓		✓		✓			S2	F3		O	O	
Miscellaneous added data (e.g., stamps, notes, etc.)	M		✓				✓	✓		S3 or S4					
Payer added	L		✓				✓	✓		S4	F3		O		
Payee/ depositor added	L		✓		✓		✓	✓		S4	F3		O		
Bank added										S5	F3		O		
Label(s)--Payment voucher data	L	✓					✓			S2 or S3	F3		N		
Payment voucher data	L		✓				✓	✓		S3	F3		O		
ACH routing code	M	✓	✓		✓		✓			S2	F3		N		
Bank product name	L	✓	✓				✓			S2	F3		N		
Lock icon & security feature notification text	VH	✓	✓							S2	F4		N		

Front of Check Attributes	Frequency	Check Design Attribute	Expect to Survive BW Imaging	Required to be Negotiable	Used by FI Pay Process	Needed to Prove Payment	Breach of RB/TB Warranty	May Affect Claim	X9.37 Usage	When Added to Check	Federal Reserve Category	Use Level I	Use Level II	Use Level III	Use Level IV
Bank logo	H	✓	✓				✓			S2	F5		N		
Check printer name	H	✓	✓				✓			S2	F3		N		
Label--MP (MicroPrint)	M	✓	✓				✓			S2	F2		N		
MicroPrinting	M	✓	✗							S2	F4		N		
Warning band	M	✓	✓				✓			S2	F3		N		
Decorative borders	M	✓	✗							S2	F5		N		
Pantographs	M	✓	✗							S2	F4		N		
Check style	L	✓	✓				✓			S2	F3		N	C	
Decorative or safety background	VH	✓	✗							S2	F5		N		
True or artificial watermarks	H	✓	✗							S1	F4		N		
Special inks (e.g., reactive)	M	✓	✗							S2	F4		N		
Chemical treatments	L	✓	✗							S1 or S2	F4		N		
Paper treatments	L	✓	✗							S1	F4		N		
Textural printing (e.g., embossing, intaglio)	L	✓	✗							S1	F4		N		
Hard to copy imagery (e.g., holograms, portraiture)	L	✓	✗							S1 or S2	F5		N	C	
Fibers and threads	L	✓	✗							S1	F4		N		

Back of Check Attributes	Frequency	Check Design Attribute	Expect to Survive BW Imaging	Required to be Negotiable	Used by FI Pay Process	Needed to Prove Payment	Breach of RB/TB Warranty	May Affect Claim	X9.37 Usage	When Added to Check	Federal Reserve Category	Use Level I	Use Level II	Use Level III	Use Level IV
Label--payee endorsement	VH	✓	✗				✓			S2	F3		O		
Endorsement area demarcation	VH	✓	✗							S2	F3		N		
Payee endorsement							✓	✓		S4	F3				
Payee signature(s) or stamp	VH		✓	✓			✓		CA	S4			C		
Payee account number	H		✓				✓		CA	S4			C		
Payee conditions/ deposit instructions	M		✓				✓	✓		S4			C		
Maker endorsement terms	L		✓			✓	✓	✓	CO	S2 or S4	F3		O		
Bank of first deposit endorsement	VH		✓				✓			S5	F3				
Bank name	VH		✓		✓	✓	✓			S5		C	O		
BoFD mark ▶ ◀ or ><	VH		✓		✓	✓	✓			S5		C	C		
RT number	VH		✓		✓	✓	✓	✓	CP	S5		C	C		
Processing (business) date	VH		✓		✓	✓	✓	✓	CP	S5		O	C		
Location (branch or processing center)	VH		✓				✓			S5		N	N		
Sequence number	VH		✓		✓	✓	✓		CP	S5		O	O		
Other data (e.g., phone #, account #, etc.)	H		✓				✓		CA	S5		N	N		
Second bank endorsement	M		✓		✓		✓			S6	F3		O		
Bank name	VH		✓				✓			S6			O		

Back of Check Attributes	Frequency	Check Design Attribute	Expect to Survive BW Imaging	Required to be Negotiable	Used by FI Pay Process	Needed to Prove Payment	Breach of RB/TB Warranty	May Affect Claim	X9.37 Usage	When Added to Check	Federal Reserve Category	Use Level I	Use Level II	Use Level III	Use Level IV
RT number	VH		✓				✓	✓	CP	S6			C		
Processing (business) date	VH		✓				✓	✓	CP	S6			C		
Location (branch or processing center)	VH		✓				✓			S6			N		
Sequence number	VH		✓				✓		CP	S6			O		
Other data (e.g., phone #, account #, etc.)	H		✓				✓			S6			O		
Third bank endorsement	L		✓		✓		✓			S6	F3		O		
Bank name	VH		✓				✓			S6			O		
RT number	VH		✓				✓	✓	CP	S6			C		
Processing (business) date	VH		✓				✓	✓	CP	S6			C		
Location (branch or processing center)	VH		✓				✓			S6			N		
Sequence number	VH		✓				✓		CP	S6			O		
Other data (e.g., phone #, account #, etc.)	H		✓				✓			S6			O		
Registered check stock sequence number	L	✓	✓				✓			S2	F3		N		
Miscellaneous added data (e.g., stamps, authentication, etc.)															
Maker added	L		✓				✓	✓		S3	F3		N		

Back of Check Attributes	Frequency	Check Design Attribute	Expect to Survive BW Imaging	Required to be Negotiable	Used by FI Pay Process	Needed to Prove Payment	Breach of RB/TB Warranty	May Affect Claim	X9.37 Usage	When Added to Check	Federal Reserve Category	Use Level I	Use Level II	Use Level III	Use Level IV
Payee added	L		✓				✓	✓		S4	F3		N		
Bank added	L		✓				✓			S5, 7, 8	F3		N		
Security text block	VH	✓	✗							S2	F4		N		
Original document screen	H	✓	✗							S2	F4		N		
Carbon strips	L	✓	✓							S1			N		

Notes on the Analysis

ANS X9.7, “Bank Check Background and Numerical Convenience Amount Field,” defines a Print Contrast Signal (PCS) requirement that should result in the $ sign, convenience amount, and MICR line remaining visible in the image. It defines the maximum background reflectivity and/or PCS for areas of interest described in the next paragraph. Many other attributes printed on the check are designed to be below the PCS and to therefore not survive in a black-and-white image.

ANS X9.7 also defines Areas of Interest (AOI) on the check, generally ¼ inch high. These are: MICR Line, Convenience Amount box, Date Line, Pay to the Order of Line, Dollar line, and Signature Line. The Convenience Amount is referred to throughout this document as the Courtesy Amount.

ANS X9/TG-2 (Technical Guideline 2), “Understanding and Designing Checks,” provides additional information regarding check design and design attributes. ANS X9/TG-8, “Check Security Guideline,” provides information on applying “traditional” check security features to combat check fraud.

A check may be issued to one or more beneficiaries. When multiple beneficiaries are named, they may be “joint” (AND) or “several” (OR). Although the beneficiary conditions are not separately analyzed, they may have a significant impact on the processing of a check and need to be preserved in the imaging process. These conditions, for selected accounts, may also be verified by the paying bank or the maker by verifying the information in the endorsement.

Checks may also be negotiated by beneficiaries to third parties. This document does not address the specific case of third (or more) party checks. Beneficiaries may also apply conditions to checks, such as “For Deposit Only.” These conditions are generally incorporated into the field “Payee conditions/ deposit instructions” as part of the Payee endorsement.

The MICR Amount field is currently listed as having a very high (> 90 percent) rate of occurrence. With the increased move to image-based systems, the amount field is being encoded in the MICR line less frequently.

This document addresses survivability in an image of a full-sized check. The Fractional TN, Account Verification Number (AVN), and some other attributes may appear in very small print. The legibility of these attributes improves with an increase in resolution. They may, however, be printed in too small a font to be legible in a substitute check or an image of a substitute check. Microprinting is too small to survive imaging at the resolutions commonly used for checks.

The AVN is simply the account number from the MICR line reprinted elsewhere on the check. It is not used by all banks. Some banks have also begun to print the ACH routing code on their checks to facilitate conversion of these items to ACH transactions.

Other payee descriptive data may include such information as address, account number, voucher number, invoice numbers, and so forth. Typically this type of information is found on business checks. There are no standards to ensure consistent placement or identification of this information. FSTC believes this makes it impractical for a truncating bank to usability test for this information on transit items.

MICR lines are remarkably complex and are typically not fully processed except by the paying bank. The Aux-On-Us field can only appear on a business-size check. It may contain other information in addition to the serial number. The last digit on a personal check is the External Processing Code (EPC) digit. The MICR line is specified in ANS X9.27, “Print Specifications for Magnetic Ink Character Recognition,” ANS X9.100-160-1, “Placement and Location of Magnetic Ink Printing (MICR)” (formerly X9.13) and ANS X9.100-160-2, “Placement and Location of Magnetic Ink Printing (MICR) Part 2: EPC Field Use” (formerly X9.13 Annex A only).

ANS X9.100-111, “Specification for Check Endorsements,” (formerly X9.53), and Regulation CC Appendix D, “Availability of Funds and Collection of Checks,” provide a detailed specification of the bank endorsements. For either bank of first deposit or subsequent bank stamps, the bank name, routing code, location, and date are required. The branch, phone number, and trace sequence number are specified as optional, but other data may also be included. In addition, for the bank of first deposit the >< or ►◄ characters are also required.

The Return Item Reason Code is shown in the table as mandatory for DSTU X9.37, “Specifications for Electronic Exchange of Check Image Data.” In that standard it is carried in record types 28, 31, and 35, and is mandatory for returned item records (Type 31) only. DSTU X9.37 is currently undergoing revision.

In DSTU X9.37 record type 26, the Payee account number is referred to as the Deposit Account Number at the Bank of First Deposit. This may not be the payee if the check is negotiated to a third party prior to deposit.

APPENDIX E:
IMAGE-DEFECT MEASUREMENTS REPORTED

Following is a list of the measurements reported by each testing vendor.

Figure E-1. Image-Defect Measurements Reported

Image Defect(s)	Measurements Reported	Image Defect(s)	Measurements Reported
Undersize Image/ Oversize Image	Image Height Front	Document Framing Error	Framing Front Bottom Height
	Image Width Front		Framing Front Left Width
	Image Height Rear		Framing Front Top Height
	Image Width Rear		Framing Front Right Width
Folded or Torn Document Corners	Torn Folded Corners Front Bottom Right Width		Framing Back Bottom Height
			Framing Back Left Width
			Framing Back Top Height
	Torn Folded Corners Front Bottom Right		Framing Back Right Width
	Torn Folded Corners Front Bottom Left		
	Torn Folded Corners Front Bottom Left		
	Torn Folded Corners Front Top Right Width	Excessive Document Skew	Skew Front Angle
	Torn Folded Corners Front Top Right Height		Skew Back Angle
	Torn Folded Corners Front Top Left Width	Piggyback Document	Piggyback Image Front
	Torn Folded Corners Front Top Left Height		Piggyback Image Back
	Torn Folded Corners Back Bottom Right	Image Too Light/Image Too Dark	Front Percent Black Pixels
	Torn Folded Corners Back Bottom Right		Front Percent Average Brightness
	Torn Folded Corners Back Bottom Left		Front Percent Average Contrast
	Torn Folded Corners Back Bottom Left		Rear Percent Black Pixels
	Torn Folded Corners Back Top Right Width		Rear Percent Average Brightness
	Torn Folded Corners Back Top Right Height		Rear Percent Average Contrast
	Torn Folded Corners Back Top Left Width	Horizontal Streaks Present in the Image	Front Number of Streaks
	Torn Folded Corners Back Top Left Height		Front Largest Streak Location
Folded or Torn Document Edges	Torn Folded Edges Front Bottom Width		Front Largest Streak Height
			Rear Number Streaks
			Rear Largest Streak Location
	Torn Folded Edges Front Bottom Height		Rear Largest Streak Height
	Torn Folded Edges Front Left Width		

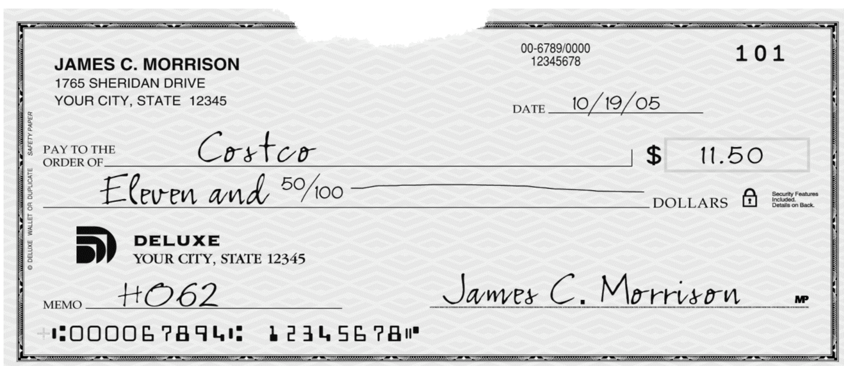
Image Defect(s)	Measurements Reported	Image Defect(s)	Measurements Reported
	Torn Folded Edges Front Left Height		
	Torn Folded Edges Front Top Width	Below Minimum/ Above Maximum Compressed Image Size	Front Size in Bytes
	Torn Folded Edges Front Top Height		Front Resolution
	Torn Folded Edges Front Right Width		Front Scheme
	Torn Folded Edges Front Right Height		Rear Size in Bytes
	Torn Folded Edges Rear Bottom Width		Rear Resolution
	Torn Folded Edges Rear Bottom Height		Rear Scheme
	Torn Folded Edges Rear Left Width	Excessive "Spot Noise" in the Image	Front Average Number of Spots
	Torn Folded Edges Rear Left Height		Rear Average Number of Spots
	Torn Folded Edges Rear Top Width	Front-Rear Image Dimension Mismatch	Width Difference
	Torn Folded Edges Rear Top Height		Height Difference
	Torn Folded Edges Rear Right Width	Carbon Strip Detected	Carbon Strip Detected
	Torn Folded Edges Rear Right Height	Image "Out of Focus"	Front Out of Focus
			Rear Out of Focus

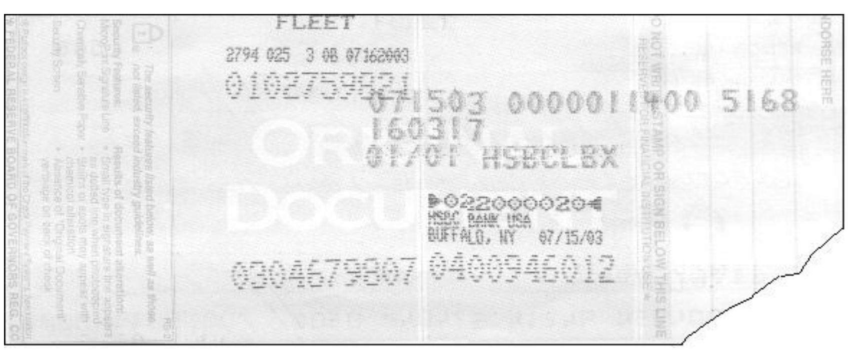
APPENDIX F:
EDGE-DEFECT TESTING SUPPLEMENT

Phase one of the FSTC IQ&U project defined sixteen check-image defects, including folds and tears for both check edges and corners. During phase two, the project team was tasked with defining the quantitative thresholds for these image-defect metrics. For folds and tears, this meant gathering empirical evidence about when a tear or fold is likely to affect check usability.

As shown in figure F-1, not every fold or tear affects meaningful data; therefore, to minimize false suspects it is critical to determine when an edge tear or fold adversely affects data. Lack of standards for check layout and differences in check usage make it difficult to predict the location of meaningful data relative to a check edge or corner. To date, no publicly available industry study provides reliable guidance regarding data location relative to check edges, thus it was necessary for the project team to perform its own statistical analysis of meaningful data location.

Figure F-1: Sample Tears That Don't Appear to Affect Data on One

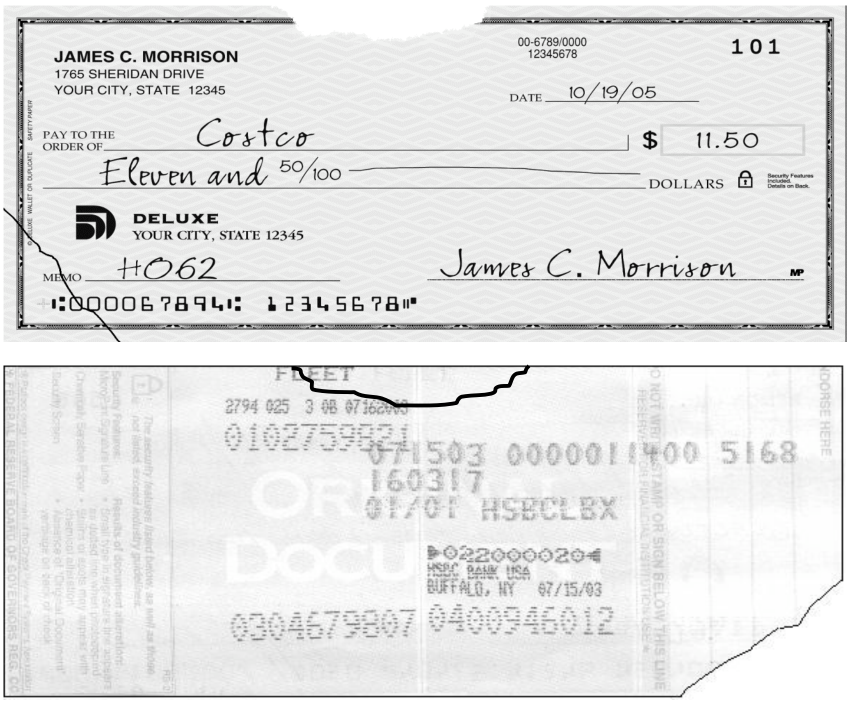




The project team measured a statistically significant sample of production checks to determine how close to a check edge meaningful data is located. Using this data, the project team created a model to predict the probability that a tear or fold will impinge on meaningful data. The team then validated this model during the live image-testing phase of the project.

The team found meaningful data close to check edges more frequently than anticipated because when both the front and back are considered, the amount of meaningful data potentially affected by an edge defect significantly increases. Figure 2 illustrates that although the tears do not affect data in figure F-1, when they are examined from the opposite side they do in fact affect meaningful data.

Figure F-2: Sample Tears Affect Data When Looking at Both Sides of the Check



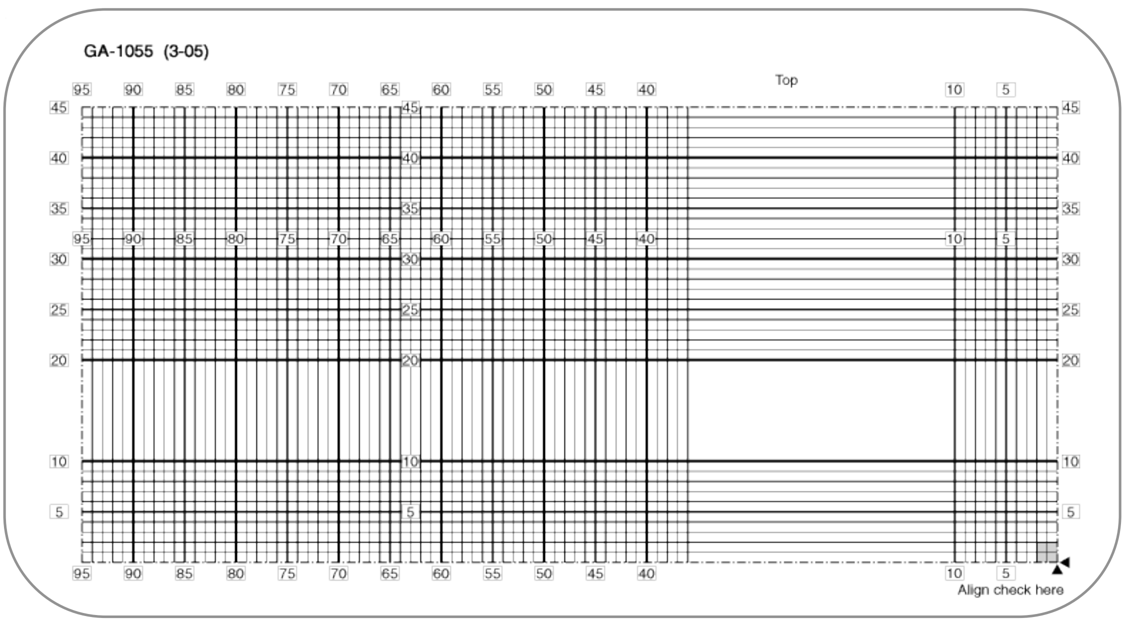
Methodology

The test required measurement of actual production checks. To accomplish this and ensure confidentiality, ten financial institutions measured their own checks. Bank operations staff made all measurements at their own sites using checks from the daily production work stream.

FSTC instructed test participants to collect a diverse sample of checks, focusing primarily on personal (wallet) and commercial (legal) checks. In addition, FSTC asked participants to include examples of money orders, travelers checks, international checks, and “others” (e.g., rebates, gift certificates, and refunds).

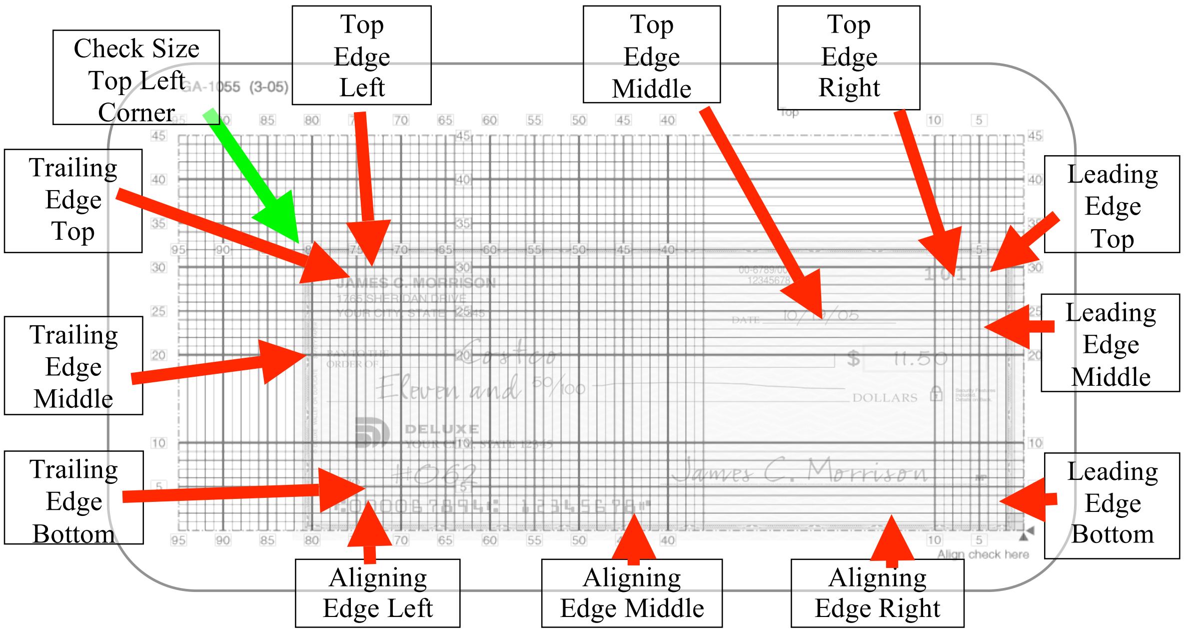
FSTC provided all test participants with an “edge measurement kit,” which included a transparent measurement tool, as shown in Figure F-3, developed for the project by Deluxe, a data collection form, and test instructions. An FSTC test coordinator conducted training via teleconference

Figure F-3: Edge Measurement Tool



As shown in Figure F-4, participants placed the measurement tool on top of each check to be measured. Using the grid on the tool, participants measured the distances to meaningful data closest to the edge from twelve check-edge locations, and they measured the check sizes. Participants also measured both the front and back of each check and recorded the data either manually or electronically onto the data collection sheet. The participants then submitted all data to FSTC, where it was combined, checked, and analyzed.

Figure F-4: Using the Edge Measurement Tool



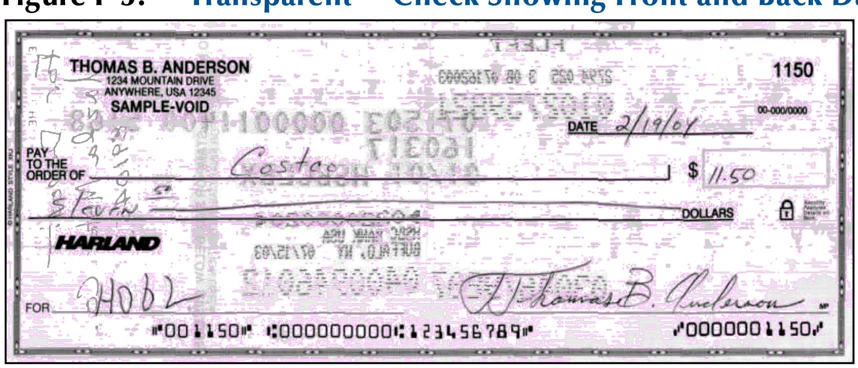
FSTC defined "meaningful data" as any information used during transaction processing. Participants did not count decorative borders and labels as meaningful data--with one exception. Based on advice from bank legal counsel, FSTC defined the label "Pay to the Order of" as meaningful data for purposes of the test.

Notes on Data Collection and Analysis

Front and Back Measurements

Edge defects affect data on both the front and back of a check. As Figures F-2 and F-5 show, either the front or back could have data closest to the edge, therefore, participants measured check fronts as well as backs. When compiling results for individual checks, participants compared front and back data at each point and used the measurement closest to the edge for making preliminary recommendations.

Figure F-5: "Transparent" Check Showing Front and Back Data



Edge Measurement Data Normalized

We measured the data as an absolute position on the grid. However, the value we were interested in was the distance from the edge to meaningful data. We normalized the data by converting all measurements to tenths of an inch from each edge, and we analyzed the data for all check types and sizes both independently and in aggregate. No differences significant enough to warrant separate thresholds for different check types emerged between check size/type and location of meaningful data relative to the check edge.

Measurement Granularity

Lines on the measurement grid were one-tenth of an inch apart. In cases where meaningful data fell between two lines, we recorded the line nearest the data. Therefore, a reference to a measurement of 0.2 inches from the edge actually includes items between 0.15 and 0.25 inches from the edge. We chose tenth-inch increments because we felt that an edge defect obscuring less than 0.05 inches of data was unlikely to affect the usability of that data adversely.

Suspect Data Discarded

The team scrutinized check data to ensure its validity, and we excluded suspect data from the analysis. For example, several checks had edge measurements that were greater than the recorded check size. As it is not possible for data to be located outside of the check, the team did not use these measurements. In all, we discarded measurements from 183 unreliable checks.

Results From the Lower Right Corner and Upper Right Corner

For the lower right front corner, FSTC instructed test participants to treat the MICR amount field as meaningful data. During preliminary test result analysis, the project team questioned whether this field should be considered meaningful data, because it is derived from other fields on a check. If the MICR amount is not considered meaningful, then the distances reported for Align Right and Lead Bottom are likely to be too small.

Similarly, for the upper right front corner, FSTC instructed test participants to consider the check number as meaningful data. Because this number is printed elsewhere on the check, it could be argued that a defect affecting this area is not critical. If the check number is not considered critical, then the distances reported for Top Right and Lead Top are also likely to be too small.

FSTC took these observations into account when making preliminarily recommendations for edge thresholds.

Participation

The following financial institutions participated in the edge testing:

- Bank of America
- Bank of New York
- Bank One
- Citigroup
- Fidelity Information Services
- JPMorgan Chase
- Southwest Corporate Federal Credit Union
- U.S. Bank
- Wachovia
- Wells Fargo

In total, the team analyzed 2,220 checks of the following types

Table F-1: Breakdown of Checks Analyzed

Check Type	Number of Checks
Wallet	333
Legal	1,395
Travelers	161
Money Order	93
International	163
Other	78
Total	2,220

Results

We did not anticipate the results of the edge measurement testing. We expected circumstances in which the distance between an edge and the beginning of meaningful data was greater than 0.3 inches. In fact, we found that for a significant number of checks, meaningful data occurred within 0.1 or 0.2 inches from every edge, meaning that any edge defect 0.2 inches or deeper is likely to affect meaningful data.

Table F-2 (shown on page 93) compares aggregate measurements for combined, fronts, and backs separately.

The table and figures below show the shortest distance after comparing front and back measurements for each point. A hypothetical check image is shown (Figure F-7) for reference purposes Results for each measurement location are expressed as a range of distances, with the percentage of "escapes" or checks with data closer to the edge associated with each distance, as explained in Figure F-6 and Figure F-7.

Figure F-6: Check-Measurement Reporting Key

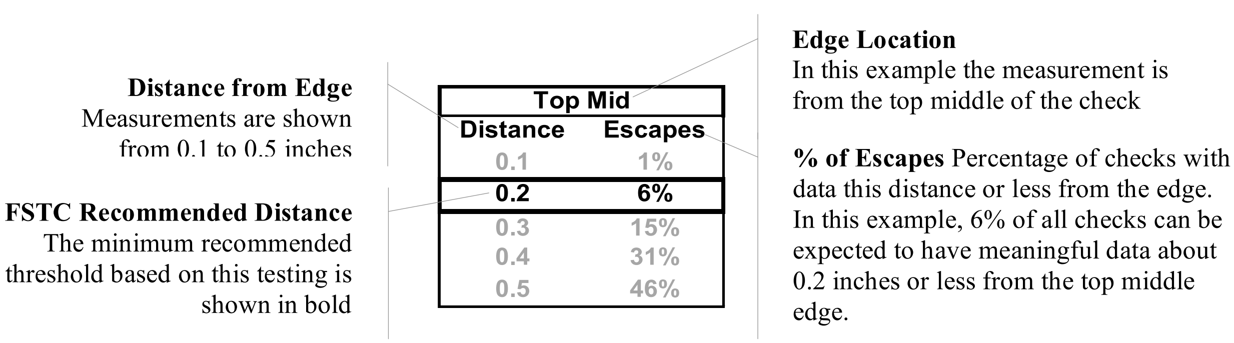


Figure F-7: Combined Distances from Edge

Top Left

Distance	Escapes
0.1	**1.9%**
0.2	9.3%
0.3	27.3%
0.4	55.9%
0.5	77.0%

Top Mid

Distance	Escapes
0.1	1.0%
0.2	**5.5%**
0.3	14.5%
0.4	31.5%
0.5	46.3%

Top Right

Distance	Escapes
0.1	1.1%
0.2	**8.1%**
0.3	27.7%
0.4	70.2%
0.5	87.0%

Trail Top

Distance	Escapes
0.1	**1.7%**
0.2	10.1%
0.3	29.1%
0.4	48.9%
0.5	64.1%

Lead Top

Distance	Escapes
0.1	2.3%
0.2	**9.7%**
0.3	24.1%
0.4	46.8%
0.5	68.8%

Trail Mid

Distance	Escapes
0.1	**2.1%**
0.2	16.2%
0.3	43.0%
0.4	72.7%
0.5	86.5%

Lead Mid

Distance	Escapes
0.1	**1.7%**
0.2	9.5%
0.3	24.9%
0.4	47.9%
0.5	62.7%

Trail Bottom

Distance	Escapes
0.1	1.1%
0.2	**8.8%**
0.3	26.5%
0.4	49.8%
0.5	62.1%

Lead Bottom

Distance	Escapes
0.1	4.4%
0.2	**16.9%**
0.3	35.5%
0.4	80.0%
0.5	97.8%

Align Left

Distance	Escapes
0.1	0.8%
0.2	**4.1%**
0.3	39.6%
0.4	58.1%
0.5	64.0%

Align Mid

Distance	Escapes
0.1	0.3%
0.2	**2.2%**
0.3	59.1%
0.4	96.9%
0.5	99.4%

Align Right

Distance	Escapes
0.1	1.1%
0.2	**4.5%**
0.3	72.1%
0.4	96.3%
0.5	98.0%

Distance is in tenths of inches from the edge. Escapes are the % of checks with meaningful data closer to the edge than the given distance.
FSTC recommendations show in bold for each location

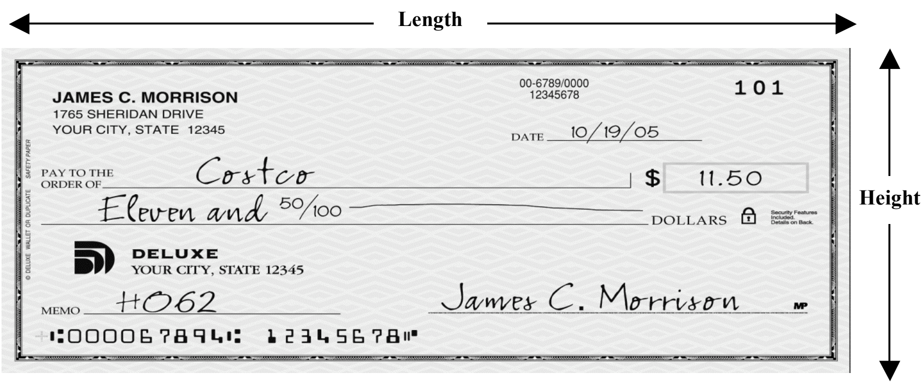
Table F-2: Edge Measurement Comparison--Combined, Front Only, Back Only

Check Edge Location	Distance (inches)	% of Checks Affected: Front/Back Combined	Front Only	Back Only
Top Left	**0.1**	**1.9%**	0.9%	1.0%
	0.2	9.3%	4.2%	5.4%
	0.3	27.3%	17.0%	13.7%
	0.4	55.9%	42.6%	26.6%
	0.5	77.0%	67.7%	35.6%
Top Middle	0.1	1.0%	0.5%	0.6%
	0.2	**5.5%**	2.9%	2.7%
	0.3	14.5%	9.3%	6.4%
	0.4	31.5%	23.8%	12.4%
	0.5	46.3%	37.3%	17.4%
Top Right	0.1	1.1%	0.7%	0.4%
	0.2	**8.1%**	6.1%	2.3%
	0.3	27.7%	24.0%	5.7%
	0.4	70.2%	67.1%	10.5%
	0.5	87.0%	85.4%	15.0%
Align Left	0.1	0.8%	0.0%	0.8%
	0.2	**4.1%**	0.8%	3.3%
	0.3	39.6%	33.2%	11.2%
	0.4	58.1%	51.4%	16.2%
	0.5	64.0%	54.4%	23.9%
Align Middle	0.1	0.3%	0.0%	0.3%
	0.2	**2.2%**	1.2%	1.0%
	0.3	59.1%	57.7%	3.2%
	0.4	96.9%	96.6%	6.4%
	0.5	99.4%	99.0%	11.2%
Align Right	0.1	1.1%	0.0%	1.1%
	0.2	**4.5%**	0.4%	4.3%
	0.3	72.1%	69.1%	10.7%
	0.4	96.3%	95.0%	19.1%
	0.5	98.0%	96.9%	24.6%
Trail Top	**0.1**	**1.7%**	0.5%	1.2%
	0.2	10.1%	2.8%	7.6%
	0.3	29.1%	7.5%	23.3%
	0.4	48.9%	21.4%	33.3%
	0.5	64.1%	33.0%	44.7%
Trail Middle	**0.1**	**2.1%**	0.8%	1.3%
	0.2	16.2%	7.5%	9.0%
	0.3	43.0%	24.8%	24.7%
	0.4	72.7%	58.6%	37.2%
	0.5	86.5%	75.7%	50.3%
Trail Bottom	0.1	1.1%	0.1%	0.9%
	0.2	**8.8%**	2.9%	5.9%
	0.3	26.5%	10.7%	17.3%
	0.4	49.8%	30.3%	29.2%
	0.5	62.1%	38.2%	41.6%
Lead Top	0.1	2.3%	0.1%	2.2%
	0.2	**9.7%**	1.5%	8.3%
	0.3	24.1%	12.1%	13.9%
	0.4	46.8%	30.4%	23.4%
	0.5	68.8%	56.1%	27.7%
Lead Middle	**0.1**	**1.7%**	0.2%	1.5%
	0.2	9.5%	1.5%	8.1%
	0.3	24.9%	9.5%	17.7%
	0.4	47.9%	25.5%	31.0%
	0.5	62.7%	42.6%	38.5%
Lead Bottom	0.1	4.4%	0.2%	4.2%
	0.2	**16.9%**	0.7%	16.5%
	0.3	35.5%	10.2%	29.4%
	0.4	80.0%	71.6%	39.0%
	0.5	97.8%	95.8%	44.1%

Check Size

Check length and height were measured for each check as shown in Figure F-8.

Figure F-8: Check-Size Measurements



The project team analyzed the results to determine the minimum and maximum check sizes expected. In addition, we derived the variance and standard deviation for measurements of each check type. Finally, measurement frequencies for each check type were plotted as histograms. As expected, the study found significant differences in size among different check types.

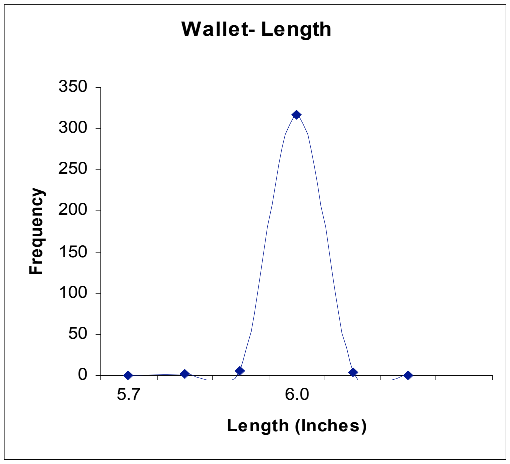
Table F-3 summarizes the results of a range of sizes for various check types. In addition, the following pages provide a breakdown of each check type and a histogram showing the size variance for each check type Tables F-4 through F-9. Finally, Table F-11 (shown on page 98) presents preliminary recommendations for check-size thresholds.

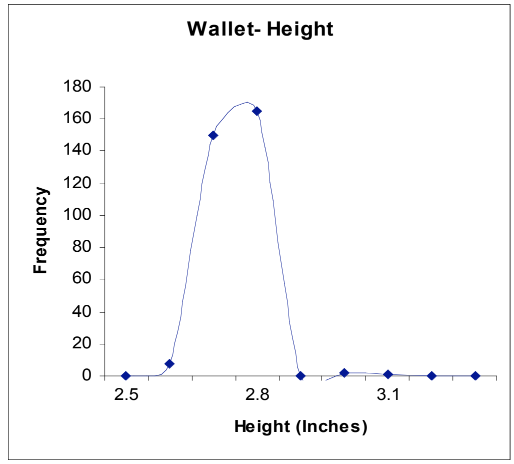
Table F-3: Check-Size Summary

Check Type	Sample Size	Length (Inches)				Height (Inches)			
		Min	Max	Var	Stnd Dev	Min	Max	Var	Stnd Dev
Wallet	330	5.8	6.1	0.04	0.2	2.6	3.1	0.39	0.6
Legal	1,395	6.8	8.8	7.89	2.8	2.6	4.3	5.08	2.3
Travelers	161	6.0	6.4	0.39	0.6	2.7	3.0	0.51	0.7
Money Order	93	6.0	8.6	111.16	10.5	2.7	3.5	8.75	3.0
International	163	7.4	8.5	8.48	2.9	2.8	4.2	6.20	2.5
Other	78	6.0	8.5	83.73	9.2	2.7	4.0	9.67	3.1
Combined	2,220	5.8	8.8	105.27	10.3	2.6	4.3	12.08	3.5

Check Type	Sample Size	Length (Inches)				Height (Inches)			
		Min	Max	Var	Stnd Dev	Min	Max	Var	Stnd Dev
Wallet	330	5.8	6.1	0.04	0.2	2.6	3.1	0.39	0.6

Table F-4: Wallet Check Size

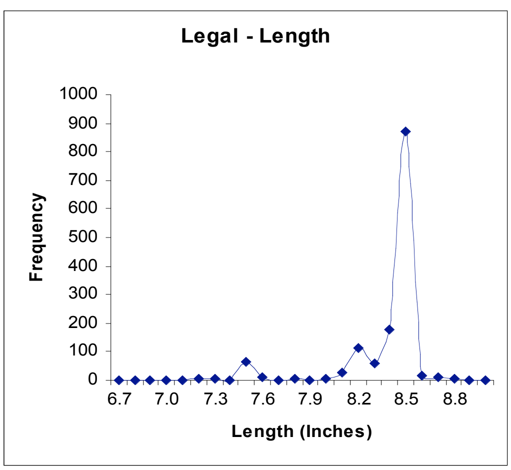


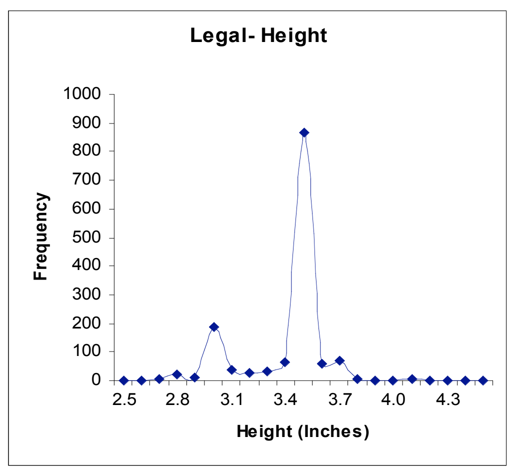


Wallet-style check varied little in size, with most checks 6.0 inches long and between 2.7 and 2.8 inches high. This is consistent with expectations and ANSI standard X9.7-1999 (Bank Check Background and Numerical Convenience Amount Field).

Check Type	Sample Size	Length (Inches)				Height (Inches)			
		Min	Max	Var	Stnd Dev	Min	Max	Var	Stnd Dev
Legal	1,395	6.8	8.8	7.89	2.8	2.6	4.3	5.08	2.3

Table F-5: Legal Check Size

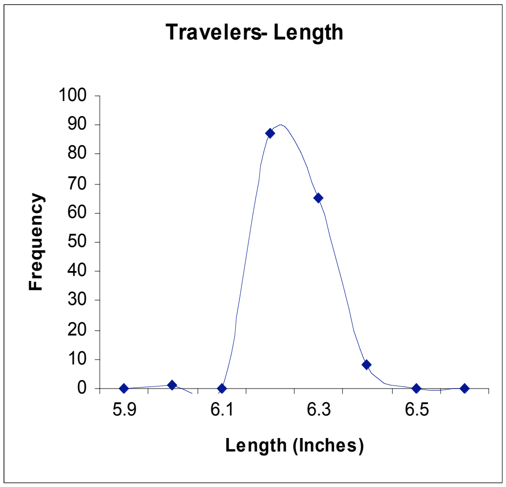


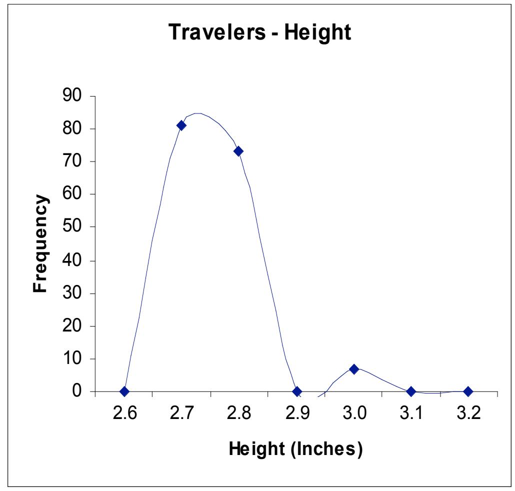


Legal and commercial checks showed more variability, although the majority of checks were 8.5 inches long and 3.6 inches high.

Check Type	Sample Size	Length (Inches)				Height (Inches)			
		Min	Max	Var	Stnd Dev	Min	Max	Var	Stnd Dev
Travelers	161	6.0	6.4	0.39	0.6	2.7	3.0	0.51	0.7

Table F-6: Travelers Check Size

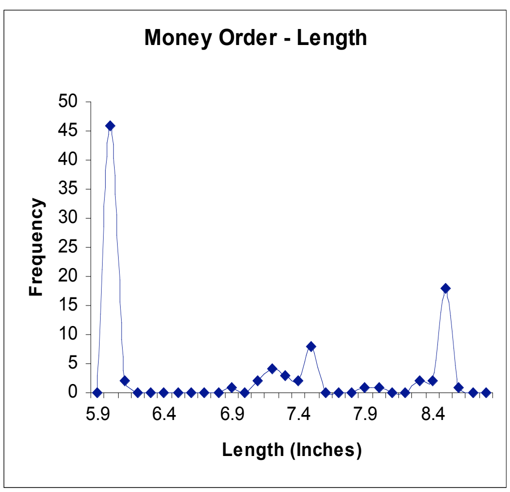


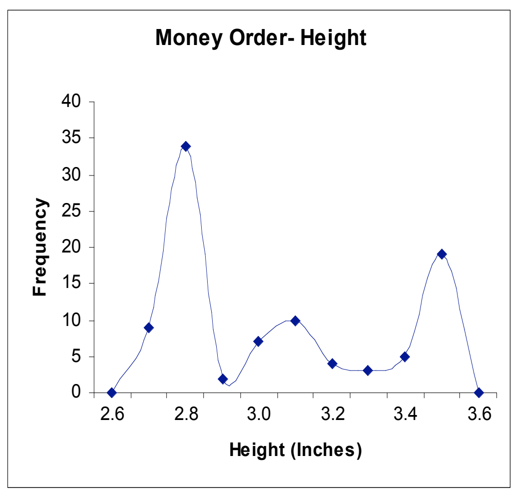


Travelers checks were quite standard, with check length at 6.2–6.4 inches and height between 2.7 and 2.8 inches.

Check Type	Sample Size	Length (Inches)				Height (Inches)			
		Min	Max	Var	Stnd Dev	Min	Max	Var	Stnd Dev
Money Order	93	6.0	8.6	111.16	10.5	2.7	3.5	8.75	3.0

Table F-7: Money Order Check Size

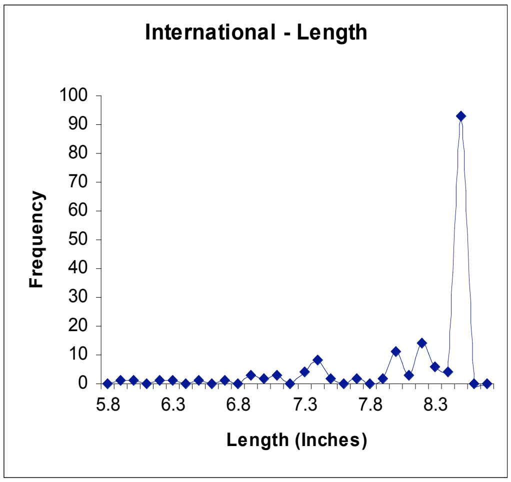


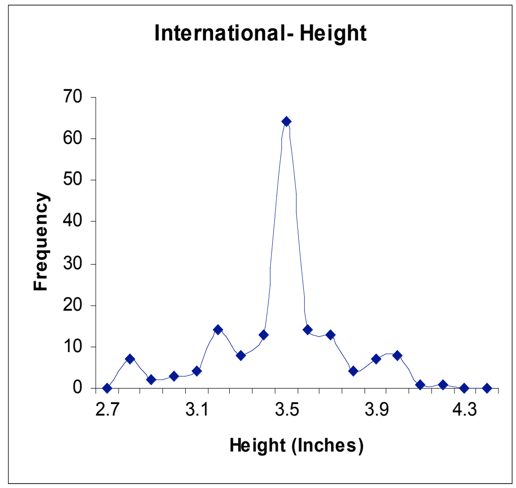


Money order check size was surprisingly variable. Anecdotal evidence from testers indicated little variability, but analysis of data from all tests shows that there are most likely several "standard" sizes for money orders: 6.0, 8.5, and possibly 7.4 inches for length; and 2.8, 3.5, and possibly 3.1 inches for height.

Check Type	Sample	Length (Inches)				Height (Inches)			
International	Size	Min	Max	Var	Stnd Dev	Min	Max	Var	Stnd Dev
	163	7.4	8.5	8.48	2.9	2.8	4.2	6.20	2.5

Table F-8: International Check Size

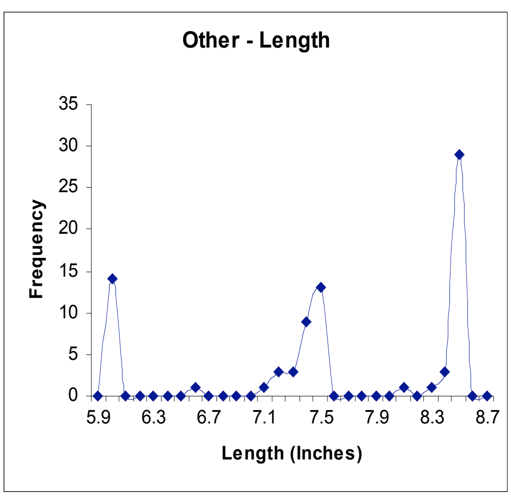


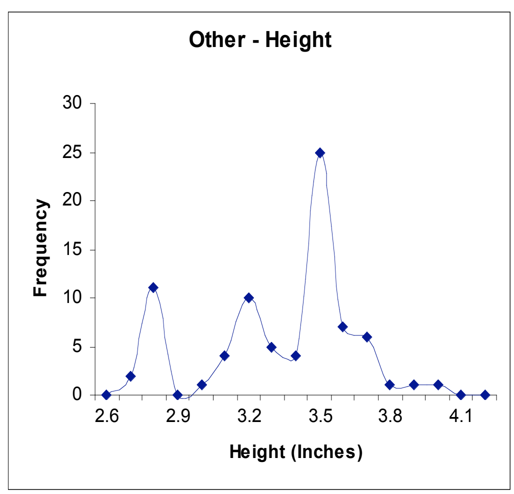


International check size varied, although the majority of international checks tended to be 8.5 inches long and between 3.2 and 3.7 inches high--however, the sample size was too small to draw reliable conclusions.

Check Type	Sample	Length (Inches)				Height (Inches)			
Other	Size	Min	Max	Var	Stnd Dev	Min	Max	Var	Stnd Dev
	78	6.0	8.5	83.73	9.2	2.7	4.0	9.67	3.1

Table F-9: Other Check Size





The category of "Other" encompassed all check types not covered by the other descriptions, and it included refunds, rebates, and gift certificates. As expected, the results for "Other" were variable, and given the small sample size, results were inconclusive. Anecdotal evidence from testers indicated that government refund checks can be quite small and may create processing challenges.

PRELIMINARY EDGE-THRESHOLD RECOMMENDATIONS

FSTC's check-edge measurement analysis led to the following preliminary recommendations (subject to confirmation during the large scale testing of live items).

Edge Defects

For a significant number of checks, meaningful data is located between 0.1 and 0.2 inches from every edge as shown in Table F-10, and preliminary threshold recommendations for each edge-defect measurement location are:

Table F-10: Edge Defects--Preliminary Threshold Recommendations

Distance from Edge of Data		
Check-Edge Location	Distance (inches)	% of Checks Impacted
Top Left	0.1	1.9%
Top Middle	0.2	5.5%
Top Right	0.2	8.1%
Align Left	0.2	4.1%
Align Middle	0.2	2.2%
Align Right	0.2	4.5%
Trail Top	0.1	1.7%
Trail Middle	0.1	2.1%
Trail Bottom	0.2	8.8%
Lead Top	0.2	9.7%
Lead Middle	0.1	1.7%
Lead Bottom	0.2	16.9%

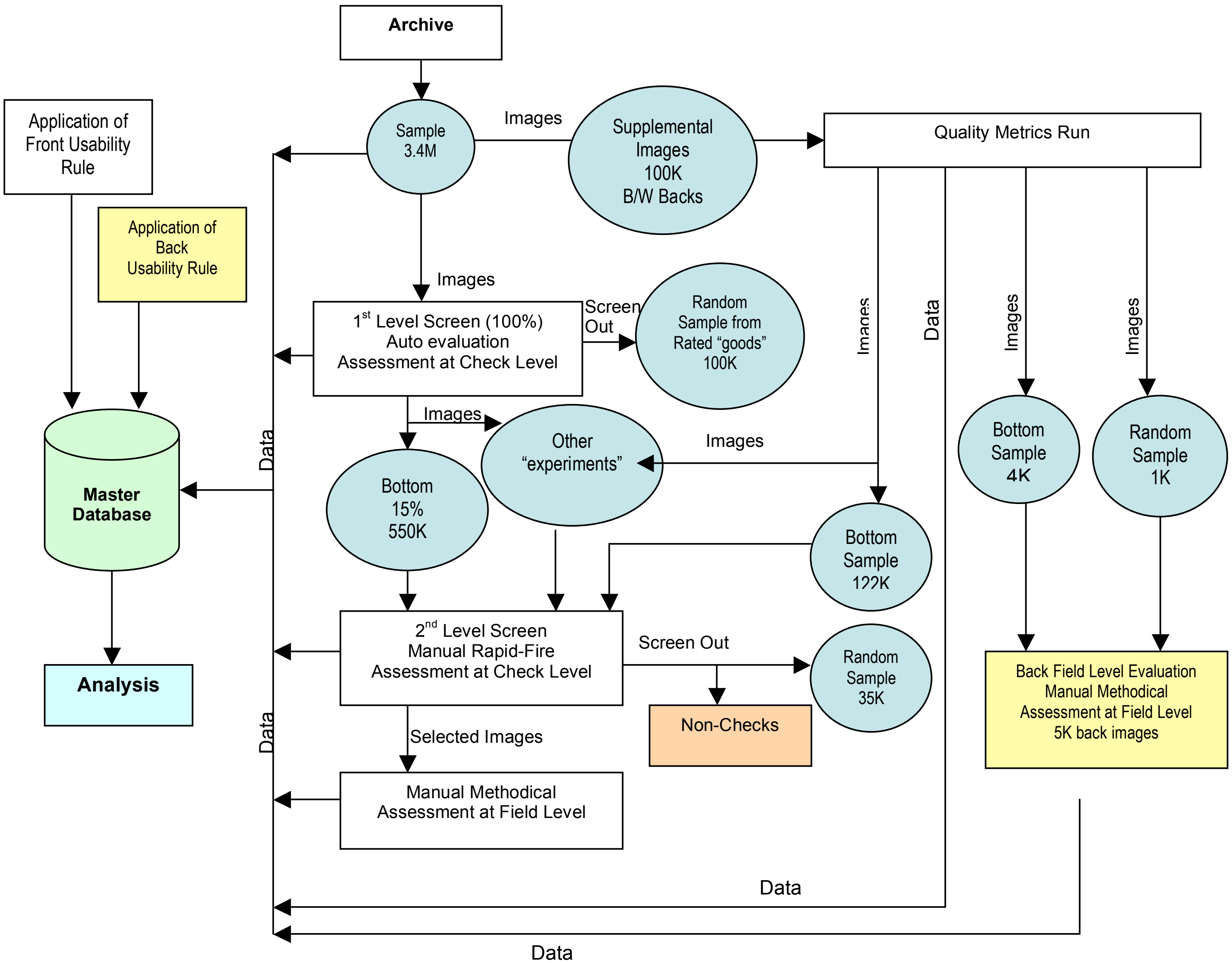
Check Size

The results of this portion of the study lead FSTC to suggest preliminary recommendations for check-size thresholds.

Table F-11: Check Size--Preliminary Threshold Recommendations

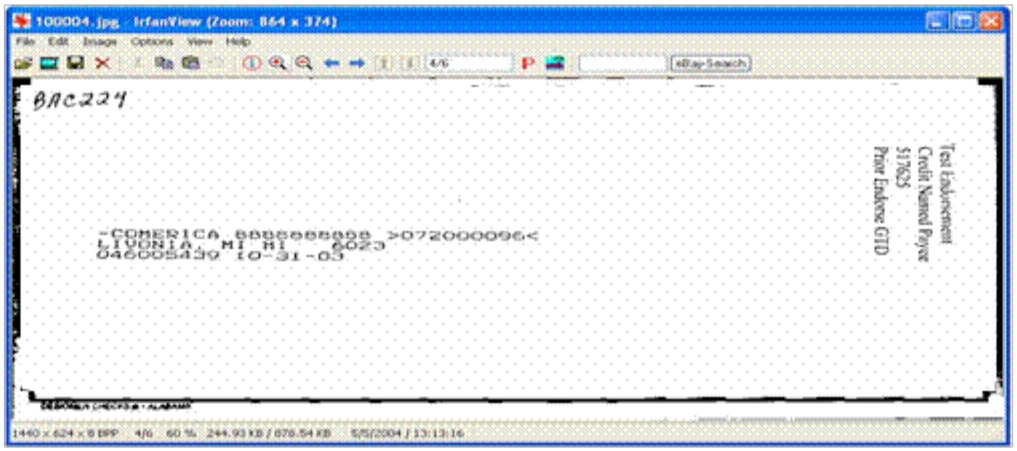
Check Size					
Length (inches)			Height (inches)		
Min	Max	% of checks	Min	Max	% of checks
5.9	8.6	98.1%	2.6	3.8	98.4%

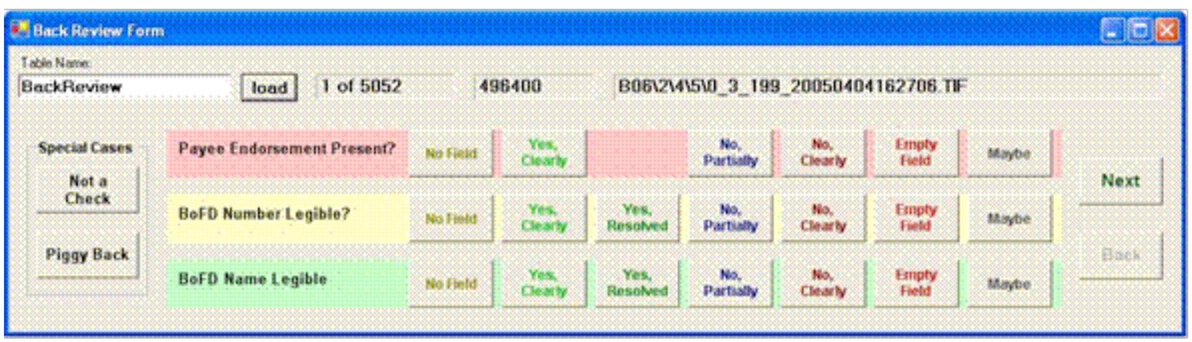
APPENDIX G: ADI WORKFLOW OVERVIEW



APPENDIX H:
METHODICAL MANUAL-REVIEW DATA COLLECTION SCREEN–FRONT

APPENDIX I:
METHODICAL MANUAL-REVIEW DATA COLLECTION SCREEN–REAR





APPENDIX J: SAMPLE IMAGES

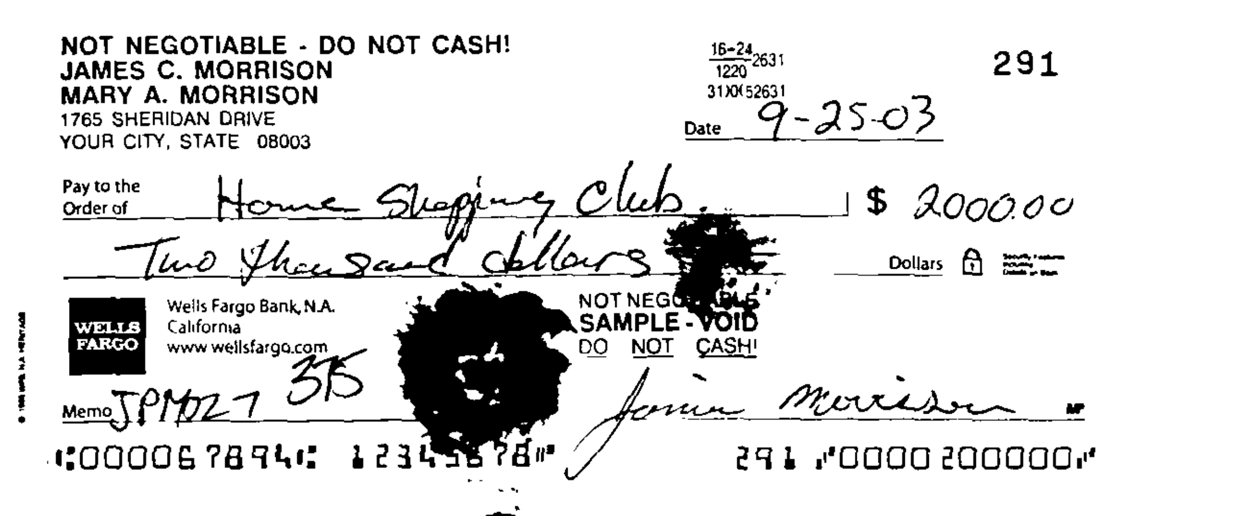
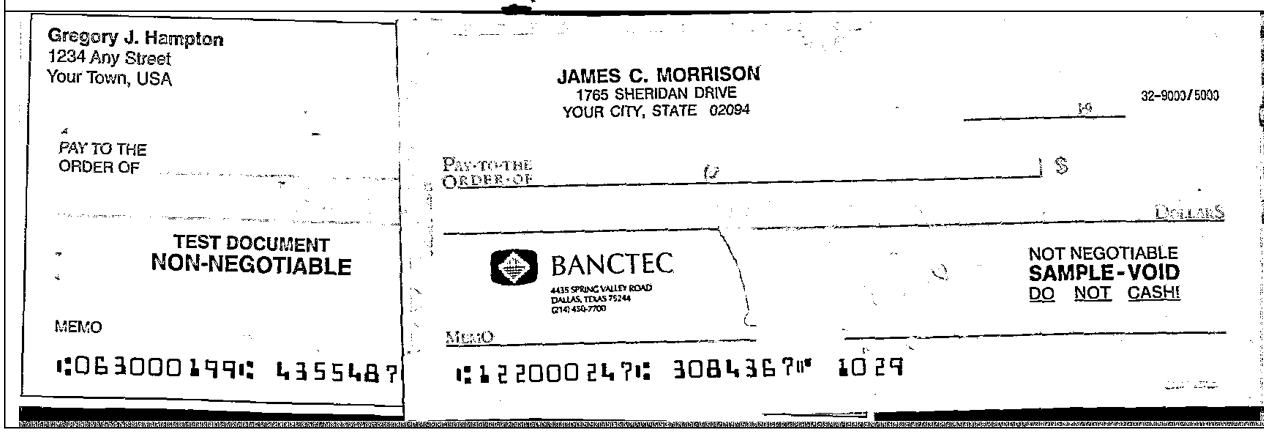
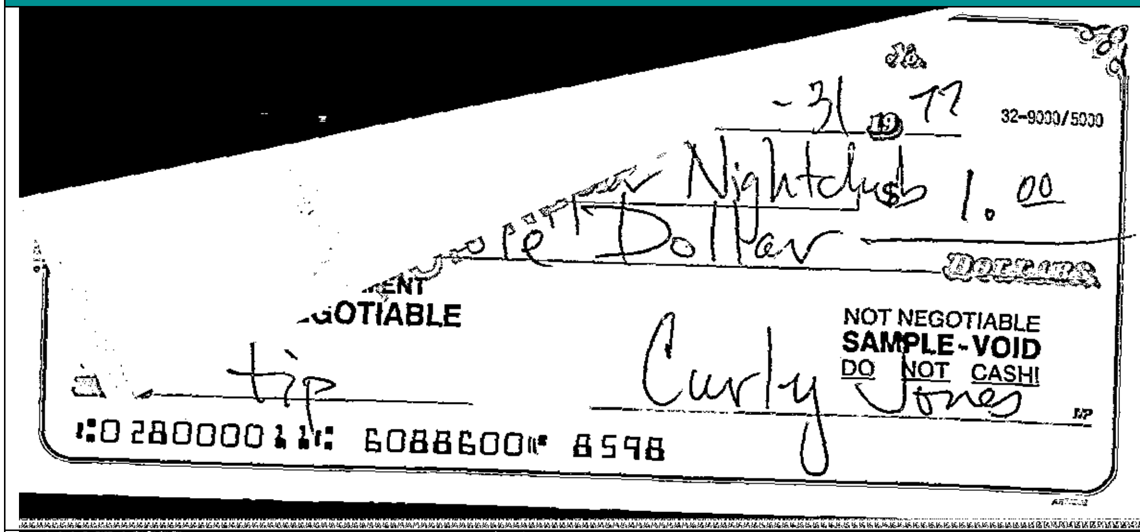
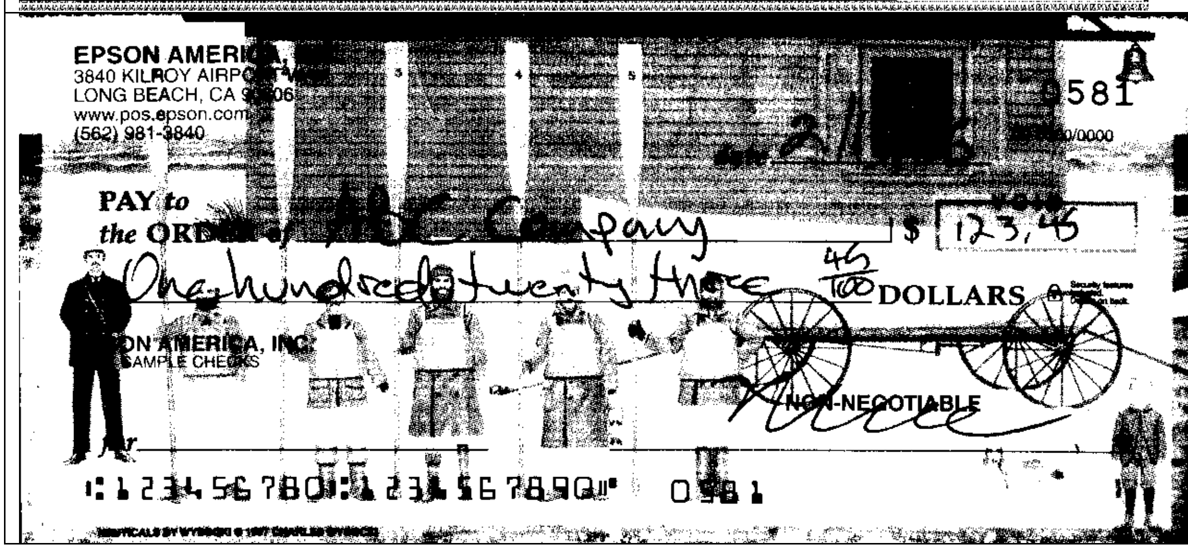
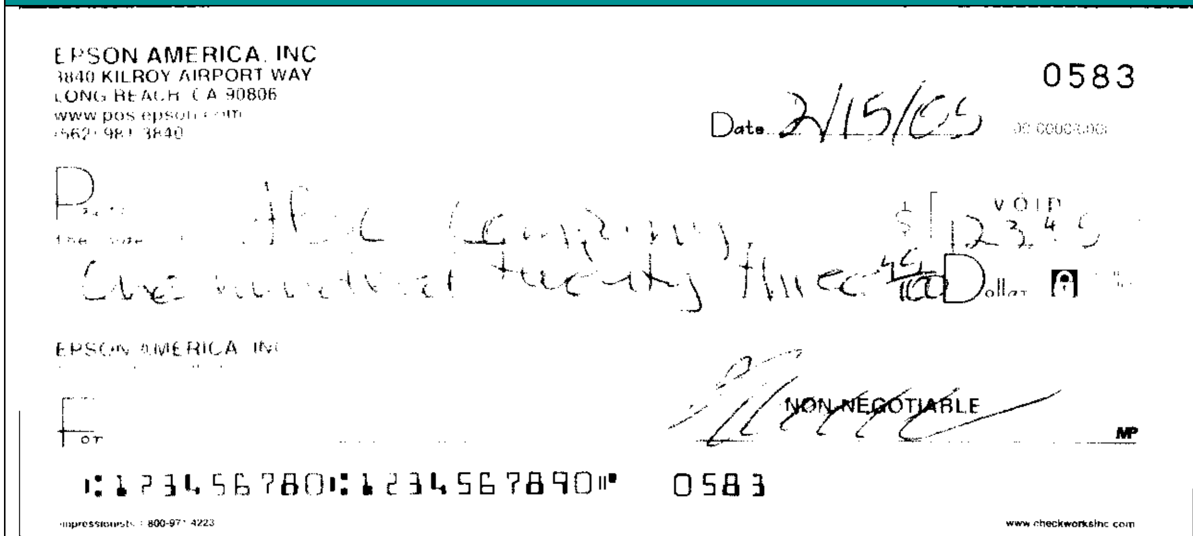
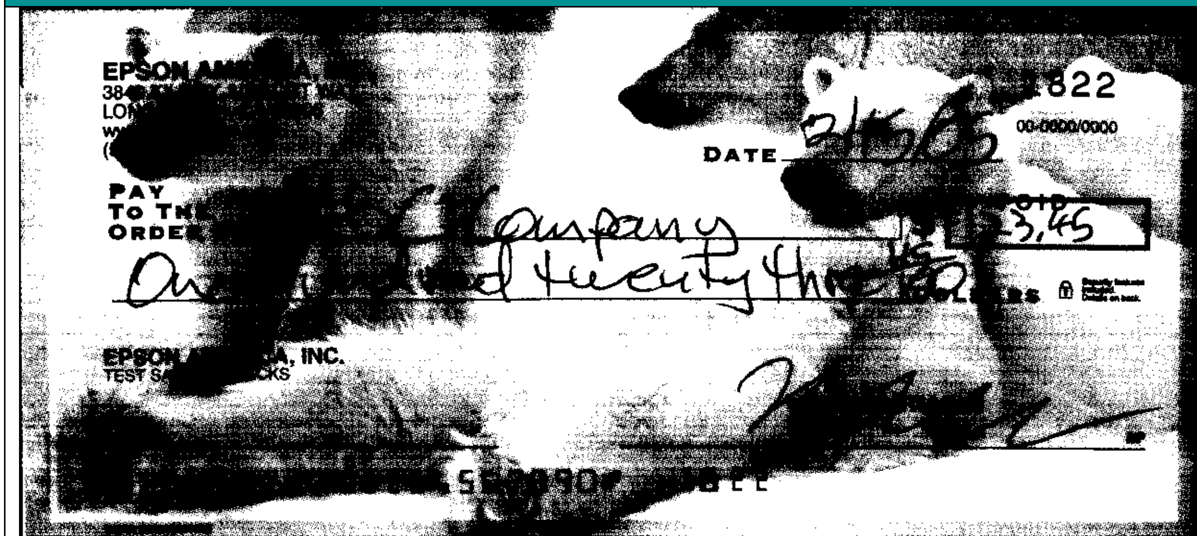
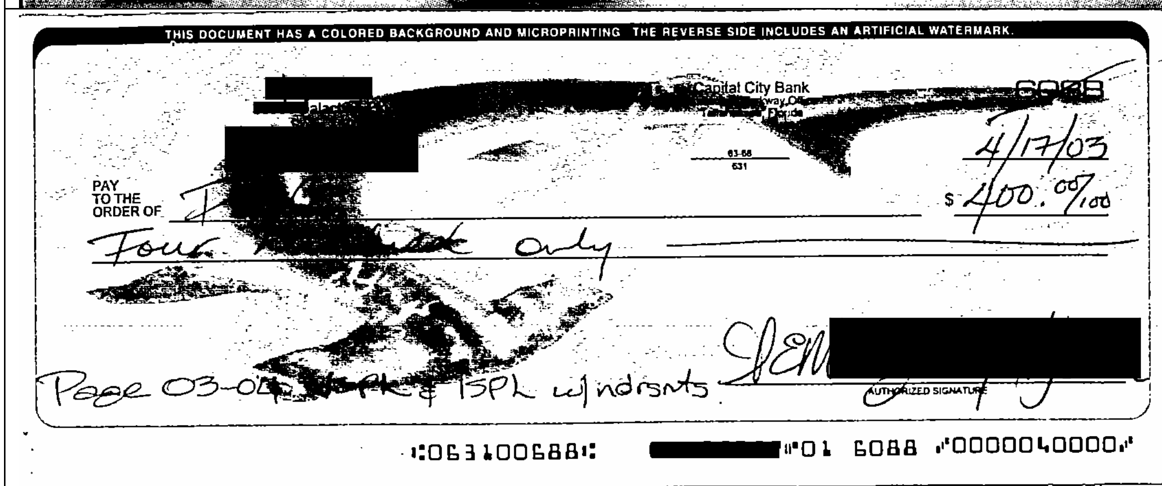
<table>
<tr><th>Image</th><th>Usability Score (Estimated)</th></tr>
<tr><td>
</td><td>2</td></tr>
<tr><td>
</td><td>Piggyback</td></tr>
</table>

Image	Usability Score (Estimated)
	9
	4

<table>
<tr><th>Image</th><th>Usability Score (Estimated)</th></tr>
<tr><td>
</td><td>8</td></tr>
<tr><td>
</td><td>15</td></tr>
</table>

<table>
<tr><th>Image</th><th>Usability Score (Estimated)</th></tr>
<tr><td>
</td><td>10 – Ambiguous, but by using both amount fields an experienced observer can figure out the amount.</td></tr>
<tr><td>
</td><td>7</td></tr>
</table>

FOR MORE INFORMATION

For more information about FSTC or the IQ&U initiatives, please contact:

Chris Nautiyal
Chris.Nautiyal@fstc.org
(201) 313 5756

Frank Jaffe
Frank.Jaffe@fstc.org or
Frank.Jaffe@MorSecure.com
(207) 879-9555

Zach Tumin
Zachary.Tumin@fstc.org
(914) 576-7629

