INTERNATIONAL STANDARD

ISO/IEC 19795-5

First edition
2011-03-01

Information technology — Biometric performance testing and reporting —

Part 5: Access control scenario and grading scheme

Technologies de l'information — Essais et rapports de performance biométriques —

Partie 5: Plan de classement pour évaluation de scénario de contrôle d'accès

Reference number
ISO/IEC 19795-5:2011(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 19795-5 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

ISO/IEC 19795 consists of the following parts, under the general title *Information technology — Biometric performance testing and reporting*:

— *Part 1: Principles and framework*

— *Part 2: Testing methodologies for technology and scenario evaluation*

— *Part 3: Modality-specific testing* [Technical report]

— *Part 4: Interoperability performance testing*

— *Part 5: Access control scenario and grading scheme*

— *Part 6: Testing methodologies for operational evaluation*

— *Part 7: Testing of on-card biometric comparison algorithms*

Introduction

This part of ISO/IEC 19795 is concerned solely with the scientific “technical performance testing” of biometric systems and subsystems to be used for access control. Technical performance testing seeks to determine error rates and transaction times with the goal of understanding and predicting the real-world error and transaction times of a biometric system. The error rates include false accept rate, and false reject rate, as well as failure to enrol (FTE) and failure to acquire (FTA) rates across the test population. These measures are generally applicable to all access control systems that contain a biometric verification subsystem.

This part of ISO/IEC 19795 defines a testing framework with the following fundamental aspects.

— This part of ISO/IEC 19795 was conceived to be a framework for a general- or multi-purpose test: “one size fits many (but not all)”. The focus is limited to access control applications.

— The framework is suitable as both a requirements statement and an evaluation report.

— The general-purpose nature of this part of ISO/IEC 19795 is centred on the common access control application requirements, and acknowledges the fact that this framework will not be suitable for specialized applications (very high levels of protection, specialized user populations like the elderly, students, etc.). Specialized applications will warrant specialized testing processes.

— The perceived benefit of the general- or multi-purpose test is economy. The supplier can submit to one testing process, and many potential customers can utilize the results, interpreting the suitability of the device (based on the results) for their application.

This testing framework assigns grades representing the tested level of performance, and these grades include a statistical confidence taking the conservative approach, that is, the performance of the system is at least as good as the grade indicated (at the 90% confidence level). Using the grading scheme to specify a required performance level of a system needs to take into account this conservative approach.

It is acknowledged that technical performance testing is only one form of biometric testing. Other types of testing not considered in this part of ISO/IEC 19795 include the following:

— reliability, availability and maintainability;

— security, including vulnerability;

— human factors, including user acceptance;

— environmental;

— safety;

— cost/benefit;

— privacy regulation compliance.

Methods and philosophies for these other types of tests are currently being considered internationally by a broad range of groups.

The purpose of this part of ISO/IEC 19795 is to capture the current understanding by the biometrics community of requirements and best scientific practices for conducting performance testing towards the end of providing consistent, structured evaluations of biometric systems intended for use in access control applications. The framework defined in this part of ISO/IEC 19795 has utility as a method for defining user requirements, for specifying the extent of performance evaluation, for conducting and for reporting.

Information technology — Biometric performance testing and reporting —

Part 5: Access control scenario and grading scheme

1 Scope

This part of ISO/IEC 19795:

— defines a common biometric access control scenario for use in scenario evaluation of biometric verification systems;

— provides a grading scheme for expressing quantitative biometric system requirements and performance levels;

— provides a common basis for conducting scenario evaluations to demonstrate that specified performance grades are being achieved which is adaptable to particular testing facilities and to specific biometric systems.

This part of ISO/IEC 19795 is applicable to performance testing of biometric systems without detailed knowledge of the comparison algorithms or of the underlying distribution of biometric characteristics in the population of interest.

The minimum false accept rate (FAR) tested by this part of ISO/IEC 19795 is 0.1%. If a lower FAR is required, customized testing (outside the scope of this part of ISO/IEC 19795) might be appropriate, and needs to be fully compliant with ISO/IEC 19795-2.

This part of ISO/IEC 19795 addresses testing a biometric system for physical access control, and the suitability of the testing for logical access devices needs to be determined on a case-by-case basis.

Not within the scope of this part of ISO/IEC 19795 is the measurement of error and throughput rates for people deliberately trying to circumvent correct recognition by the biometric system (i.e. active impostors). In addition, this part of ISO/IEC 19795 does not assess the following:

— reliability, availability and maintainability;

— security, including vulnerability;

— human factors, including user acceptance;

— environmental impacts;

— safety;

— cost/benefit/suitability;

— privacy regulation compliance.

These assessments are the responsibility of the procuring authority.

2 Conformance

A test conforms to this part of ISO/IEC 19795 if the scenario used (including test crew demographics, environmental controls, time separation between enrolment and revisit, numbers of attempts and transactions), test conduct, and test reporting all conform to the mandatory requirements in Clauses 5 through 7.

3 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 19795-1:2006, *Information technology — Biometric performance testing and reporting — Part 1: Principles and framework*

ISO/IEC 19795-2:2007, *Information technology — Biometric performance testing and reporting — Part 2: Testing methodologies for technology and scenario evaluation*

ISO/IEC TR 19795-3, *Information technology — Biometric performance testing and reporting — Part 3: Modality-specific testing*

4 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 19795-1 and the following apply.

4.1
access control system
ACS
entire electro-mechanical suite that performs the granting or denying of access at controlled entry points of a facility

4.2
biometric subsystem
portion of a biometric system that is present at each access entry point, including the biometric sensor or sampling subsystem

4.3
grade levels
measurement associated with the quantified levels of biometric subsystem performance

NOTE Grade levels are defined, ranging from 0 to 3, or 0 to 6. It is possible that additional grade levels above these values will be defined at a future date.

4.4
FAR level
scale for the relative level of resistance to false accepts in a form associated with three specific false accept rate (FAR) values

4.5
transaction time
time required for the biometric system portion of an access control transaction

NOTE Transaction time is measured in seconds.

5 Definition of testing scenario

5.1 Overview

The goal of testing and evaluating biometric access control systems against the standard set of criteria documented in this part of ISO/IEC 19795 is to ensure that the technical performance of every biometric access control system is evaluated fairly, accurately and equivalently.

Testing shall be performed in a consistent, unbiased manner under conditions that are well understood and documented. Test controls shall be applied to ensure reproducible test results to the most practical extent possible (considering the involvement of human crew members). To accomplish this, every candidate biometric access control system shall be tested in accordance with the same general test protocol.

The procedures to be used shall be based upon a "framework" consisting of specific metrics extracted from biometric system operations and accompanying evaluation criteria which provides for graded evaluation against different levels of false accept rate. The evaluation framework shall accommodate biometric subsystems that output similarity scores or that output only the final match/no match decision.

NOTE 1 Throughout this part of ISO/IEC 19795, where reference is made to similarity scores, it should be understood that for those test results in the form of decision output, the equivalent, suitable process is applied.

NOTE 2 Throughout this part of ISO/IEC 19795, where reference is made to similarity scores, it should be understood that devices that generate dissimilarity scores will be accommodated by making the appropriate threshold comparison matching decisions.

To facilitate the testing of a specific biometric access control system, a specialized biometric test procedure shall be developed. It shall be identical to the general procedure with the exception that any additional information (for example sliding the cover to allow placement of a finger to a sensor) needed in the real-world operation of a particular biometric access control system shall be identified.

5.2 Relationship of biometric system / subsystem to access control system

A biometric access control system is an access control system that contains a biometric system as a subsystem. This biometric system can be, for instance:

— a verification or identification system with centralized biometric template storage;

— a verification system with decentralized biometric template storage in the biometric subsystem; or

— a verification system with localized biometric template storage (e.g. on an ID card).

NOTE 1 The evaluation of identification performance metrics is outside the scope of this part of ISO/IEC 19795.

Figure 1 illustrates the components and information flows in a generic access control system that includes a biometric system. Following Figure 1 is a key to the circled letters representing information flows. Real deployed systems may vary from this general model.

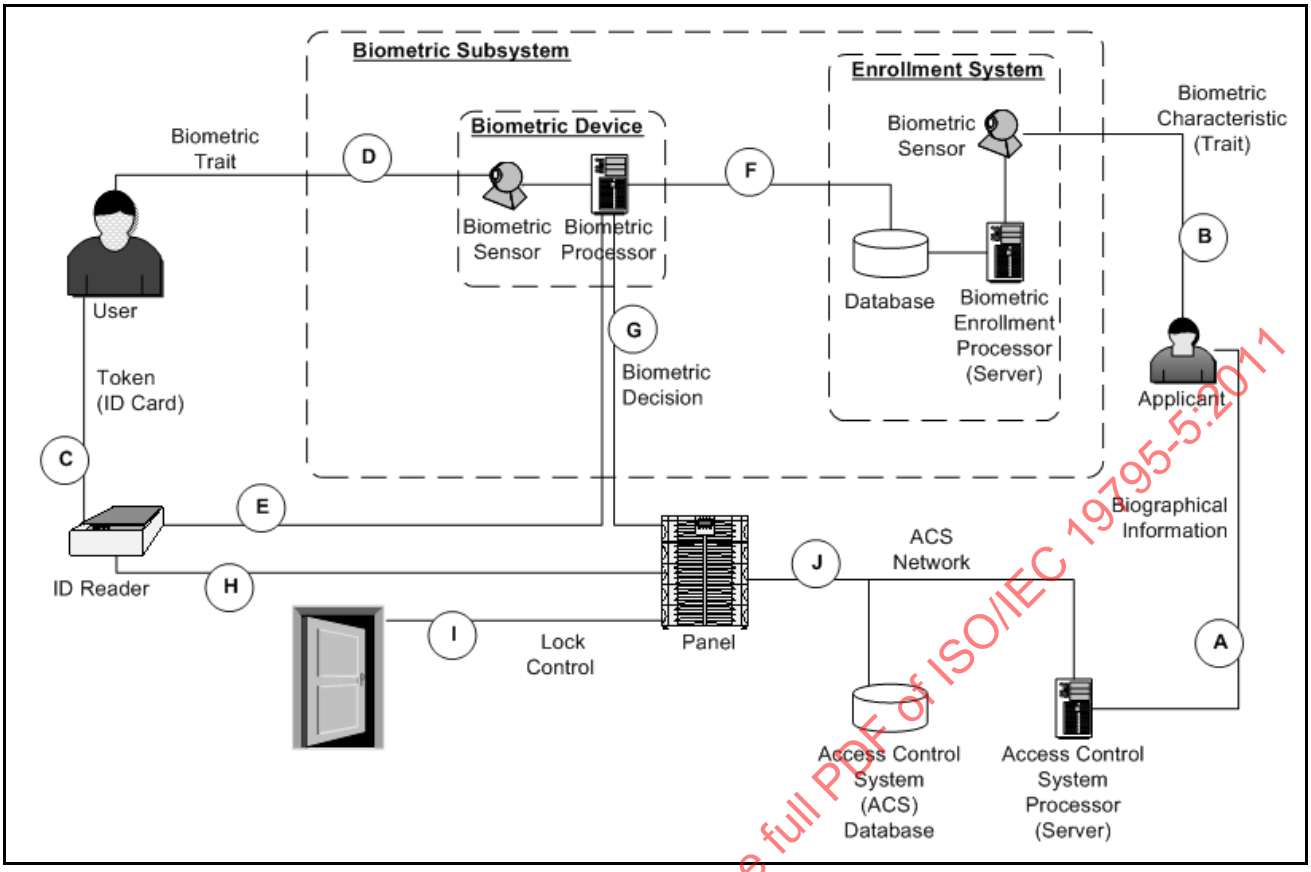

Figure 1 — Generic Biometric Access Control System

A description of the information flow in Figure 1 is as follows:

A. Biographical information: applicant-supplied information (name, address, etc.) obtained during Access Control System (ACS) enrolment via the ACS Processor. This flow is part of a typical legacy ACS.

B. Biometric characteristic (trait): the body part or human behaviour presented by the applicant to the biometric sensor during enrolment (e.g. fingerprint, iris, voice, signature). This flow may also include any interactions between applicant and sensor such as indicator lights or audio feedback.

NOTE 2 An applicant becomes a user only after the enrolment process is completed and access privileges are granted by the access control authority.

C. Token (ID card): any form of machine-readable credential presented by the user to the ID reader to claim an identity.

D. Biometric trait: the body part or human behaviour presented by the user to the biometric sensor during an access transaction (e.g., fingerprint, iris, voice, or signature). This flow may also include any interactions between user and sensor such as indicator lights or audio feedback.

E. User identity code: (ID number, card number, ACS ID) read from the token by the ID reader and sent to the biometric processor as the claim of identity. This flow also includes user template data for template on card architectures.

F. Biometric template data: from enrolment database to biometric processor (for implementations using server-stored templates). This flow is architecture-specific, may be per user transaction or periodic pre-loads.

G. Biometric decision: Yes/No indication (electrical signal or message) from biometric processor to panel conveying the result of the user verification transaction.

H. User identity code: (ID number, card number, ACS ID) read from the token by the ID reader and sent to the panel for the ACS to determine access privilege. This flow is part of a typical legacy ACS.

I. Lock control: electrical signal from the panel used to command the door electro-mechanical locking mechanisms. This flow may also include other signals such as door-open indicators, emergency lock override, etc. This flow is part of a typical legacy ACS.

J. ACS network data: (physical) communication channel (Ethernet, RS485, etc.) enabling data interchange between the panel, ACS processor, and ACS database. The ACS network (logically) depends upon site-specific implementation, and includes a user identity code from panel and user access authorization from ACS processor.

5.3 Evaluation metrics overview

The framework is based on the necessary and sufficient metrics for evaluation of a biometric system for use in an access control application. These metrics are:

— (Single-attempt) false reject rates (FRR_1) at specific values of FAR;

— (Transaction-level) false reject rates (FRR) at specific values of FAR;

— failure to enrol rate (FTE);

— verification transaction time.

To serve many (not all) applications, a range of protection levels, expressed as specific values of FAR, shall be used in this framework.

For each metric, the framework establishes a quantitative grading scheme, using numerical grades, ranging from 0 to 3 (or 0 to 6 for FRR), where a higher score shall indicate better performance and a lower score shall indicate poorer performance. In Clause 7, the metrics are fully defined and the quantitative grading values are established.

NOTE 1 Different metrics may have different grading, as it can be seen in the examples shown in 7.1.5.2 or 7.1.5.3.

NOTE 2 In the kind of test under the scope of this part of ISO/IEC 19795, is not always possible to isolate failure to acquire rate (FTA) cases from false non-match rate (FNMR). Therefore, for the purpose of this part of ISO/IEC 19795, FRR and FRR_1 always include FTA. In case FTA can be obtained, evaluators are encouraged to detail FTA results in the evaluation report.

5.4 Evaluation approach

5.4.1 Tests

The testing defined in this part of the standard shall be Scenario testing under controlled, indoor conditions. The test consists of determination of failure to enrol rate (FTE), verification time, and matching error rates at the single-attempt and transaction levels. The test consists of 10 specific graded metrics: transaction level error rates at three different levels of FAR, attempt level error rates at three different levels of FAR, determination of FTE, and verification transaction time at three different levels of FAR.

A system may, based on supplier request, undergo additional optional testing beyond the graded test. Optional testing designed to generate additional metrics may be conducted depending on the method of operation of the specific system. Such optional testing is not defined by this part of ISO/IEC 19795.

5.4.2 Universality of the test

The rationale for using grade levels versus pass/fail relates to the “universality” or variety of user applications of the evaluation results. Each application is expected to have its own set of required metric grades. A pass/fail evaluation of a system against any particular set of metrics could be developed. However, the introduction of a grade level-based evaluation may provide several advantages. First, a standard test can be defined and used for several different applications. More importantly, the results of a single evaluation can be used by all potential users of the system to judge the suitability of the tested system to their specific application. The system supplier could theoretically reduce overall evaluation cost by submitting to one test, which would optimize the test organization’s time and resources. The overall cost for a single graded evaluation may be higher, but could apply to a variety of user applications.

5.4.3 Levels of effort and decision policies

The experimenter shall report enrolment and verification levels of effort and decision policies as follows.

Minimum and maximum number of placements, attempts, and transactions required or permitted to enrol may be somewhat dependant on the enrolment subsystem under test. An enrolment subsystem may allow enrolment after one attempt, or may require multiple presentations, attempts, and transactions. Unless otherwise dictated, the following shall apply:

— three enrolment transactions of up to three attempts each shall be allowed (if unable to enrol on the first or second transaction);

— an enrolment transaction shall be defined by the supplier, consistent with their operational enrolment practices. For modalities with multiple instances (e.g. fingers, irises), the enrolment policy may include attempts with a primary instance (e.g. right index finger), and if that attempt fails, then secondary instances may be used to enrol;

— three attempts shall be allowed for each verification transaction.

Minimum and maximum duration permitted or required to enrol within a given enrolment attempt or transaction may be somewhat dependant on the enrolment subsystem under test. A biometric subsystem may terminate an enrolment attempt or transaction after a fixed duration. This may be due to (1) inability to acquire sufficiently distinctive data or (2) inability to sense any biometric data input. Incident (1) means that a biometric subsystem has acquired and processed data but found it lacking; incident (2) means that the data was not acquired and processed. It is not feasible to allow a biometric subsystem to attempt to acquire data indefinitely; therefore for subsystems that do not time out, a time of 45 seconds shall be established as the default time-out.

5.4.4 Controlled Indoor Environment

In order to allow for comparability of test results and establish one scenario that has common features for testing, some environmental conditions shall be specified. The test environment shall be controlled, representative of an indoor/office environment, and within the specification for conditions for the system under test.

NOTE 1 The environment is a factor that can affect biometric system performance. It is out of the scope of this part of ISO/IEC 19795 to analyse its influence, however, some environmental conditions have to be controlled to reach a common basis for obtaining comparable and repeatable test results.

The following environmental conditions shall be controlled for all tests:

- temperature: 22ºC ± 4ºC;
- relative humidity: 40% to 60%.

The other environmental factors to control (e.g. illumination, noise, vibration, etc.) shall be specified by the test organization taking into account the biometric system under test consistent with ISO/IEC TR 19795-3. The test

organization shall report on the controlled environmental conditions and values. Any non-controlled conditions considered to have a significant influence on the test shall be reported.

EXAMPLE 1 For an audio-prompted iris recognition system, the experimenter should control as necessary, record and report the following:
- temperature;
- relative humidity;
- presence of natural and artificial lighting, direction and intensity;
- level of noise.

EXAMPLE 2 For a speaker recognition system, the experimenter should record the following:
- temperature;
- relative humidity;
- level of noise.

Required environmental conditions shall be reached before tests are conducted and shall be controlled during enrolment and verification processes with suitable devices. Such conditions shall be recorded and reported.

NOTE 2 Regarding noise, it may be unrealistic to test in near silence, and uncomfortable for the staff to work in the presence of continuous, high background noise.

NOTE 3 Good testing practice is to avoid noisy, distractive activity in the vicinity of a test activity, such as could result from multiple devices being tested together in close proximity.

NOTE 4 The recommended best practice is to suspend testing if out-of-limits environmental conditions are present.

5.5 Crew characteristics and management

5.5.1 Crew demographics

5.5.1.1 General

Demographic characteristics that shall be controlled are the crew age and gender. If other demographic controls are instituted, the controlled parameters, values and results shall be reported.

5.5.1.2 Age

The age distribution of the crew used shall adhere to the ranges of values shown in Table 1.

Table 1 — Age distribution

Age				
<18	18-30	31-50	51-70	>70
0%	25-40%	25-40%	25-40%	0%

NOTE This age distribution does not include younger than 18 or older than 70. If a different age distribution is required, this part of ISO/IEC 19795 is not applicable. ISO/IEC 19795-2 provides more general scenario testing guidance.

5.5.1.3 Gender

The gender distribution of the crew used shall adhere to the ranges of values shown in Table 2.

Table 2 — Gender distribution

Gender	
Male	Female
40-60%	40-60%

NOTE 1 If a different gender distribution is required, this part of ISO/IEC 19795 is not applicable. ISO/IEC 19795-2 provides more general scenario testing guidance.

NOTE 2 The testing organization is encouraged to also control the age distribution within each gender.

5.5.2 Crew size

5.5.2.1 Minimum crew size

To provide the most statistically significant relevant results the crew size should be the maximum number that can be accommodated within the project budget and data collection capabilities of the testing organization.

The minimum crew size of 230 crew members shall be required for conducting verification tests under this part of ISO/IEC 19795.

NOTE If each crew member carries out 15 impostor transactions, 230 crew members will carry out a total of 3450 impostor transactions. Due to correlations among different transactions performed by the same crew member, these 3450 impostor transactions will not be completely independent. For the sake of simplifying statistical analysis, however, it is common to assume that they are statistically independent. If not a single false acceptance occurred among 3450 independent impostor transactions, then the FAR is with 95% confidence not higher than 0.001 ("Rule of 3"). If false acceptances occur, then the FAR is above 0.001. For these reasons, the minimum FAR that could be established with 230 crew members is 0.001.

5.5.2.2 Crew size control

The crew size for enrolment should exceed the target verification crew size and should consider the expectation of 10-25% "drop-out", with the number increasing as time between visits increases. Both the number of individuals that participate in verification testing and the dropout rate shall be reported.

NOTE 1 Testing organizations may elect to include a compensation incentive to reduce drop-out level. Caution must be exercised in that underestimating the drop-out rate potentially undermines the entire test if insufficient revisits are performed, so being conservative is warranted.

NOTE 2 Test results may be biased by selective "drop-out", such as crew members generating errors being encouraged to drop out, or not reminded about future appointments. These practices are discouraged, but crew members have the right to withdraw from testing at any time.

The composition of the crew shall be reported in terms of age and gender for both the enrolled crew and the crew that participated in the revisit testing.

5.5.3 Test crew selection

The experimenter shall assemble a crew of human test subjects to carry out the testing. The demographics of the crew shall be controlled in terms of gender and age (see clause 5.5.1). Controlling these factors will allow for more defensible test results across various crew populations. The same individual shall not be enrolled under different identities. Test organization should be aware that some people may come back twice under different identities and should take precautions to prevent that from happening.

5.5.4 Test crew training

Test crew members shall be trained in the use of the biometric system under test during the enrolment process. The specific crew member training materials should be provided by the supplier and should be representative of the training provided to operational users. The training shall be determined to be completed when the crew member has demonstrated their capability to present their biometric sample as instructed. Additional training shall be conducted when a crew member fails to enrol on their first enrolment transaction.

NOTE The time allocated for crew training is in addition to the time needed to capture the enrolment data as described in clause 5.4.3.

5.5.5 Operator - crew member interaction

Experimenters shall determine and report operator-crew member interaction required and permitted as follows:

- whether biometric system test operation is intended to be attended or unattended;
- whether the operator is to provide specific guidance above that provided by the system during enrolment or comparison;
- the amount of information given to the crew member regarding the evaluation;
- the amount of feedback given to the crew member during the evaluation.

5.5.6 Habituation

The test shall approximate to the extent possible usage by highly-habituated end users.

NOTE 1 The target population for this test may be highly-habituated end users, such as employees. The usage of biometric subsystems evaluated in this test may typically be a several-times-a-day occurrence for a large segment of the population.

NOTE 2 The universal nature of the approach to testing used in this part of ISO/IEC 19795 brings about compromise situations, where it is recognised that not all applications have highly-habituated users.

The experimenter shall specify the degree of habituation of the crew, both a priori and as introduced through the course of the testing, as well as the degree to which habituation effects are accounted for in test design.

NOTE 3 See ISO/IEC 19795-2:2007 for additional detail on habituation reporting.

5.6 Privacy

5.6.1 General

The testing shall be conducted to ensure that the crew identity and all biometric data collected are protected from misuse. Steps taken to protect the privacy of test crew members shall comply with national legislation or, in the absence thereof, with the Organisation for Economic Co-operation and Development (OECD) Guidelines on the Protection of Privacy and Trans-border Flows of Personal Data.

5.6.2 Crew identity protection

The test organization shall maintain privacy practices that ensure the protection of crew identity from unauthorized access or disclosure.

5.6.3 Data protection

The test organization shall maintain privacy practices that ensure the protection of biometric data and all forms of personally identifiable information from unauthorized access or disclosure.

5.6.4 Proprietary information

The testing organization shall not disclose any supplier-proprietary information obtained during testing.

6 Testing approach and conduct

6.1 Planning

6.1.1 General

The activities described in this clause shall be performed to ensure that biometric system evaluations are efficient, expedient, equivalent, unbiased and reliable.

6.1.2 Test objectives

Test objectives shall be to quantify and grade the performance of a biometric component of an access control system in the operating range FAR = 0.1% to FAR = 1% in a controlled environment scenario that is not modality-specific or biased.

NOTE For biometric subsystems that output decisions, the full operating range will not be tested, but only the operating point associated with the fixed threshold setting used for testing.

6.1.3 Inputs to and outputs from the test process

The suppliers shall provide their biometric systems, processes and documentation as test input.

For test output, the facility shall report test findings and substantiating data. Additional output shall consist of detailed data summaries.

6.1.4 Concept of operations

The operation should utilize an accredited testing organization/facility to conduct controlled scenario performance testing of biometric systems to provide an unbiased accurate measure of system performance. See ISO/IEC 17025, General requirements for the competence of testing and calibration laboratories, for information on testing organization/facility accreditation.

It is the responsibility of the supplier to assert that its equipment meet all safety requirements of the test organization.

As a part of the pre-test application process, suppliers shall be required to provide system operating manuals and describe how the supplier intends to train test facility administrators and operators.

To operate the supplier's biometric system, training of system administrators by suppliers shall at least address the following:

— overview of the principles and operation of the biometric system;

— protection of sensitive personal data;

— system installation procedures;

— operator skills required for successful system operations;

— system start-up procedures, normal operating procedures, human interface procedures, shutdown procedures, reboot procedures (e.g. in case of power failure etc.);

— procedures for providing raw similarity scores (or decisions) to the test facility network;

— system error code and exception response activities;

— system tear-down procedures, including permanent deletion of all sensitive personal data.

It is the supplier's responsibility to ensure that adequate training shall be provided to test facility staff prior to actual testing.

When the above actions have been completed, a test schedule should be developed accommodating the availability of the supplier and facility resources. The schedule milestones shall include:

— biometric system delivery to the test organization;

— biometric system installation and integration with the facility test and evaluation network;

— biometric system training of operators;

— a preliminary period of biometric system operation and debugging;

— certification by the supplier that its biometric system is ready for test;

— reach the environmental conditions specified for the test scenario;

— commencement of testing with the biometric enrolment of the crew and the collection of enrolment data, time trials, and statistics;

— continuation of tests including biometric verification testing of genuine and impostor crew member submissions;

— analysis of the data, the construction of the DET curves, and assignment of grades;

— validation of test results and out-briefing with the supplier;

— tear-down and removal of supplier's biometric systems.

6.1.5 Adherence to native system operations

The experimenter shall instruct the biometric system supplier to implement their match thresholds in a fashion consistent with that of operational systems. (Also, see Clause 6.3.3). The experimenter shall instruct the biometric system supplier to implement user interface features (i.e. feedback such as lights, aural cues, and/or visual cues) during enrolment and verification attempts and transactions in a fashion consistent with that of operational systems.

6.2 General test approach

6.2.1 General

This clause defines the general approach and standard practices for the conduct of all biometric system testing performed in conformance with this part of ISO/IEC 19795. Specific areas include: pre-test activities, system operability verification, data collection, problem reporting, and post test activities.

Supplier should perform a 72 hour continuous burn-in period prior to delivery to the testing site. This shall be reported.

6.2.2 Pre-test activities

6.2.2.1 General

Pre-test activities should include inspection, a configuration audit, the test readiness review with supplier participation, and a pre-test briefing which should be conducted prior to the start of each test period with participation from the test personnel only.

6.2.2.2 Inspection

6.2.2.2.1 General

Prior to testing, an evaluation program elements check list should be completed to ensure proper installation and normal operation of all equipment being evaluated, test equipment and equipment interfaces. The purpose of this activity is to ensure the integrity of the test configuration.

It is anticipated that each supplier shall support the installation of its system in the test organization's facility, in close cooperation with the test organization. The supplier shall be required to certify that the installation is functional, shall provide specific configuration and settings data at the completion of setup, and any instructions to the test personnel shall be written. Supplier's personnel should not be present during the actual testing period (unless contacted by the experimenter).

6.2.2.2.2 Physical layout of test environment

The experimenter shall record the physical layout of the test environment, including but not limited to the following:

— dimensional area dedicated to scenario test execution;

— controls and measurement devices related to temperature, relative humidity and other relevant environmental factors that have been specified by test facility for the biometric system under test;

— positioning of biometric acquisition systems;

— relative location of each biometric system in the test environment.

Continuous monitoring of these conditions through the test period may be necessary.

6.2.2.2.3 Specifications

The experimenter shall record the following elements of the biometric system:

— acquisition subsystem: supplier, model, version, and firmware as applicable;

— if integrated within a third-party subsystem, acquisition subsystem: supplier, model, version, and firmware of the acquisition components;

— biometric algorithms: version, revision;

— if the scenario test incorporates a biometric software application, such as a demonstration application or logical access interface then record: provider, title, version, and build of the software application;

— systems tested on or through personal computers, personal digital assistants, or other computing subsystems: processing power, memory, supplier, and model of computing subsystem.

6.2.2.2.4 Architecture

The experimenter shall record the following elements for the biometric system:

— biometric data acquisition, processing, and storage architecture;

— data flow between biometric system and test organization's facility equipment components;

— test management application: design and functions of any application into which the test system is integrated for test management;

— data analysis application: design and functions of any application used to analyze performance results;

— schematics: acquisition subsystems, workstations, server components, layout of test components.

6.2.2.2.5 Implementation

The experimenter shall record system implementation information corresponding to each of the following:

— method of biometric and platform system acquisition;

— level of supplier's involvement in system implementation.

6.2.2.3 Configuration audit

Following successful system installation and checkout, but prior to the test readiness review, there should be a system configuration audit. The audit should be performed after an operational system configuration has been established by the supplier's representative. This configuration shall represent the system baseline at the start of testing.

The configuration audit should be performed by the appropriate test personnel with the assistance of a supplier's representative. The primary goal of the audit is to identify and record system hardware, firmware, and software configuration accurately. The configuration log shall be utilized to document all applicable hardware and software version and serial numbers. This log shall be created and maintained by the appropriate test personnel throughout the evaluation. Data collected during the configuration audit shall be recorded on a configuration log input form.

At the conclusion of the configuration audit, the appropriate test personnel shall place quality assurance seals at the appropriate locations on the biometric system to maintain configuration control throughout the test. In addition, the supplier's representative shall sign a system readiness form documenting approval of system configuration and preparedness for starting testing.

6.2.2.4 Test readiness review

A test readiness review should be held after completion of biometric system installation and checkout and test personnel training. This review shall be conducted by the administrator and should be attended by all personnel involved with test conduct. The purpose of the test readiness review is to verify system and test personnel readiness for conducting testing. During the test readiness review, the administrator should:

— verify (by inspection) that the biometric system meets all safety requirements;

— verify completion of configuration audit;

— review established system and test equipment configuration;

— verify completion of test personnel training;

— identify and review any changes to the biometric system test plan or procedure;

— review biometric system test schedule;

— review test crew characteristics.

Minutes of the test readiness review should be recorded by a designated member of the test personnel and validated by the appropriate test personnel. The minutes should provide enough information to completely and accurately document the review.

6.2.2.5 Pre-test briefing

A pre-test briefing should be conducted by the administrator prior to each test period and evaluation test category. The pre-test briefing should be attended by all test personnel. During the pre-test briefing the administrator should:

— identify test personnel and assign specific responsibilities;

— review system configuration;

— review the specific biometric system test procedure(s);

— review the results of relevant tests;

— identify and review any existing or expected problems;

— provides all necessary test documentation and associated forms;

— review test equipment configuration.

Minutes of the pre-test briefing shall be recorded by a designated member of the test personnel with validation of the accuracy of the recorded information by the person conducting the meeting – i.e. review of the draft minutes.

6.2.2.6 Configuration management

System configuration shall be strictly controlled throughout biometric system testing. To aid in this control, quality assurance seals shall be placed in the appropriate locations on each biometric system during the configuration audit. Should there be a need to make any changes to the system configuration, or access a sealed area due to equipment failure or for maintenance, the administrator and the appropriate test personnel shall be notified and the appropriate information shall be recorded in the configuration log. The information recorded in the configuration log shall indicate the reason for accessing the area, the detail of any configuration changes made, the responsible individual, the date and time. All entries into the configuration log shall be approved by the administrator. If changes are made to the system configuration, the administrator reserves the right to perform regression testing to verify that system performance has not been affected.

6.2.3 System operability verification

System operability verification shall be conducted periodically during testing. These tasks shall be performed by the test personnel according to supplier's procedures at the start of every test period and at supplier's recommended time intervals thereafter.

6.2.4 Data collection

Complete, accurate, and reliable collection of certain data is an integral part of the biometric system test. To facilitate this, the test data collection shall be automated to the maximum extent possible. For any required manual data recording, the administrator shall assign test personnel with specific data collection responsibilities prior to execution of each test category. Data collection assignments should be made at the pre-test briefing. Designated members shall be provided with the necessary data collection forms and/or

equipment, and shall be responsible for recording data as required in the general biometric system test procedure.

During testing, both quantitative and qualitative data shall be collected and recorded in a manner that is specific to the test organization infrastructure and that achieves the biometric system test requirements and objectives.

The quantitative enrolment error rate shall be recorded manually and data shall be entered into the test-organization-information system.

For verification with systems that output similarity scores, the quantitative capture of the system native similarity score should be automated and used to determine, off-line, associated FAR and FRR. The system native similarity scores shall be used to generate full-range DET curves which will be used to compute single-attempt and transaction-level error rates. In addition, each verification attempt time for all crew members shall be recorded. For systems that output only a match/no match decision, the decision should be recorded automatically.

The automated data collection system and the test personnel shall collect data. The data shall include, at a minimum:

— test description;

— test date(s) and time(s);

— crew member ID;

— system indications (if any);

— timeline data for transaction time analysis;

— system operational or failure data (if any);

— comparison results (similarity scores or match/no match decision output).

NOTE Storage of biometric probe samples is not mandatory for the tests specified in this part of ISO/IEC 19795.

To facilitate the correlation of biometric system response to the test crew member, a bar code reader or other appropriate electronic means may be used to read each test crew member's identifier before they interact with the biometric system.

At the conclusion of each test session all data collection forms and additional media shall be validated and signed by the appropriate test personnel. All data shall be treated as being sensitive and stored in accordance with the requirements of Clause 5.6.

6.2.5 Problem reporting and tracking

Any test anomaly or equipment problem that occurs during biometric system testing shall be documented. Specific details of the event shall be recorded by the administrator and validated by the appropriate test personnel immediately following the occurrence. The information should be recorded on a test observation log input form and should include:

— exact details of the event;

— equipment serial numbers;

— biometric system hardware and software version numbers;

— time to repair equipment;

— name of the individual who observed the event;

— any known effects on test outcome;

— times and dates of all significant maintenance and repair actions.

A further review of all logged anomalies should determine if a problem report needs to be written. A problem report shall be written if the event involves a functional problem with the biometric system or could affect the results of a test. Each problem shall be assigned a unique number and shall be tracked throughout testing. Priority levels shall be used to document the severity of the problem. These priorities are identified as:

— Priority I – Performance-Critical - Affects the performance of a critical operational function of the biometric system or results in a severe degradation of performance;

— Priority II – Test-Critical - Does not affect the performance of a critical function of the biometric system, but has an unsatisfactory effect on the test;

— Priority III – Non-Critical - Involves a non-critical operational function or a non-operational system function.

Testing should be terminated if the system fails and cannot be repaired by the supplier within two working days, unless additional time is approved by the experimenter. At the conclusion of each test all problem reports shall be compiled by the test personnel and included in the final test report.

6.2.6 Post-test briefing

There should be an internal post-test briefing at the conclusion of each test period and test category. All test personnel should attend the post-test briefing. The administrator should present a summary of the test activities and relevant issues. During the post test briefing the administrator should review, as necessary:

— deviations from the planned test procedure(s);

— test anomalies;

— configuration management issues;

— equipment failures;

— upcoming test schedule.

Minutes of the post-test briefing should be recorded by a designated member of the test personnel and validated. The minutes should provide enough information to completely and accurately document the meeting.

6.3 Testing methodology

6.3.1 Introduction

For the purposes of this part of ISO/IEC 19795, a controlled environment scenario test configuration shall be utilized, reducing the number of variables and resulting in improved test repeatability and increased reliability for product performance comparisons.

Testing shall generate enrolment, biometric verification accuracy, and transaction time results for participating crew members.

NOTE Crew members are the participants in the testing whose behaviour is monitored, controlled and trusted to be in accordance with the planned test.

The biometric subsystem should be capable of providing up to three consecutive verification attempts (if needed) without having the crew member make additional claims of identity. The subsystem shall also provide a clear indication of success or failure for each attempt so that crew members can proceed with additional attempts (if required) with a minimum delay.

6.3.2 Enrolment transactions and results generation

Prior to enrolment, the crew member's identity and the identifier (e.g. crew member ID number) used to relate the crew member to his biometric data shall be established. The crew member shall be enrolled into the biometric system and his biometric template(s) may be made capable of being stored externally to the biometric system. The success of the crew member's enrolment shall be based on the supplier's established enrolment process. If the enrolment process fails, the crew member shall be considered to have failed this enrolment process. (See clause 5.4.3 for enrolment policy details.)

NOTE For biometric modalities with multiple instances (e.g. left and right iris, multiple fingers), the supplier's established or typical enrolment process should be employed and documented. In these cases, a failure to enrol is declared based on the supplier's recommended practice, which could result in declaring a failure only if none of the instances are able to be enrolled.

6.3.3 Verification attempts, transactions, and results generation

The crew member shall conduct biometric verification transactions on the day of enrolment and a minimum of 1 week after enrolment. Biometric verification transactions that occur after the day of enrolment are referred to as revisit tests as described in clause 6.3.4.3.2.

The crew member shall conduct 5 genuine biometric verification transactions and up to 15 zero-effort impostor biometric verification transactions during the revisit. Impostor transaction generation and the number of impostor transactions are discussed in clause 6.3.5. The operator shall inform the crew member as to whether an impostor or genuine transaction is being conducted.

Each biometric transaction shall be composed of one or more attempts. Therefore, up to 3 (the specified maximum amount of attempts) shall be conducted (never more than that). As soon as a successful result is given, no further attempts are performed within that transaction. A decision threshold that is sufficiently demanding shall be chosen (to achieve the strictest FAR of interest to the supplier) and shall not be changed during the whole test campaign.

NOTE 1 If the chosen decision threshold is too lenient (i.e. results in a transaction-level FAR of more than 0.001), then it will not be possible to find out what would have been the transaction-level FRR and the transaction time at a more demanding threshold (e.g. at an FAR of 0.001), as within each transaction no further attempts are performed as soon as the decision threshold is met.

NOTE 2 For biometric modalities with multiple instances enrolled, the sequence of presentations will be defined by the supplier in accordance with normal operational practices.

Each verification transaction results in a decision of match or non-match, which shall be recorded. For each transaction, it shall also be recorded (if possible) how many attempts have been conducted until a decision of match has been reached.

NOTE 3 A transaction-level decision of non-match is always the result of three non-matching attempts.

NOTE 4 The number of allowed attempts is prescribed to be 3 for the purpose of obtaining comparable test results. In an operational setting, a different number of allowed attempts may be chosen depending on the security and usability requirements on the biometric system.

For purposes of determining FAR, FRR, and transaction time values, if any verification attempt generates a similarity score greater than the threshold (or a match decision), then the transaction is considered to be complete.

If the first verification attempt fails to generate a similarity score greater than the threshold or fails to acquire, then the crew member shall conduct the second verification attempt. If the second attempt fails to generate a

similarity score greater than the threshold or fails to acquire, then the crew member shall conduct the third verification attempt. If the third attempt fails to generate a similarity score greater than the threshold or fails to acquire, then the transaction is complete.

For testing based on similarity scores, results from each biometric verification attempt shall be stored such that the experimenter can determine which, if any, of the three biometric verification attempts was the first to generate a similarity score. If each of the three biometric verification attempts fails to generate a similarity score, the biometric verification transaction shall be declared a transaction level failure to acquire.

Any FTA case shall be considered as a FRR case, and if the detection of whether it has been a FTA or a FNMR is available, the rate for FTA should be reported.

NOTE 5 For testing based on matching decision output it may be impossible to differentiate between a failure to acquire and a non-match decision.

Similarity scores as available from all of the verification attempts shall be stored such that it can be determined which, if any, similarity score was the first to exceed the match threshold for a given level of FAR. Transaction-level FAR and FRR shall be based on the first comparison score that meets the decision threshold for a given FAR. Such calculations are based on revisit tests for both FRR and FAR, and takes place once all crew members' revisit tests are completed. (See Clause 7.1 for further details on data analysis and grading.)

EXAMPLE 1 If any verification attempt generates a similarity score that passes the criteria of the threshold level, then the transaction is complete after the first attempt. This value is saved for "Attempt 1-Transaction N". No values are saved for "Attempt 2-Transaction N" or "Attempt 3-Transaction N", as these attempts were never executed.

EXAMPLE 2 The crew member fails to generate a similarity score that passes the criteria of the threshold level on his first attempt because no sample is acquired. On his second attempt, the crew member generates a similarity score above the threshold. An "FTA" value is saved for "Attempt 1-Transaction N". The similarity score that exceeded the threshold is saved for "Attempt 2-Transaction N". No value is saved for "Attempt 3-Transaction N" as this attempt is never executed.

EXAMPLE 3 The crew member fails to generate a similarity score that passes the criteria of the threshold level on his first attempt because the similarity score failed to exceed the threshold. On his second attempt, the crew member generates a similarity score above the threshold. The similarity scores for "Attempt 1-Transaction N" and "Attempt 2-Transaction N" are saved. No value is saved for "Attempt 3-Transaction N" as this attempt is never executed.

EXAMPLE 4 The crew member fails to generate a similarity score that passes the criteria of the threshold level on each of his three attempts because each similarity score fails to exceed the threshold. The similarity scores for "Attempt 1-Transaction N", "Attempt 2-Transaction N", and "Attempt 3-Transaction N" are saved.

Biometric verification transaction time shall be calculated as the duration from the point at which the first attempt is initiated to the reporting of the result from the first attempt that generates a similarity score greater than the threshold (see 7.1.3.5) or a match decision.

The biometric system may generate a single similarity score for each biometric verification attempt, or may generate a series of similarity scores during the course of a biometric verification attempt. In the latter case, the strongest similarity score from the series shall be retained as the one similarity score associated with this biometric verification attempt.

6.3.4 Enrolment and verification temporal separation

6.3.4.1 General

Ideally, the performance of a biometric system would be evaluated with a significant time period between enrolment and verification attempts. However, in light of the cost, complexity and timeliness of completion of these evaluations, a compromise must be achieved.

6.3.4.2 Enrolment-verification

The successful enrolment of a user shall follow the supplier prescribed process, and typically includes an immediate verification that the candidate enrolment template can be successfully matched. This may take place prior to template storage. The time separation between candidate enrolment template generation and the "enrolment-verification" shall be as little as possible, or based on the supplier's directions. In essence it is immediate, but with disengagement from the device. These verifications are considered part of the enrolment process and shall not be included as part of the verification testing.

6.3.4.3 Verification testing

To generate data for calculation of false reject rate, verification testing must be performed. The temporal separation between enrolment and verification can be categorized as "same-day" or "revisit".

6.3.4.3.1 Same-day verification testing

For the test, each crew member shall conduct 5 same-day verification transactions.

NOTE 1 Although same day verification results are not used for obtaining FAR and FRR figures, this verification is very important for the methodology defined in this part of ISO/IEC 19795, because it will improve the crew training, and also it can be used for detecting if some users need re-enrolment.

The minimum time between the enrolment and same-day verification shall allow any convenient timing based on testing operations as long as there is effective disengagement from the device.

NOTE 2 To force such effective disengagement, the laboratory may consider one of the following strategies:

— making the user wait for 5 minutes between his/her enrolment and the first verification;

— if more than one system is evaluated at the same time, make the user change to the following system, until he/she completes a whole round among all systems;

— if a single system is being evaluated, and more than one crew member is in the evaluation room, rotate among crew members.

NOTE 3 The easiest verification FRR data to obtain is that which is taken on the same day as the enrolment, as the crew members are physically available.

Following the initial time separation following enrolment, multiple same-day verification transactions shall be carried out with brief disengagement between transactions. Similar approaches as the ones given for the lapse between enrolment and first same-day verification may be considered.

Disengagement between attempts within each transaction shall be done following the supplier's policy.

6.3.4.3.2 Revisit verification testing

One single revisit session shall be allowed for each crew member. The test revisit verification minimum time separation from enrolment shall be no less than 1 week and no more than three months. At least 75% of revisit times shall occur in the 4 to 8 week interval, as measured from the first visit date. During the revisit, each crew member shall perform 5 verification transactions.

NOTE Revisit time separation may be a function of the ability of the test facility to reschedule visits.

Multiple revisit verification transactions shall be carried out with brief disengagement between transactions. Similar approaches as the ones given for the lapse between enrolment and first same-day verification should be considered.

Disengagement between attempts within each transaction shall be done following the supplier's policy.

6.3.5 Impostor tests

6.3.5.1 General

To evaluate false accept rate, testing shall be performed where the ground truth is known, and the identity claim for verification is intentionally a false claim.

6.3.5.2 Crew composition

The same test crew as used for genuine verification testing shall be used for impostor testing.

6.3.5.3 Method of analysis

All false accept rate evaluation data shall be collected on-line, randomly assigning false claims of identity from the remainder of enrolled test crew (and without replacement).

6.3.5.4 Number of impostor transactions

During the revisit, the number of impostor transactions shall be 15 per individual for a test crew size of 230 test subjects.

NOTE Impostor transactions are made only at revisit, which simplifies random selection, as the test crew is fully enrolled at that stage.

However, if the test crew size is higher, it is possible to achieve the same confidence level reducing the number of impostor transactions per crew member.

When the test crew size is over 230 test subjects, the number of impostor transactions to be performed per test crew member may be reduced providing that a total number of at least 3,450 impostor transactions are executed. The test organization should (to the extent possible) maintain a constant number of imposter transactions per crew member.

6.4 Errors and exception cases

For a given interaction of an individual crew member with the system under test, errors or exception cases may occur including the following:

— biometric system provides the operator with a functional error code. The operator should react to the error code as mandated by the supplier's manuals. The time associated with reacting to the error code should be included as part of the crew interaction time interval. If the system error is not corrected within a reasonable time, then the specific system with the error is removed from service and as appropriate, the entire enrolment transaction or subsequent verification transaction is restarted from the beginning with a substitute system of the same production make and model.

NOTE Since this test involves only one unit (from a larger production population), it is assumes that each unit will perform similarly (toward identically). Given that underlying assumption, it would be acceptable to replace a device at any time without expectation of creating any inconsistency. It is clearly outside the scope of this part of ISO/IEC 19795 to prove that all units in a production lot perform equally.

— during enrolment, an individual crew member is not successful in obtaining an acceptable enrolment result. If enrolment is not possible, this result is recorded and the specific individual crew member is removed for revisit verification test purposes, and the crew subsequently adjusted so as to maintain the validity required of test statistics.

6.5 Incremental performance evaluations

In some instances (for example when a change is made only to the comparison algorithm, not to the acquisition hardware), it might be appropriate to conduct an incremental performance evaluation of a biometric system, rather than a full re-evaluation. The conditions under which an incremental evaluation is permitted shall be as described in this part of ISO/IEC 19795. In general, an incremental evaluation will be permitted only where the acquisition portion of the biometric system has not changed whatsoever, and the biometric images or templates (as applicable) were saved during the initial evaluation. In this case an off-line evaluation may be conducted by cycling the stored images or templates through the modified comparison algorithm running on the processor within the biometric system. If incremental performance evaluation has been carried out, this shall be reported and documented.

7 Grading and reporting

7.1 Grading

7.1.1 Data analysis

The test personnel shall perform all data analysis upon completion of testing. The same data analysis methods used are performed for each biometric system tested to provide consistent, unbiased results. Specific biometric system performance parameters that shall be calculated include, but are not limited to, FTE, FRR and FAR for single attempts, FRR and FAR for transactions (multiple attempts), and transaction time.

NOTE Single attempt FRR is not the same as FNMR because it also includes any failures to acquire, which in this type of testing cannot usually be distinguished.

7.1.2 Using statistical analysis methods

Due to the variability inherent in human subject based testing, it is necessary to use statistical analysis based on confidence intervals and the variability in the measured test data to arrive at a defendable grade determination. Otherwise, it would be possible for a system to be tested, and then re-tested, and the two grading decision results may be different. To mitigate this undesirable situation, the measured performance will need to be “significantly” (in the statistical sense) better than the required grading values.

The statistical methods that shall be used to establish grades for each of the graded metrics are described in detail in Annex B. This analysis shall use a 90% confidence level that true system performance (not just this limited test measured performance) is below the specific grading levels. The specific method for computing the confidence interval to be applied for each metric is:

— for FRR (at each FAR), use the Correlated Binary method (Annex B.1);

— for FAR, use the Correlated Binary method (Annex B.1);

— for FTE, use the Beta Distribution method (Annex B.2);

— for transaction time, use the Z-statistic (Annex B.3).

7.1.3 Performance measures

7.1.3.1 General

The test consists of 10 specific graded metrics: transaction level error rates at three different levels of FAR, attempt level error rates at three different levels of FAR, determination of FTE, and verification transaction time at three different levels of FAR. (See clause 7.1.3.5.)

7.1.3.2 FAR levels

To serve many (not all) applications, a range of protection levels, expressed as specific values of FAR, shall be used in this framework. These 3 levels are 0.1%, 0.3% and 1.0% FAR.

7.1.3.3 Comparison error rates

Grades for FRR shall be assigned based on the values in Table 3. FRR values used in this context represent the statistical upper bound of the measured FRR. (See 7.1.2 and Annex B.)

Table 3 — False reject error rate grading

Grade	False Reject Rate (FRR), %
6	≤ 0.33%
5	≤ 1.00%
4	≤ 2.00%
3	≤ 3.33%
2	≤ 5.00%
1	≤ 7.00%
0	≤ 100.00%

The framework provides for evaluation of error rates at two access policy levels: single-attempt and multiple-attempt (also referred to as "transaction-level"). The multiple-attempt policy allows up to three attempts per transaction for identity confirmation. Both single-attempt and transaction-level results are graded using the same quantitative grading methods.

7.1.3.4 Failure to Enrol Rate

The framework includes the measurement and grading of the biometric enrolment subsystem performance in terms of its ability to enrol crew members. The grading shall utilize specific FTE values as shown In Table 4. The scale for grading the system performance in FTE is expressed in the form of percentage. The framework includes the measurement of the failure to enrol rate of the biometric subsystem as well as the statistical analysis of the confidence in that measurement. FTE shall be graded as in Table 4, using the upper bound of the measured FTE rate.

Table 4 — Failure to enrol grading

Grade	Failure to enrol rate (FTE), %
3	≤ 1.0%
2	≤ 3.5%
1	≤ 7.7%
0	≤ 100.0%

NOTE To obtain comparable results among tests, certain constraints are applied to the enrolment process, as stated in clause 5.4.3.

7.1.3.5 Transaction time

The metric for the time required for the biometric system portion of an access control transaction shall be measured in terms of verification (1:1) transaction time.

The objective of biometric transaction time testing is to measure, under simulated operating conditions, the verification time of the biometric subsystem, exclusive of subsequent actions of the access control system. Biometric transaction time includes only the biometric system-related times. Not included in the timing is any access control processing or mechanical/electrical activity associated with granting access.

Timing begins when the biometric capture subject commences interaction with the biometric capture device, i.e. the earlier of starting to position themselves to present their biometric characteristic, or making their claim of identity (such as PIN entry or reading an ID badge). Timing ends when the biometric system renders a final transaction decision (i.e. a successful verification, or the denial after the allowable number of attempts). The test report shall indicate the procedure used for biometric transaction timing.

To accurately measure biometric transaction time, the test personnel shall analyze the full set of up to 3 attempts for genuine verification transactions in accordance with biometric subsystem operating procedures.

NOTE 1 The timing of the full transaction is measured. If more than one attempt is needed (up to 3 attempts) for ending a transaction, the time for each of the attempts, plus the time needed by the user to interact with the system, will be measured as a single transaction time.

Biometric transaction time metrics shall then be computed offline for each FAR level based on revisit genuine transactions. Only transactions that result in a successful verification will be included in this computation.

NOTE 2 The offline computation requires analysis of each attempt in transactions of more than one attempt, to determine which attempt, given the FAR level being computed and its associated threshold value first achieves a successful match. This may lead to reducing the time for that transaction for higher levels of FAR.

The transaction time value, using data from all subjects and the statistical analysis defined in clause 7.1.2, shall be used to determine the biometric subsystem's grade level as specified in Table 5.

NOTE 3 Depending on the system, it may happen that for FAR level of 1.0%, the timing measured will be really close to the one of a single attempt configuration.

Table 5 — Mean genuine transaction time grading

Grade	Mean Successful Transaction Time (sec)
4	≤ 3
3	≤ 4
2	≤ 5
1	≤ 6
0	≤ ∞

Average impostor transaction times shall be calculated and reported, but not graded.

As noted in ISO/IEC 19795-1, clause 6.3.K, verification timings of error rates are sometimes dependent on database size. If this is the case, a database of 1000 enrolments shall be used in the test. The device supplier shall provide a database of 1000 enrolments as part of the test article.

7.1.3.6 Test graded performance metrics

For each subsystem test, both attempt-level and transaction-level FRR shall be computed and reported. These error rates correspond to the test metrics numbered 1 to 6 in Table 6. Metrics 1 through 3 show transaction-level FRR. Associated with these 3 metrics are the threshold values (in the units native to the subsystem under test) corresponding to the FAR-FRR point on the DET curve. Metrics 4 through 6 show attempt-level FRR. The threshold value_1 parameters are the associated threshold values for the 3 points on the single attempt DET curve. Results derived solely from revisit verification attempts and transactions shall be reported and graded.

Metric 7, FTE, address the percentage of the population of crew members who were unable to generate a usable template for the biometric subsystem under test.

Metrics 8 through 10, transaction time addresses the time interval from the initiation of a first verification attempt to the availability of the corresponding transaction result at each of the 3 FAR levels.

NOTE For decision based systems the transaction time will be computed directly from measured transaction time data.

Table 6 — Test metrics

Target FAR	0.1%	0.3%	1.0%
FRR	Metric #1	Metric #2	Metric #3
Threshold value			
FRR_1	Metric #4	Metric #5	Metric #6
Threshold value_1			
FTE	Metric #7		
Transaction time	Metric #8	Metric #9	Metric #10

7.1.4 Grading of matching performance illustration

For the graded matching performance test, device similarity scores shall be plotted if possible. The associated subsystem error shall be computed and the result shall be used to construct the quantitative error rates plotted as a Detection Error Trade-off (DET) curve. The DET portrays the relationship between the device's FRR and its FAR. In addition, a transaction is defined as the aggregate "logical OR" of up to three verification attempts by an individual crew member. Results derived from both verification attempts and transactions shall be plotted (i.e. 2 DET's), reported and graded. (See A.5).

When plotting DET curves, all measured points shall be shown on the curve. When plotting DET curves, the graph shall only be shown within the range of the actual data points.

For biometric subsystems that output match/no match decisions (rather than similarity scores), the result will be an operating point on a DET graph rather than a curve. Grades shall be assigned by extrapolating the operating point horizontally to the right on the DET graph, to the first FAR level encountered. No grades can be assigned for the FAR levels below the measured FAR at the operating point.

7.1.5 Uses (of grading)

7.1.5.1 General

This framework is suitable to use for two different purposes or perspectives: user requirements, and evaluation and reporting.

7.1.5.2 User requirements

The system owner or system requirements-defining organization can use this framework to set quantitative performance requirements for a biometric system based on the needs of their particular application. The example below depicting FAR=0.10% criteria could represent a high-security facility requirement. That application requires low false reject rates at very low false accept rates. In this particular application, the FTE rate and verification transaction time metrics are not specified at the higher grade levels, because a higher rate of failure to enrol or a slower transaction time is not critical to the operation. Other applications may have a significantly different selection of required grades.

EXAMPLE Table 7 provides sample requirements for high-security applications.

Table 7 — Sample Test Performance Requirements at FAR =0.10%

Metric Requirement	Grade Level
Single-attempt false rejection rate of <2%	4
Transaction-level false rejection rate of <1%	5
Failure to enrol < 3.5%	2
Verification transaction time < 6 seconds	1

7.1.5.3 Evaluation and reporting

A summary of test results as depicted in Table 8 below may be generated for a system evaluated using the grading methods in this part of ISO/IEC 19795 (based on Table 3, 4 and 5).

Table 8 — Sample summary of test results for all grades

Target FAR	0.1%	0.3%	1.0%
FRR Grade	3	4	5
Threshold value	120	97	90
FRR_1 Grade	2	3	4
Threshold value_1	98	73	64
FTE Grade	3		
Transaction time Grade	1	2	4

Reporting of the threshold values associated with the graded performance points reported should be helpful guidance to system implementer when selecting a threshold level expected to achieve a specific level of FAR.

NOTE See Annex A for additional information on interpretation of grades.

7.1.5.4 Comparing and interpreting grades

Grading results based on this part of ISO/IEC 19795 should be broadly comparable if test conditions are identical. However, even in identical test conditions, it should be noted that a single performance grade difference may not be statistically significant.

Situations leading to reduced comparability across graded tests include:

— differences in test conditions, e.g. different test labs may operate using different environmental controls;

— differences in the guidance provided to the test crew, and differences in their level of habituation;

— systems tested on using match / non-match decisions based on threshold settings set by the biometric supplier, may not be graded as favourably as those returning comparison scores;

— systems with the same observed FRR may receive different grades if tested using different sized test crews (as the standard error reduces as crew size increases, resulting in a smaller value for the upper confidence bound for FRR);

— systems with the same observed FRR may receive different grades if they have different intra-individual correlations (intra-individual correlation is defined in Annex B.1, and as this increases so does the standard error, resulting a greater value for the upper confidence bound for FRR).

7.2 Documentation requirements and control

7.2.1 General

The test shall be executed according to the guidelines established in this part of ISO/IEC 19795. Effective documentation and control of all events shall occur during the test process to ensure compliance with stated objectives. This Clause provides a brief description of the documents that shall be utilized during testing.

7.2.2 Test control

7.2.2.1 General

Extensive control of test activities and documentation of results is required throughout all phases of testing. To accomplish this, test control documentation shall be utilized. This documentation shall include a configuration log, test briefing minutes, test data, problem reports, and a biometric test observation log.

7.2.2.2 Configuration log

The configuration log shall be used to document and control biometric system configuration throughout the evaluation. The log shall contain biometric system configuration data from the configuration audit and document any system baseline changes that occur during testing. Entries into the log shall be performed by the test personnel and validated by the appropriate test personnel.

7.2.2.3 Test briefing minutes

The minutes of the test briefings shall provide enough information to completely and accurately document the pre-test and post-test meetings. The minutes shall be compiled for each biometric system and shall be used as a reference to document the daily test activities. These meeting minutes shall be assembled by the test personnel and shall be included in the test report.

7.2.2.4 Test data

The test data shall be collected throughout testing and shall be retained for subsequent analysis. Proper recording of this data is the responsibility of the administrator and shall be validated by the appropriate test personnel. Any biometric system proprietary information obtained during test shall be used for test and analysis purposes only. The test personnel shall make no disclosure of proprietary information to anyone or to any contractor who has not entered into a nondisclosure agreement with the supplier.

7.2.2.5 Test problem report

All entries into the test observation log shall be reviewed daily to determine if a test problem report needs to be written. In general, a test problem report shall be written if the observation is related to equipment failure or if it may affect test results. The problem reports shall be written by the administrator, assigned a unique number, and shall be closely tracked throughout testing. All problem reports shall be addressed in the test report. This shall include a description of the problem, disposition, and the results of any retests performed, if applicable.

7.2.2.6 Test observation log

A test observation log shall be compiled during testing to document all test anomalies and equipment failures. Specific details shall be recorded on a test observation log input form. All entries shall be recorded by the administrator (or operator) and validated by the appropriate test personnel immediately following the occurrence. The log shall be used as an attachment to the evaluation report.

7.3 Reporting performance results

7.3.1 Reporting requirements

All reporting requirements shall be documented in a test report. The reporting of the process and other documentary aspects of the evaluation shall be summarized and reported in accordance with ISO/IEC 19795-1:2005 and ISO/IEC 19795-2:2007. Environmental factors reporting shall be consistent with ISO/IEC TR 19795-3.

The biometric system test report shall contain a detailed analysis of the test results, an evaluation of the operability of the system, and a summary of any problems encountered during the test. The report shall contain all the information necessary to evaluate the system against the criteria and provide information concerning the operational performance of the system. It shall be prepared by the test personnel and approved by the experimenter.

7.3.2 Report structure

The final test report shall include but not be limited to the following:

- executive summary;
- system description / configuration, including the modality;
- test objectives;
- reference documents;
- test date and location;
- physical layout and environmental factors;
- supplier information;
- test personnel;
- crew size and profile;
- test results summary, including , at least, metric tested (following Table 6), and grade for each of those metrics;
- detailed test results;
- data collection techniques;
- data analysis techniques;
- procedural deviations;
- problems encountered;
- regression test requirements;
- conclusions and recommendations;
- test briefing minutes;
- test observation log.

Annex A
(informative)

Grading information

A.1 Equivalence of tests

The purpose of strictly defining testing methods in the body of this part of ISO/IEC 19795 is to maximize the likelihood that tests of biometric systems will produce grades that can be compared across different testing organizations. However, there are a few situations that could lead to reduced comparability across graded tests. This information is provided to alert users of the testing results of these possibilities.

A.2 Comparison of test results

The performance grading shows statistical confidence that a system meets the performance required of that grade. However this does not mean that systems may be directly compared to each other by the grades given.

The following sources of possible comparability issues have been identified:

— differences in test conditions, e.g. different test labs may operate using different environmental controls

— differences in the guidance provided to the test crew, and differences in their level of habituation

— systems tested on using match / non-match decisions based on threshold settings set by the biometric supplier, may not be graded as favourably as those returning comparison scores

— systems with the same observed FRR may receive different grades if tested using different sized test crews (as the standard error reduces as crew size increases, resulting in a smaller value for the upper confidence bound for FRR)

— systems with the same observed FRR may receive different grades if they have different intra-individual correlations (intra-individual correlation is defined in Annex B.1, and as this increases so does the standard error, resulting a greater value for the upper confidence bound for FRR).

To compare performance of tested systems, it is advised that the full test report is used, allowing comparison between the performance levels actually achieved, as well as the conditions of the test.

A.3 Grading values for enrolment performance

Table 4 grade value ranges represent a variety of access control applications. Some applications may require very complete coverage with few failures to enrol (Grade 3). Others may have a second biometric modality available, or adequate secondary procedures, therefore requiring less complete coverage.

These values are representative of testing results achieved in recent tests of biometric subsystems intended for access control applications. In these tests, conducted with a very similar testing protocol as this part, 7 systems were evaluated.