

INCITS M1/03-0563

INCITS 1602-D  
Biometric Performance Testing and Reporting  
Part 3: Scenario Testing and Reporting  
Draft 3.0

Task Group M1.4  
Biometric Performance Testing and Reporting

## Revision History

<b>Revision</b>	<b>Date</b>	<b>M1 Document</b>	<b>Comments</b>
1.0	4/14/2003	M1/03-0204	Initial draft
2.0	7/21/2003	M1/03-0351	Incorporates comments from week of June 9, 2003; initial distribution for internal review
3.0	11/5/2003	M1/03-0563	Incorporates IBG contribution M1/03-0477, replacing Section 6.N

INCITS M1/03-0563

## Foreword

tbd

Contents

**1. Scope..... 1**

**2. Conformance ..... 1**

**3. Normative References..... 1**

**4. Terms and Definitions..... 1**

4.1 Adaptation .....1

4.2 Attempt .....1

4.3 Enrollment .....1

4.4 Enrollment Attempt.....1

4.5 Enrollment Limit .....2

4.6 Enrollment Presentation .....2

4.7 Enrollment Sequence.....2

4.8 Matching .....2

4.9 Match Attempt .....2

4.10 Match Limit.....2

4.11 Match Presentation .....2

4.12 Match Sequence .....2

4.13 Extraction .....3

4.14 Match Template Data .....3

4.15 Presentation .....3

4.16 Reference Template.....3

4.17 Scenario Script .....3

4.18 System Description .....3

**5. Requirements for Scenario Testing..... 4**

**5.1 Introduction .....4**

**5.2 Test Design.....4**

5.2.1 Test Objectives .....4

5.2.2 Scenario Tested: Generic vs. Specific .....5

5.2.3 Identification, Verification, and Watchlist Testing .....5

5.2.4 COTS vs. Customized Systems Testing .....5

5.2.5 Controlling for Environmental Conditions .....6

5.2.6 Online and Offline Testing .....6

5.2.7 Habituation .....7

5.2.8 Acclimation.....7

5.2.9 Test Size.....8

5.2.10 Composition of Test Population .....9

5.2.11 Number of Visits / Time between Visits.....9

5.2.12 Number of Systems and Modalities Tested .....10

5.2.13 Use of Variable Matching Thresholds .....10

5.2.14 Physiological/Behavioral Data to be Enrolled .....11

**5.3 Test Execution .....12**

5.3.1 System Configuration.....12

5.3.2 System Maintenance .....13

5.3.3 Scheduling and Test Subject Management .....13

5.3.4 System Ordering.....13

5.3.5 Test Operator Roles and Responsibilities .....13

5.3.6 Directing Test Subjects in Enrollment and Matching.....14

5.3.7 Data Collection .....14

- 6. Scenario Test Reporting..... 15**
- 6.1 Systems and environment ..... 15**
- 6.1.1 Biometric system specifications ..... 15
- 6.1.2 Platform specifications ..... 15
- 6.1.3 System implementation..... 15
- 6.1.4 Physical layout of test environment ..... 15
- 6.1.5 Operational environment ..... 16
- 6.1.6 Match and enrollment outputs..... 16
- 6.1.7 Scenario Narrative ..... 17
- 6.2 Test processes..... 17**
- 6.2.1 Test subject management ..... 17
- 6.2.2 Performance data collection and management ..... 17
- 6.2.3 Use of Identifiers ..... 17
- 6.2.4 Enrollment processes..... 18
- 6.2.5 Matching processes ..... 19
- 6.3 System-Level results ..... 20**
- 6.3.1 System-Level enrollment results ..... 20
- 6.3.2 System-level 1:1 match results ..... 21
- 6.3.3 System-level 1:N match results..... 23
- 6.4 Subject-level results..... 24**
- 6.4.1 Subject-Level enrollment ..... 24
- 6.4.2 Subject-Level 1:1 match results ..... 24
- 6.4.3 Subject-level 1:N match results..... 25

INCITS M1/03-0563

## Introduction

tbd

## 1. Scope

The objective of this standard is to establish requirements for scenario-based biometric testing and reporting.

The goal of scenario testing is to determine the overall system performance in a prototype or simulated application. Testing is carried out on a complete system in an environment that models a real-world target application of interest. Each tested system will have its own acquisition sensor and so will receive slightly different data. Consequently, care will be required that data collection across all tested systems is in the same environment with the same population. Depending upon the data storage capabilities of each device, testing might be a combination of offline and online comparisons. Test results will be repeatable only to the extent that the modeled scenario can be carefully controlled. (BPTRP Par.20).

## 2. Conformance

No known conformance in this area.

## 3. Normative References

- Best Practices in Testing and Reporting Performance of Biometric Devices Version 2.01 by A. J. Mansfield, National Physical Laboratory and J. L. Wayman, San Jose State University

**[Ed. Note: I no longer think this belongs as a normative reference, but I cannot make that decision unilaterally. Input welcome.]**

## 4. Terms and Definitions

Note: terms and definitions may be modified to ensure alignment with standard terminology lists being developed within SC37 and M1.

General testing definitions may only reside in Part 1, with definitions specific to scenario testing residing below. Terms in bold are those which, in our initial estimation, are specific to scenario testing and reporting. Terms from Best Practices are labeled BP, terms from X9.84 are labeled X9.

### 4.1 Adaptation

The process by which a biometric system automatically updates or refreshes a reference template, normally to account for temporal changes in the biometric samples.

### 4.2 Attempt

The submission of a biometric sample on the part of an individual for the purposes of enrollment, verification, or identification in a biometric system. An individual may be permitted several attempts to enroll, to verify, or to be identified. An attempt may be comprised of one or more presentations. (X9)

### 4.3 Enrollment

The process of collecting biometric samples from a test subject and subsequently generating and storing one or more reference templates associated with that subject. X9

### 4.4 Enrollment Attempt

The submission of one or more biometric samples on the part of an individual for the purposes of enrollment in a biometric system. An individual may be required or permitted several attempts to enroll. An enrollment attempt may be comprised of one or more enrollment presentations.

## **4.5 Enrollment Limit**

The maximum number of failed enrollment sequences permitted in a scenario test effort, subsequent to which a test subject is declared a failure to enroll. X9

## **4.6 Enrollment Presentation**

The provision of a single biometric sample on the part of a test subject for the purpose of enrollment in a biometric system. Multiple enrollment presentations may be required or permitted to enroll in a biometric system, or a single presentation may be sufficient, depending on the specific system.

## **4.7 Enrollment Sequence**

The number of enrollment attempts a test subject is permitted before an attempt to enroll is declared unsuccessful. Depending on test protocols, one or more enrollment sequences may be permitted before a subject is declared to have failed to enroll for a given system.

## **4.8 Matching**

The process by which a biometric system compares a match template against one or more previously stored reference templates. X9

## **4.9 Match Attempt**

The submission of one or more biometric samples on the part of an individual for the purposes of matching in a biometric system. An individual may be required or permitted several attempts to match. A match attempt may be comprised of one or more match presentations.

## **4.10 Match Limit**

The maximum number of failed match sequences permitted in a scenario test effort, subsequent to which a test subject is declared a failure to match. This terminology is applicable to both genuine and imposter attempts. X9

## **4.11 Match Presentation**

The provision of a single biometric sample on the part of a test subject for the purpose of matching in a biometric system. Multiple match presentations may be required or permitted to match in a biometric system, or a single presentation may be sufficient, depending on the specific system.

## **4.12 Match Sequence**

The number of match attempts a test subject is permitted before an attempt to match is declared unsuccessful. Depending on test protocols, one or more match sequences may be permitted before a subject is declared to have failed to match for a given system.

## **4.13 Extraction**

The process of converting raw biometric data into processed biometric for use in template comparison or reference template creation. X9

## **4.14 Match Template Data**

Match Template Data which represents the biometric measurement of a claimant, extracted from a claimant's biometric sample and used by a biometric system for comparison against one or more stored reference templates. X9

## **4.15 Presentation**

The provision of a single biometric sample on the part of a test subject for the purpose of enrollment or matching in a biometric system.

## **4.16 Reference Template**

Data, which represents the biometric measurement of an enrollee, extracted from an enrollee's biometric sample and typically stored and used by a biometric system for comparison against subsequently submitted match templates. X9

## **4.17 Scenario Script**

A script utilized by a test operator in the direction of a test subject during enrollment and verification sequences.

## **4.18 System Description**

A description utilized by a test operator to introduce a test subject to general system usage, including placement and interaction methods.

## 5. Requirements for Scenario Testing

*Note: all references to verification below are applicable to both verification and identification applications unless otherwise noted.*

Scenario testing differs from technology testing primarily as follows:

- Scenario testing evaluates biometric sensors (cameras, scanners, microphones, etc.) in conjunction with their associated matching algorithms as a full biometric systems; technology testing acquires datasets through one or more sensors and processes these identical datasets through a plurality of algorithms.
- Scenario testing defines error rates for enrollment, imposter test subject verification, and genuine test subject verification in accordance with the generic requirements of a real-world target application. Accordingly, multiple attempts may be allowed to enroll, to be verified, or to be identified.
- Scenario testing is predicated on live interaction between a test subject and one or more biometric systems, and as such is therefore more likely to incorporate online testing than offline testing (in contrast to technology testing). As a result test populations are likely to be smaller in scenario testing than in operational testing.

Scenario testing differs from operational testing primarily as follows:

- Scenario testing utilizes systems installed and configured specifically for a test effort as opposed to “live” operational systems.
- Scenario testing introduces imposters specifically for the purpose of attempting verification and/or identification as legitimate users. As such, scenario testing can often generate more extensive FMR data than operational testing.
- These factors impact the manner in which scenario tests are conducted and results are reported.

### 5.1 Introduction

This section discusses Test Design, or areas to be addressed and/or considered by a Test Organization prior to executing scenario testing, and Test Execution, or the logistics and mechanics of conducting the test. Section 6 addresses Test Metrics and Reporting Requirements.

### 5.2 Test Design

Prior to executing a scenario-based biometric test, it is necessary to account for the following design-related elements that can impact test results.

#### 5.2.1 Test Objectives

Beyond the generation of performance metrics such as FMR, FNMR, and FTE for biometric systems in an emulated usage scenario, biometric scenario testing can be designed to accomplish a number of objectives. Understanding the objectives of a scenario test is central to decision-making at various points in test design. Such objectives may include measuring performance of a single biometric system across multiple populations, measuring performance of multiple biometric systems or technologies across one or more target populations, determining error rates over time, determining the relation between enrollment rates and matching rates, and many other potential objectives.

Because test objectives will drive decision-making at various points of test design, such objectives should be discussed in Test Reports.

## **5.2.2 Scenario Tested: Generic vs. Specific**

A test organization must clearly define the biometric usage scenario being emulated and tested in order to effectively render the scenario's concept of operations within the test process. Biometric usage scenarios may be generic or specific. Generic scenarios, such as 1:1 authentication, have the advantage of generating results of interest to a broader range of deployers. Specific scenarios, such as 1:1 authentication for access control utilizing token-based biometric storage, generate results more directly applicable to a narrower range of deployer requirements.

To avoid misinterpretation of data, Test Reports should contain discussion of the assumptions and preconditions underlying the scenario under evaluation and the potential applicability of test results to both general and specific scenarios.

## **5.2.3 Identification, Verification, and Watchlist Testing**

Depending on the scenario and modalities being evaluated, a scenario test may include biometric systems that operate in 1:1, 1:N open set, and 1:N closed set modes. Scenario testing that involves both 1:1 and 1:N systems requires a careful approach to test process development and data collection in order to ensure equitable results reporting. Comparing results from verification and identification systems can be challenging, as the task of biometric identification is more difficult than that of verification. A system may be susceptible to higher FMR and FNMR when performing identification as opposed to verification. However, certain scenarios can utilize both verification and identification modes (e.g. physical access), such that it is not unreasonable to construct a test approach that compares such operations. This issue is complicated to a degree by systems that perform a 1:N search upon enrollment in order to ensure that duplicate enrollments do not occur. Such systems may require a reordering of FNMR, FTE, and FMR testing.

Test Reports associated with scenario evaluations that evaluate both 1:1 and 1:N systems should address the manner in which test design and results reporting account for the use of such systems.

## **5.2.4 COTS vs. Customized Systems Testing**

Scenario testing may incorporate commercial off-the-shelf (COTS) systems, customized systems, or a mixture of both. There are benefits to enforcing COTS-only and customized-only system requirements. By testing COTS systems, a test organization has an increased certainty that device performance is reflective of a sensor/algorithm combination as available on the market. By allowing for customization, a test organization has an increased certainty that a sensor/algorithm combination can be modified to meet the requirements of a given test scenario. Customization of a biometric system could entail modifying enrollment thresholds to accommodate specific test populations. Customization of systems for a test effort is not generally considered an ideal process, as the results may be more reflective of a vendor's ability to customize a system to address a specific scenario than of the biometric system's core abilities to enroll and verify users. However it should be noted that there may be substantial interest in how well a vendor can customize a system to meet a certain test protocol.

If a distinction is made between testing COTS and customized systems, this should be detailed in the Test Report. In addition, vendors should provide information regarding the degree to which a customized system has been modified to meet the requirements of a given scenario test.

## 5.2.5 Controlling for Environmental Conditions

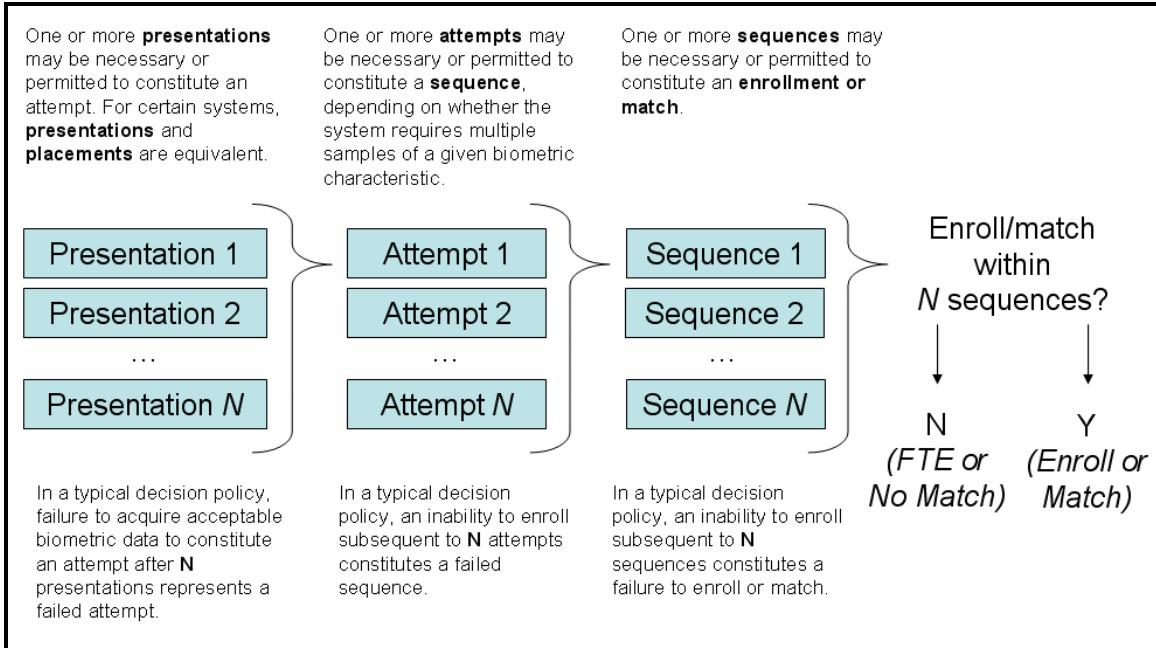
Depending on the scenario being tested, environmental conditions may be manipulated or introduced in order to emulate a specific usage environment. Scenario testing provides opportunities for introducing such environmental conditions and constraints, given the use of a specific test population providing data specifically for the purpose of biometric testing. Such controls may be established for temperature, lighting, humidity, and other factors known or suspected to impact biometric performance.

Such controls established for environmental conditions in a scenario test should be provided as part of Test Reports. This provides readers interested in deploying biometric technology with an understanding of the degree to which a test environment mirrors their specific usage environment. Should a test organization decide that establishing environmental controls is unnecessary for a given scenario test, it is still necessary to report the basic conditions (indoor, outdoor, air conditioned, office, industrial) under which a test took place.

## 5.2.6 Online and Offline Testing

Scenario testing is generally conceived of as being executed in an online fashion, such that test subjects interact with systems in real time such that the results of enrollment and matching attempts drive the subject's subsequent interaction with a test system. This approach is in contrast to collection of data for the purposes of offline testing. Scenario testing captures the transactional element of enrollment and matching, consistent with the concept that a subject may be allowed or required to provide multiple presentations to enroll or match, and may be allowed or required to execute multiple sequences (comprised of multiple presentations) to enroll or match.

If testing is conducted in an offline fashion, such that biometric data is compared and results generated at a time or location that does not impact test execution, then it is normally advisable to collect sufficient data during enrollment and matching such that one can derive scenario-specific data. For example, collecting only a single placement for enrollment and a single placement for matching on each system or device would only enable reporting of enrollment and matching for a single placement. Most systems either allow or require multiple placements to enroll and match; collection of N samples or templates to be used in enrollment and collection of N samples or templates to be used in matching allows generation of offline data in a fashion compliant with the transactional nature of scenario testing. This offline approach does reduce the impact of positive and negative feedback on enrollment and matching, and may thereby skew results.



The manner in which biometric extraction and matching processes are executed should be discussed in the Test Report to ensure that the reader understands whether performance was measured in an online or offline mode.

### 5.2.7 Habituation

The degree of habituation introduced both before testing and during test execution can impact test results. In a habituated biometric system users interact with the device on a regular basis and therefore presumably gain operational insights into the most effective mode of device interaction. Systems such as time and attendance and network login would in most cases be presumed to be habituated applications. Systems such as voter ID, border entry, and entitlements may require much less regular interaction with a biometric system or device and therefore would be more likely to be categorized as non-habituated. Determining whether the test is meant to emulate a habituated or non-habituated usage environment is an essential aspect of defining the parameters of a scenario test. Emulating habituated usage requires management of test subjects to facilitate repeated usage over a period of time, and implies that a subject will interact with a system for a longer period of time than just the initial enrollment and authentication and that a subject will return for  $N$  number of visits.

Because no firm parameters exist to define levels of habituation, test operators should ensure that all relevant data is collected for the purposes of reporting the number of visits, the duration between visits, the period of time over which visits occurred, and the trends in performance (such as reduced FNMR) present over the course of habituated usage.

### 5.2.8 Acclimation

Because scenario test results can be heavily impacted by the level of effort required in interacting with test systems, the question of acclimation is central to scenario testing. At a minimum, survey data must be collected from test subjects regarding their degree of acclimation with the biometric systems being tested (the degree of acclimation may vary if multiple biometrics are being tested). This information will in most cases be of interest to test readers. For the purposes of results reporting, a statement of disclosure or intent should accompany scenario test results that describes the degree of acclimation required of the users, how such acclimation was measured (i.e. self-disclosed vs. objectively assessed), and what steps were taken to train users not

acclimated to system usage. Data collection, disclosure, and reporting in this area is complicated by the following factors, each of which can impact results:

- *Acclimation with one biometric vs. with multiple biometrics.* If multiple biometric technologies are tested in a given scenario test, then a subject may be acclimated with one and not others, impacting results (assuming results are meant to be comparative).
- *Acclimation with one device vs. multiple devices.* If multiple devices within a given biometric modality are tested, then a subject may be acclimated to one device type but not with others. For example, with fingerprint technology, a user may be acclimated to placement but not swiping a fingerprint. Therefore performance may differ for placement vs. swipe-based devices. As acclimation with biometric systems across the general population increases, it may be necessary to develop more granular categories of acclimation than modality-specific.
- *Relevance of acclimation to a given modality.* Certain biometric technologies are more heavily impacted by the degree of acclimation than others. For example, facial recognition may be implemented in a fashion that requires little or no effort on the part of the end user. Fingerprint, on the other hand, generally requires deliberate, direct action on the part of the user. Results may be skewed to favor modalities in which acclimation is more relevant if acclimation is a test precondition.

A discussion of acclimation that addresses these issues, as appropriate, is a necessary component of a Test Report. Sufficient data should be present for a test reader to assess the degree to which acclimation was a factor for different systems and technologies.

### 5.2.9 Test Size

Factors that can particularly impact test size within scenario testing include the following:

- *Generating a sufficient number of errors for a test to maintain statistical significance.* The question of test size relative to the number of errors generated for a given performance metric is addressed for the most part in BP, with its references to rules of 3 and 30. Although one cannot necessarily know that a given test population will generate X errors, it is necessary to anticipate testing a sufficient number of subjects such that statistically relevant test results will be generated. *This factor drives requirements for increased test size.*
- *Generating results for a specific sub-population or target population.* An extension of the parameter above, generating results for a specific sub-population can impact test size as both controlled and generic test results need to be generated to provide a reference point for population-specific data. *This factor drives requirements for increased test size.*
- *Test duration.* A scenario test may need to be executed within a specific timeframe in order to generate results that are meaningful to a report reader. The speed with which biometric systems incorporate new sensors and algorithms can limit the effective shelf life of biometric test data. If constraints are present which limit the time allocated to recruitment, enrollment, matching, results assessment, and reporting, the number of subjects tested may need to be reduced. *This factor drives requirements for reduced test size.*
- *Limiting costs associated with testing.* A testing organization may be working within a specific budget to cover test execution and administration. *This factor drives requirements for reduced test size.*

Factors that drive test size decisions, as well as the impacts (positive and negative) of test size should be incorporated within Test Reports.

## 5.2.10 Composition of Test Population

Scenario testing affords an opportunity to the testing entity to completely control the composition of the test population. This differentiates scenario testing from certain technology and operational tests in which the test population is externally dictated, either through the composition of a target database or through the makeup of an employee or customer base. Controlling for a certain population can provide insights as to performance in a scenario environment for individuals of a particular gender, ethnicity, age, occupation, or other factor of interest.

If one intends to execute testing to determine performance on the part of a target user group, it is necessary to be capable of deriving result for a control group. Without this control group it cannot be determined whether the performance for a target group was anomalous in any fashion or whether target results differed from results gathered across the general population. This control group should be comprised of an equitable distribution of test subjects relative to the target group.

Controlling for a specific target user group in a scenario test is complicated by the expense and effort of executing scenario testing as well as by the need to maintain statistically significant test populations. If one were to test for performance across age groups, for example, it is necessary to test a sufficient number of individuals within each age group such that a small number of subjects do not dramatically skew test results. For this reason it may be difficult to test for a large number of target populations within a given scenario test effort.

The degree to which test population composition is controlled, and the objectives of such control, should be discussed in Test Reports such that readers have an understanding of the degree to which the results of a test effort are extensible to one's own population.

## 5.2.11 Number of Visits / Time between Visits

Both the number of visits and the time between visits can impact scenario test execution and reporting. Scenario testing can be executed with a single visit on the part of test subjects, an approach which reduces the logistical problems of subject management at the cost of generating less robust false non-match data. Within this single visit, the test subject can execute enrollment, false match, and false match testing (not necessarily in that order). Data resulting from this testing would provide results on a system's ability to match users, in effect, immediately subsequent to enrollment – a best-case usage scenario. Since it is suspected that most biometric characteristics are subject to change over time, and that the manner in which subjects interact with systems may be impacted by the introduction of time, then performing true-match testing on one or more later dates to generate FNMR over time. False match testing can also be executed in conjunction with true match testing, although intuitively there is no reason to anticipate that FMR would be any lower or higher on day N than on the day of initial enrollment and false match testing. The discussion of habituation, above, informs considerations around the number of visits involved in a scenario test.

The time between visits is an additional consideration that can impact performance results. Maximizing the time between initial enrollment and subsequent long-term matching will in most cases generate the most pronounced performance challenges. While Best Practices recommends that sufficient time be introduced between a first and last visit (regardless of the number of visits in between) should be commensurate with the projected time required for a biometric characteristic to be regenerated. For example, regrowth of a layer of skin requires approximately 6 weeks, such that fingerprints should be tested at some point approximately 6 weeks subsequent to enrollment. However, such a practice may introduce a degree of subjectivity and speculation in gauging the regeneration of a biometric characteristic. If a scenario can be associated with an inter-visit time lapse consistent with its concept of operations, then such a time lapse can be enforced across all tested systems.

## 5.2.12 Number of Systems and Modalities Tested

Scenario testing can incorporate a single biometric system, multiple biometric systems, or multiple biometric modalities (as well as multiple systems within multiple modalities). Testing multiple systems provides the advantage of potentially establishing a range of performance against which different systems can be evaluated. Anomalous performance can be difficult to gauge from a single-system test. Similarly, testing multiple modalities provides a comparative backdrop for determining whether a test methodology for a given biometric modality was excessively restrictive or forgiving.

When testing multiple systems and/or technologies, system order is a major consideration. As the degree of acclimation to biometric systems can presumably improve within the span of a handful of interactions with biometric systems, the first systems tested within a given modality would likely be at a disadvantage relative to later systems, as users learn within the course of a test session the most effective method of interacting with a given modality. Similarly, an individual may “tire” over the course of a test session, particularly if required to sign or recite pass phrases across a number of systems. Therefore the order in which subjects interact with systems should be arranged such that each system is tested a roughly equivalent number of times in the first position, second position, through to the Nth test position.

In addition, the order in which systems are tested should be adjusted such that no two systems consistently precede or follow each other. Certain systems may require substantial effort, such that users become frustrated and grow unable to concentrate on systems that directly follow. Randomizing system interaction to the degree possible reduces the risk of collecting skewed data.

## 5.2.13 Use of Variable Matching Thresholds

Biometric systems utilize matching thresholds to determine if two or more templates are from the same individual. For many systems, these thresholds are adjustable to reduce FMR/FAR or FNMR/FRR as required. Scenario tests can be conducted against a single matching threshold or against multiple thresholds. When testing multiple systems against single thresholds, test organizations should establish the performance objectives associated with the threshold tested (e.g. minimizing FMR, minimizing FNMR, minimizing EER) such results can be compared across systems.

More variables are involved when testing at multiple matching thresholds. A test organization may decide to execute tests at each security level, moving through the required number of presentations, attempts, and sequences for each system at each security level for both imposter and genuine match tests. A test organization may also decide to test at the level most likely to elicit errors – i.e. at low security for FMR testing and high security for FNMR testing – and only progress to subsequent security levels if an error occurs. The presumption in this case is that is a false match does not occur at low security, then by definition the same presentations would not have generated false matches at medium or high security. Regardless of the approach, the approach to matching thresholds should be enforced in a uniform fashion across all systems and modalities involved in a scenario test.

The discussion of match outputs, below, also informs the use of variable matching thresholds. If match scores are recorded for genuine and imposter transactions, then a test organization can execute a single full transaction, inclusive of the maximum number of permitted attempts, and subsequently determine what error rates would have been generated at low, medium, and high thresholds supplied by the vendor.

**5.2.14 Physiological/Behavioral Data to be Enrolled**

For most biometric modalities, a determination must be made on what data to collect to associate with a test subject. The following table provides some of the options available for applicable biometric disciplines.

Modality	Characteristic Options
Fingerprint	1-10 fingers; finger selection, if less than 10
Iris	1-2 irises; iris selection if less than 2
Signature Verification	Fixed signature; signature of own selection
Voice verification	Fixed spoken phrase; spoken phrase of own selection
Hand Geometry	1-2 hands; hand selection if less than 2 (most deployed systems are designed to enroll only the right hand)

In determining how to test a given modality, a test organization must consider the scenario under evaluation and the general applicability of the characteristic acquisition method. If acquisition of a certain type of sample is dictated by a scenario’s concept of operations, then the test should mirror such an acquisition process. For example, a fingerprint system may be deployed that requires all users to enroll with their right index finger (for example to identify duplicate enrollments). Dictating that all test subjects utilize the right index finger will then provide relevant results. In dictating the physiological data to be utilized, an additional variable of physiological preference may be introduced which impacts data collection. Left-handed users, for example, may not place their right index finger in as intuitive a fashion as their left index, such that more placement effort or guidance is necessary. The same applies to certain iris-based systems, in which location of the eye in a mirrored imaging device is necessary to enroll. Subjects often favor one eye over the other in their everyday life, such that locating a weaker eye in an iris device can be challenging. Dictating the eye to be enrolled can increase failure to enroll and false non-match rates due to the difficulty of interacting with systems with the weaker eye.

However, by not dictating the physiological characteristic to be enrolled, additional problems emerge. First, there is the problem of subject management: the need to track for all imposter and genuine attempts across all visits which biometric characteristic should be provided for which system. This increases the likelihood that provision of the wrong sample on the part of the subject will undermine the validity of match results. If allowed to enroll in a discretionary fashion, users may also switch between right and left hand fingerprints as they move across devices, which limits the comparability of performance across devices (assuming such comparability is a test objective). By dividing a given system into separate tests of right index fingerprints, left index fingerprints, etc., the statistical validity of data for a given system may be reduced.

Therefore, unless there is a compelling reason to allow for discretionary interaction on the part of the test subject, then the physiological characteristics involved in enrollment and verification should be dictated and compliance so enforced; the variable of discretionary sample selection should be minimized to those users lacking a given physiological characteristic (e.g. missing a hand or finger).

The question of characteristic and sample dictation is more pronounced in behavioral-oriented systems such as voice and signature, particularly in the area of imposter testing. In these modalities, the subject provides biometric data over a period of time and is also capable of masking or altering the submission in a fashion difficult to detect from the perspective of the test operator. There is therefore a much broader range of variable and/or discretionary subject-system interaction than with physiological biometric systems such as fingerprint, face, and iris (though the distinction is not absolute). In determining what sample to utilize, a test organization can allow the user to select a sample – for the purposes of this discussion, a signature or voice sample. This is consistent with behavioral systems in which the user provides a sample of his or her “choosing”,

such as the signing of a name with or without a middle initial or the selection of a secret password for a voice-protected account.

However, by not dictating the content of the behavioral characteristic to be tested, other problematic variables are introduced, including the following:

Tracking sample data utilized by subject. If users are allowed to select their enrollment sample, then some method of tracking this sample is necessary. The test is then informed by the test operator's ability to accurately record the sample provided by the subject.

*Interpretation of enrollee data on the part of the imposter.* If users are allowed to enroll in a voice system with their name as opposed to a fixed and easily comprehensible phrase, then the testing may be heavily influenced by the imposter's ability to read and correctly pronounce a person's name. This mandates usage of a phrase which all imposters will be capable of reciting without error.

A central object of biometric testing is to determine the ability of biometric systems to differentiate or to determine the degree of similarity between two individuals' characteristics, as opposed to determining an imposter's ability to guess at what sample an enrollee might have used. An imposter in a test effort must be assumed to have some knowledge of the biometric characteristic against which he or she is attempting to break into.

Therefore, unless there is a compelling reason to allow for discretionary interaction on the part of the test subject, then the behavioral characteristics involved in enrollment and verification should be dictated and compliance so enforced; the variable of discretionary sample selection should be minimized. A consistent sample, such as a specific passphrase, should be utilized for all subjects. It may be desirable to introduce several fixed passphrases to utilize for enrollments to ensure that the specific phrase utilized does not skew data, but such a practice would be contingent on test size (larger tests allow for more variation in enrollment phrases).

Further complications in behavioral biometric testing cannot be readily resolved through standards, and can only reasonably be reported on in test results presentation. For example, the manner in which subjects are instructed to break in as imposters against enrolled users can impact test results. Assuming that an imposter may be allowed to view a signature against which he or she attempts to validate, then this signature may be present for the entire time or may be removed once the imposter attempt begins. If verifying against a voice system, the test operator may recite the phrase to be used for matching; the manner in which the phrase is read may have a more dramatic impact than similar demonstration of a physiological characteristic.

## **5.3 Test Execution**

### **5.3.1 System Configuration**

Biometric systems are configured prior to scenario tests to ensure stable platforms for operations, regardless of whether the scenario being tested is logical access, physical access, point of sale, surveillance, or another biometric scenario. The following are among the considerations involved in system configuration for scenario tests.

As opposed to technology testing, in which a single platform is leveraged to test multiple systems, and operational testing, in which only one device may be evaluated, scenario testing may entail multiple systems being tested on different platforms, ranging from standalone devices to multi-processor workstations. While in certain scenario tests a single test platform, e.g. a series of identical PCs, can be established for N systems, other scenario tests allow for provision of full systems on the part of technology providers.

It is incumbent on the test organization to ensure that vendors do not gain unfair advantages by providing systems whose computing power is grossly incommensurate with the scenario being tested. At the same time, a scenario test may be capable of fairly accommodating systems with divergent computing platforms and processing architectures. For example, in a logical access scenario test, one could envision a test inclusive of smart-card based matching, local PC-based matching, and server-based matching systems. The litmus test is to determine in an objective fashion whether a proposed test system would indeed be reasonably implemented in a fashion consistent with how it was tested. In addition, the metrics associated with a test may inform the degree of flexibility accorded vendors in providing test systems. If match speed is not a primary consideration, as would be the case in many 1:1 tests, then less of an advantage is gained by supplying a PC with excessive computing power. In all cases the details of the systems used for testing, whether at the vendor's discretion or standardized across all systems, should be provided in the test report.

### **5.3.2 System Maintenance**

- Emergency measures: backup and recovery, power failure, device failure
- Processes for identifying system failures (dead scanners, etc.)

### **5.3.3 Scheduling and Test Subject Management**

- Test subject management can be a challenge in scenario testing, as users may need to be accompanied through the course of a test effort and a limited number of subjects are likely to be interacting with test systems at any given time. Test subject management should be designed such that each test subject's experience in interacting with the testing organization

### **5.3.4 System Ordering**

Scenario testing is generally designed to derive, at a minimum, false match and false non-match rates (and corresponding false accept and false reject rates for 1:N testing). Scenario testing is also capable of generating failure to enroll rates. Test results can be impacted by the order in which testing is executed, such that consideration must be given to ordering and results reporting. It is a given that enrollment must precede "true-match" testing from which FNMR is derived. False match testing can occur with the subject's first interaction with the system, preceding enrollment, can take place between enrollment and genuine authentication, or can take place after enrollment and genuine authentication. At a minimum, the order in which such testing occurs should be recorded and reported if not controlled for test subjects. A preferred approach is to control the test process such that enrollment, true matching, and false matching always occur in a certain order. A consideration in this area is the fact that an enrollment in certain biometric systems may impact the composition of a global reference biometric. This can occur in neural net systems in which the biometric system "learns" the characteristics of it enrolled subjects and uses this information to execute matching. If such systems are tested then imposter match results are impacted by the fact that a subject is already enrolled; false match testing would need to be executed prior to enrollment to eliminate this impact.

This discussion assumes that test subjects execute both true match and false match testing, a characteristic of many scenario tests. If resources allow, a preferred approach may be to ensure that the genuine and imposter populations are disjoint.

### **5.3.5 Test Operator Roles and Responsibilities**

Test Operators are individuals charged with the management and execution of testing, and are the primary point of interface between the test organization and test subjects. Test Operators are an important part of scenario testing, as the roles and responsibilities of Test Operators can directly impact test results.

Test operators do not necessarily *need* to be present during enrollment and matching, depending on (1) the nature of the scenario being tested, (2) the degree of supervision of test processes required on the part of the test organization, and (3) resources available. However, the reliability of test data is greatly enhanced through close monitoring of test subjects' interaction with biometric systems on the part of Test Operators. Certain data points, such as failures to acquire data during presentations, can only viably be manually acquired by Test Operators. Self-reporting on the part of Test Subjects is a questionable practice, as Test Subjects may fail to understand what data to report or may report incorrectly for any variety of reasons.

Test Operator responsibilities include but are not limited to the following:

- Ensuring that systems are operating properly during the course of testing, i.e. looking for application- or device-level anomalies
- Directing Test Subjects in their interaction with systems and devices
- Recording number of attempts and sequences permitted to match at each security level against each target Test Subject
- Recording visually rendered scores resulting from biometric comparisons
- Enforcing placement and time limitations during enrollment and matching sequences
- Recording enrollment and match sequence times
- Observing interaction with systems and devices to note and if necessary correct incidents of gross misrepresentation
- Collecting and/or ensuring the accuracy of biographical data
- Recoding and if necessary correcting incidences of egregiously incorrect biometric sample submission
- Unsolicited comments provided by the Test Subject

### **5.3.6 Directing Test Subjects in Enrollment and Matching**

The degree of direction, both instructive and corrective, provided by Test Operators to Test Subjects can bear directly on scenario test results. In most biometric scenarios, important usage instructions are provided during initial enrollment, which is often a supervised process. Matching, on the other hand, is often unsupervised, such that providing instructions prior to or in conjunction with matching may be inconsistent with real-world operations. However, numerous exceptions to this rule could be cited: enrollment may be hastily executed in certain applications due to processing constraints, with subsequent matching processes heavily supervised. Therefore the primary requirements in this area are (1) to ensure that the level of direction is consistent across all systems, such that no system is given a performance advantage due to the quality or extent of instruction during enrollment and matching and (2) to ensure that Test Operators carefully adhere to established scripts and practices as opposed to providing discretionary advice and direction.

### **5.3.7 Data Collection**

- Automated vs. manual data collection
- Test subject response/user perception

## 6. Scenario Test Reporting

In addition to the normative and informative Reporting elements addressed in Part 1 of the Standard, the following normative and informative Reporting elements apply to Scenario Testing.

### 6.1 Systems and environment

The test report shall document system and environment-related elements of the Scenario under evaluation, including at a minimum the following:

#### 6.1.1 Biometric system specifications

For each biometric system tested, the test report shall document the following elements:

- Acquisition devices: manufacturer, model, version, and firmware as applicable. If the acquisition device tested as part of the biometric system is a peripheral or device with a 3<sup>rd</sup> party modular sensor, such as a fingerprint sensor, then the sensor manufacturer, model, version, and firmware as applicable shall be documented in the test report.
- Biometric algorithms: provider, version, revision as applicable.
- Biometric applications: provider, version, build as applicable.
- Use of off-the-shelf acquisition and processing components vs. components customized or optimized for test effort.
- Biometric data acquisition, processing, and storage architecture, specifying (1) components tested to comprise the system, (2) data flow between components, and (3) location of sample processing functions, template generation processes, template matching processes, and template storage.

#### 6.1.2 Platform specifications

As applicable for each biometric system tested, the test report shall document the following elements of the platform on or through which the biometric systems were evaluated:

- Operating system and database
- Hardware specifications including CPU speed, RAM, and hard drive capacity
- Network type and throughput  
*e.g.* 10/100 LAN

#### 6.1.3 System implementation

For each biometric system tested, the test report shall document system implementation information corresponding to each of the following:

- Method of biometric and platform system acquisition  
*e.g.* use of previously fielded systems such as those from within a test organization; procurement of systems from vendors; use of vendor-loaned systems
- Vendor involvement in system implementation  
*e.g.* hardware and software installed and configured offsite by vendor, then shipped to Test Organization; hardware and software installed and configured onsite by vendor; *e.g.* hardware and software installed and configured by Test Organization

#### 6.1.4 Physical layout of test environment

The test report shall document the physical layout of the test environment, providing at a minimum information as follows:

- Dimensional area dedicated to scenario test execution

# INCITS M1/03-0563

- Presence of natural and artificial lighting
- Positioning of biometric acquisition devices  
e.g. all devices positioned on desks of equal height; devices mounted at identical fixed height; devices mounted as per vendor instructions
- Use of chairs or other apparatus on the part of test subjects
- Relative location of each system in the test environment, rendered through a system schematic

## 6.1.5 Operational environment

The test report shall document operational conditions in the Test environment corresponding to each of the following:

- Temperature
- Lighting, including type, direction, intensity
- Humidity
- Variations in any operational environment element over time or from system to system
- Methods by which and frequency with which operational environment data was measured and logged

## 6.1.6 Match and enrollment outputs

For each biometric system tested, the test report shall document whether the biometric systems evaluated utilized 1:1 functionality and/or 1:N functionality.

For each 1:1 system tested, the test report shall document each of the following:

- Type of match outputs reported  
e.g. raw match scores; BioAPI match scores; match/no match decisions
- For systems reporting match scores, range of potential match scores
- For systems reporting match scores, vendor-reported thresholds used for purposes of match score processing  
e.g. low, medium, and high security
- mapping of match thresholds to match scores to execute match decisions
- Method(s) through which match outputs were provided by the biometric system  
e.g. logged by application; visually indicated through GUI/LCD

For each 1:N system tested, the test report shall document the following:

- Type of match outputs reported  
e.g. candidate lists; single-record best match; match scores
- For systems reporting match scores, range of potential match scores (if known)
- Method(s) through which match outputs were provided by the biometric system  
e.g. logged by application; visually indicated through GUI/LCD
- Use of open set and/or closed set identification during imposter trials

For all systems, the test report shall document the following:

- Type of enrollment quality scores reported, if applicable  
e.g. numerical raw enrollment quality scores; generic ranges such as low, medium high
- For systems reporting numerical or range-based enrollment quality scores, range of potential enrollment quality scores (if known)
- Method(s) through which enrollment quality outputs were provided by the biometric system  
e.g. logged by application; visually indicated through GUI/LCD

## 6.1.7 Scenario Narrative

The test report shall incorporate a narrative that describes the scenario test as well as the fundamental elements that comprise the concept of operations under evaluation. The narrative shall, at a minimum, address the following:

- Scenario evaluated  
e.g. identification – surveillance, screening, duplicate detection  
e.g. verification – logical access, physical access
- Test results generated  
e.g. enrollment, false accept, false reject, false match, false non-match
- Number of systems tested
- Number of technologies tested
- Rationale for system selection
- Attended vs. unattended operation
- Overt vs. covert biometric data acquisition
- Use of acclimated vs. non-acclimated subjects

**Note: it is assumed that requirements for documenting test subject population composition will be addressed in Part 1.**

## 6.2 Test processes

The test report shall document process-related elements of the Scenario under evaluation, including at a minimum the following:

### 6.2.1 Test subject management

The test report shall document information related to test subject management, including at a minimum the following:

- Method of determining order in which test subjects interact with systems
- Sequence of subject interaction with systems  
e.g. enrollment in all systems, followed by imposter matching in all systems, through to process N in all systems  
e.g. enrollment and matching in system A, followed by enrollment and matching in system B, through to enrollment and matching in system N.
- Typical and maximum number of test subjects and test operators within Test Environment during test process
- Method of selecting gallery record against which imposter attempts are executed
- Amount and type of personal data collected  
e.g. name, race, gender, occupation

### 6.2.2 Performance data collection and management

The test report shall document the following information:

- Methods of recording data for each performance element:  
e.g. manual; automated through custom-built logs; automated through vendor-designed logs
- Processes for auditing and validating performance data collection
- Use of online and offline sample processing, template generation, and template matching functions

### 6.2.3 Use of Identifiers

For each biometric system tested, the test report shall document information as related to identifiers and identity claims.

# INCITS M1/03-0563

For each 1:1 system tested, the test report shall document each of the following:

- Mechanisms by which test subjects claimed identities for the purposes of 1:1 matching  
e.g. username, PIN, token
- Manner in which this mechanism was used to claim an identity  
e.g. test subject-provided; test operator-provided
- Impact of identity claim processes on test subject interaction with systems  
e.g. if test subject-system interaction was timed, did timing of interaction commence with identity claim; did erroneous identity claims impact results
- Variances in identity claim processes for genuine and imposter trials

For each 1:N system tested, the test report shall document the following:

- Type and range of unique identifiers used for test subjects  
e.g. usernames, ID numbers
- Manner in which this mechanism was used to claim an identity  
e.g. test subject-provided; test operator-provided
- Impact of identity claim processes on test subject interaction with systems  
e.g. if test subject-system interaction was timed, did timing of interaction commence with identity claim; did erroneous identity claims impacted any test execution or results

## 6.2.4 Enrollment processes

For each biometric system tested, the test report shall document the following:

Basic methods and policies

- Methods by which test subjects provide enrollment data  
e.g. live interaction with device
- Type of biometric data stored during enrollment  
e.g. samples and/or templates
- Number of distinct biometric samples enrolled (e.g. 2 fingerprints, 2 irises)
- Note: test report should also indicate whether multiple sample acquisition was due to system requirements or at the discretion of the test designer
- If multiple samples collected, policies for usage and processing
- Specification of which biometric sample(s) are enrolled  
e.g. right iris, fixed voice pattern, variable voice pattern
- Policy for use of fallback samples if primary samples cannot be enrolled  
e.g. reversion to alternative fingerprint or signature pattern
- Policy for use of test subjects unable to enroll in execution of false match testing
- Use of reference model-based methods by which test subject enrollment data informs basis of match results
- Execution of 1:N searches upon enrollment to identify similar records  
Note: systems that execute such 1:N matching may prevent the enrollment of the 2nd sample
- Number of times each test subject enrolls  
e.g. can test subjects enroll more than once for any reason

Definition of enrollment events

- Definition of enrollment presentation
- Definition of enrollment attempt
- Definition of enrollment sequence
- Definition of enrollment trial
- Minimum number of presentations necessary to constitute an enrollment attempt
- Minimum number of attempts necessary to constitute an enrollment sequence

# INCITS M1/03-0563

- Minimum number of sequences necessary to constitute an enrollment trial
- Maximum number of presentations permitted to constitute a match attempt
- Maximum number of attempts permitted to constitute an enrollment sequence
- Maximum number of sequences permitted to constitute an enrollment trial
- Criteria by which successful enrollment is declared
- Criteria by which failure to enroll is declared
- Time-out policies for systems unable to acquire data, including system-based and test operator-based

## Guidance and instruction

- Attended/unattended nature of enrollment
- Information provided by test operator to test subjects prior to enrollment regarding Test effort and biometric system operations
- Range of feedback provided by biometric system to user during enrollment process
- Sequence and content of instructions pertaining to interactions with each system provided by test operator to test subjects prior to enrollment
  - e.g. are test subjects instructed in interaction with each system prior to interaction with any system, or are test subjects instructed in interaction with each system immediately prior to interaction
- Range of guidance, if any, test operator was permitted to provide to test subjects during enrollment process
- Use of vendor-provided scripts, instructions, guidance tools, or other mechanisms to inform test subjects as to the optimal method(s) of interacting with the system during enrollment
- Processes implemented in case of egregiously improper interaction with device, such as speaking into wrong end of telephone
  - e.g. disregard presentation, count as presentation
- Constraints on test subject appearance and apparel
  - e.g. removal of glasses, hats, contacts
- Composition of gallery database, including source(s) of gallery data
  - e.g. enrollment data collected in same environment as testing, from same subjects; enrollment data collected from external databases

### 6.2.5 Matching processes

For each biometric system tested, the test report shall document the following:

#### Basic methods and policies

- Imposter matching methodology
  - e.g. real-time test subject matching against a single gallery record, with new data acquired for each trial against a new gallery record; recording of scenario-based sample data for automated matching vs. N gallery records
- Genuine test subject matching methodology
  - e.g. real-time test subject matching against self-same gallery record; recording of scenario-based sample data for automated matching vs. N self-same gallery records
- Number of imposter trials executed by each test subject (probe)
- Number of imposter trials executed against each test subject (gallery)
- Methods by which test subjects provide matching data
  - e.g. live interaction with device
- For genuine and imposter trials, policy for matching against fallback enrollments if primary samples are unavailable
- For genuine and imposter trials, policy for matching against fallback enrollments if unable to match against primary enrollment
- For genuine and imposter trials, policy for executing additional presentations, attempts, and sequences subsequent to successful match

# INCITS M1/03-0563

*e.g.* in an imposter trial, after a successful match, does the test subject continue to repeat placements against the same enrolled imposter data to gather more data

- For each genuine match event, time elapsed since initial enrollment

## Guidance and instruction

- Attended/unattended nature of matching
- Range of guidance, if any, test operator was permitted to provide to test subjects during matching process
- Sequence and content of instructions pertaining to interactions with each system provided by test operator to test subjects prior to matching  
*e.g.* are test subjects instructed in interaction with each system prior to interaction with any system, or are test subjects instructed in interaction with each system immediately prior to interaction
- Use of vendor-provided scripts, instructions, guidance tools, or other mechanisms to inform test subjects as to the optimal method(s) of interacting with the system during enrollment
- Processes implemented in case of egregiously improper interaction with device, such as speaking into wrong end of telephone  
*e.g.* disregard presentation, count as presentation

## Definition of match events

- Definition of match presentation
- Definition of match attempt
- Definition of match sequence
- Definition of match trial
- Minimum number of presentations necessary to constitute a match attempt
- Minimum number of attempts necessary to constitute a match sequence
- Minimum number of sequences necessary to constitute a match trial
- Maximum number of presentations permitted to constitute a match attempt
- Maximum number of attempts permitted to constitute a match sequence
- Maximum number of sequences permitted to constitute a match trial
- For genuine trials, criteria by which successful match is declared
- For genuine trials, criteria by which false non-match is declared
- For imposter trials, criteria by which correct non-match is declared
- For imposter trials, criteria by which false match is declared
- Presence of time-out logic

## 6.3 System-Level results

System-level results provide information regarding each biometric systems' ability to enroll test subjects as well as to match or identify test subjects in their role as genuine users and/or imposters. For each biometric system tested, the test report shall document the following:

### 6.3.1 System-Level enrollment results

For each biometric system tested, the test report shall document the following:

- Number of subjects attempting to enroll
- Total number and percentage of test subjects unable to enroll
- Number and percentage of test subjects able to enroll at each incremental presentation interval  
*e.g.* if a minimum of 2 placements are required to successfully enroll, and a maximum of 9 placements are permitted to successfully enroll, test report shall document number and percentage of test subjects able to enroll within 2 placements, 3 placements, continuing to report up to 9 placements.

## INCITS M1/03-0563

- Number and percentage of test subjects able to enroll at each incremental attempt interval e.g. if a minimum of 1 attempt is required to successfully enroll, and a maximum of 3 attempts are permitted to successfully enroll, test report shall document number and percentage of test subjects able to enroll within 1 placement, 2 placements, and 3 placements.
- Number and percentage of test subjects able to enroll at each incremental sequence interval e.g. if a minimum of 1 sequence is required to successfully enroll, and a maximum of 2 sequences are permitted to successfully enroll, test report shall document number and percentage of test subjects able to enroll within 1 sequence and 2 sequences.
- Number and percentage of test subjects unable to enroll due to lack of biometric characteristic
- Number and percentage of test subjects unable to enroll due to system's inability to acquire minimum number of samples necessary to constitute an enrollment  
Note: System-driven time-outs and operator-driven time-outs shall be reported separately
- Number and percentage of test subjects unable to enroll due to low-quality enrollment scores subsequent to acquisition
- Number and percentage of test subjects able to partially enroll in systems requiring two distinct samples
- Number and percentage of enrollees able to enroll via fallback biometric, if applicable
- Instances of duplicate detection in 1:N enrollment search, if applicable
- For attended tests in which test operator was permitted to provide guidance to test subjects during enrollment process, number of test subjects for whom such guidance was provided.
- Time required to enroll / duration after which FTE was declared, measured from initiation of test subject interaction with acquisition device.
- Time required to enroll / duration after which FTE was declared
- Number of incidences of egregiously improper interaction with devices and resolution of such instances
- Range and frequency of enrollment quality scores reported for successful enrollees, if reported
- Range and frequency of enrollment quality scores reported for unsuccessful enrollees, if reported
- Range and frequency of enrollment quality scores reported for all enrollees, if reported

### 6.3.2 System-level 1:1 match results

For each 1:1 biometric system tested, the test report shall document the following:

#### Genuine trials

- Number of genuine trials executed
- Total number and percentage of test subjects falsely rejected
- Number and percentage of test subjects able to match at each incremental presentation interval e.g. if a minimum of 1 placement is required to successfully match, and a maximum of 9 placements are permitted to successfully match, test report shall document number and percentage of test subjects able to match within 1 placements, 2 placements, continuing to report up to 9 placements.
- Number and percentage of test subjects able to match at each incremental attempt interval e.g. if a minimum of 1 attempt is permitted to successfully match, and a maximum of 3 attempts are permitted to successfully match, test report shall document number and percentage of test subjects able to match within 1 placement, 2 placements, and 3 attempts.
- Number and percentage of test subjects able to match at each incremental sequence interval e.g. if a minimum of 1 sequence is required to successfully match, and a maximum of 2 sequences are permitted to successfully match, test report shall document number and percentage of test subjects able to match within 1 sequence and 2 sequences.

## INCITS M1/03-0563

- Number and percentage of test subjects unable to match due to system's inability to acquire a sample
- Number and percentage of test subjects unable to match due to match scores below minimum matching threshold
- Number and percentage of test subjects unable to match against primary enrollment subsequently able to match against fallback enrollment
- Number and percentage of test subjects unable to match against primary enrollment subsequently unable to match against fallback enrollment
- For attended tests in which test operator was permitted to provide to provide guidance to test subjects during matching process, number of test subjects for whom such guidance was provided
- Number of incidences of egregiously improper interaction with devices and resolution of such instances
- Range and frequency of average time lapsed between test subjects' enrollment and matching events.
- For systems in which minimum and maximum template aging between test subjects enrollment and match data varies by more than 25%, test report shall document the number and percentage of test subjects associated with each of series of equivalent template aging ranges such that no one range accounts for more than 20% of the total enrolled test subject population. The test report shall document the average match score for each of these equivalent template aging ranges.
- Time required to match / duration after which non-match was declared, measured from initiation of test subject interaction with acquisition device
- Range and frequency of match scores reported for successful matches, if reported
- Range and frequency of match scores reported for unsuccessful matches , if reported
- Range and frequency of match scores reported for all enrollees, if reported

### Imposter trials

- Number of imposter trials executed
- Number and percentage of test subjects able to match at each incremental presentation interval (FAR)  
e.g. if a minimum of 1 placement is required to successfully match, and a maximum of 9 placements are permitted to successfully match, test report shall document number and percentage of test subjects able to match within 1 placements, 2 placements, continuing to report up to 9 placements.
- Number and percentage of test subjects able to match at each incremental attempt interval  
e.g. if a minimum of 1 attempt is permitted to successfully match, and a maximum of 3 attempts are permitted to successfully match, test report shall document number and percentage of test subjects able to match within 1 placement, 2 placements, and 3 attempts.
- Number and percentage of test subjects able to match at each incremental sequence interval  
e.g. if a minimum of 1 sequence is required to successfully match, and a maximum of 2 sequences are permitted to successfully match, test report shall document number and percentage of test subjects able to match within 1 sequence and 2 sequences.
- Total number and percentage of test subjects unable to match
- Number and percentage of test subjects unable to match due to system's inability to acquire a sample
- Number and percentage of test subjects unable to match due to match scores below minimum matching threshold
- For attended tests in which test operator was permitted to provide to provide guidance to test subjects during matching process, number of test subjects for whom such guidance was provided.
- Number of incidences of egregiously improper interaction with devices and resolution of such instances
- Time required to match / duration after which non-match was declared, measured from initiation of test subject interaction with acquisition device

# INCITS M1/03-0563

- Range and frequency of match scores reported for successful matches, if reported
- Range and frequency of enrollment quality scores reported for unsuccessful matches, if reported

## Summary Results

- Receiver Operating Characteristic Curves based on range and frequency of FAR/FRR/match scores

### 6.3.3 System-level 1:N match results

For each 1:N biometric system tested, the test report shall document the following:

- Number of trials executed with test subject in gallery (i.e. closed-set trials)
- Number of trials executed with test subject not in gallery (i.e. open-set trials)
- Number and percentage of test subjects identified at each of N positions (n=database size) at each incremental presentation interval  
e.g. if a minimum of 1 placement is required to successfully match, and a maximum of 9 placements are permitted to be identified, test report shall document number and percentage of test subjects identified at each of N ranks within 1 placements, 2 placements, continuing to report up to 9 placements.
- Range and frequency of match scores
- Based on the match scores associated with the comparisons above, the test report shall provide the following information at matching thresholds sufficient to represent more than multiple security levels.
- Number and percentage of “test subject in gallery” trials scored above threshold (a correct match)
- Number and percentage of “test subject in gallery” trials scored above threshold as a different test subject (an incorrect match)
- Number and percentage of “test subject in gallery” trials not scored above threshold (an incorrect non-match)
- Number and percentage of “test subject not in gallery” trials scored above threshold (an incorrect match)
- Number and percentage of “test subject not in gallery” trials not scored above threshold (a correct non-match)
- Number and percentage of test subjects able to match at each incremental sequence interval  
e.g. if a minimum of 1 sequence is required to successfully match, and a maximum of 2 sequences are permitted to successfully match, test report shall document number and percentage of test subjects able to match within 1 sequence and 2 sequences.
- Number and percentage of test subjects for whom system was unable to acquire a sample
- Number and percentage of test subjects not identified against primary enrollment subsequently able to match against fallback enrollment
- Number and percentage of test subjects not identified against primary enrollment subsequently unable to match against fallback enrollment
- For attended tests in which test operator was permitted to provide to provide guidance to test subjects during matching process, number of test subjects for whom such guidance was provided.
- Number of incidences of egregiously improper interaction with devices and resolution of such instances
- Time lapsed between test subjects’ enrollment and matching events
- Time required to match / duration after which non-match was declared, measured from initiation of test subject interaction with acquisition device
- For systems in which minimum and maximum template aging between test subjects enrollment and match data varies by more than 25%, test report shall document the number and percentage of test subjects associated with each of series of equivalent template aging ranges such that no one range accounts for more than 20% of the total enrolled test subject

# INCITS M1/03-0563

population. The test report shall document the average match score for each of these equivalent template aging ranges.

- Range and frequency of match scores reported for successful matches, if reported
- Range and frequency of enrollment quality scores reported for unsuccessful matches, if reported

## 6.4 Subject-level results

Subject-level results provide information regarding each test subjects' ability to enroll, genuine-user system interactions, and imposter system interactions. For each test subject, the test report shall document the following:

### 6.4.1 Subject-Level enrollment

For each test subject, the following data shall be documented:

- Number of enrollment trials
- Number and percentage of systems in which the test subject was able to enroll
- Number of placements, attempts, and sequences required to enroll in each system
- Enrollment quality scores for each system, plotted against range of other test subjects' scores for each system
- For systems in which test subjects were unable to enroll through a primary biometric, the test subject's ability to enroll via a fallback biometric
- For systems in which test subject was unable to enroll, number and percentage of systems in which failure to enroll was attributable to failure to acquire
- For systems in which test subject was unable to enroll, number and percentage of systems in which failure to enroll was attributable to insufficient enrollment quality scores
- Number and percentage of systems with which the test subject was found to have egregiously improper interaction

### 6.4.2 Subject-Level 1:1 match results

For each test subject, the following data shall be documented:

For each probe subject

- Number of false match and true match trials executed, by system
- FRR and correct match rate for genuine transactions, by system
- The FAR and correct non-match rate for imposter transactions, by system
- Number and percentage of systems in which test subject was able to match correctly
- Number of placements, attempts, and sequences required to match in each system
- Match scores for each system, plotted against range of other test subjects' scores for each system
- For systems in which test subject was unable to match through a primary biometric, test subject's ability to match via a fallback biometric
- For systems in which test subject was unable to match, the number and percentage of systems in which failure to match was attributable to failure to acquire
- For systems in which test subjects were unable to match, the number and percentage of systems in which failure to match was attributable to insufficient scores
- Number and percentage of systems with which the test subject was found to have egregiously improper interaction

For each gallery subject

- Number of false match and true match trials executed against, by system

## INCITS M1/03-0563

- Number and percentage of successful and unsuccessful imposter attempts against, by system
- Number and percentage of successful and unsuccessful genuine attempts against, by system
- Number of placements, attempts, and sequences required to match against, by system
- Range and frequency of match scores for each test subject against the gallery subject

### 6.4.3 Subject-level 1:N match results

- Number of open and closed set trials executed, by system
- Number and percentage of systems in which test subject was identified at each of N positions (n=database size) at each incremental presentation interval
- Number and percentage of systems in which “test subject in gallery” trials scored above threshold (a correct match)
- Number and percentage of systems in which “test subject in gallery” trials scored above threshold as a different test subject (an incorrect match)
- Number and percentage of systems in which “test subject in gallery” trials not scored above threshold (an incorrect non-match)
- Number and percentage of systems in which “test subject not in gallery” trials scored above threshold (an incorrect match)
- Number and percentage of systems in which “test subject not in gallery” trials not scored above threshold (a correct non-match)