Standard Test Methods For Response Robots
ASTM International Standards Committee on Homeland Security Applications;
Operational Equipment; Robots (E54.08.01)

Guide for Evaluating, Purchasing, and Training with Response Robots Using DHS-NIST-ASTM International Standard Test Methods

Sponsored by: Philip Mattson, Tod Companion, Matt Hickman

**Office of Standards, Science and Technology Directorate,
U.S. Department of Homeland Security**

Developing and coordinating the adoption of national standards and appropriate evaluation methods to meet homeland security mission needs.

Performed By:

**Intelligent Systems Division, Engineering Laboratory,
National Institute of Standards and Technology**

Promoting the development and dissemination of advanced technologies, guidelines, and services through measurement science research, performance metrics, tools and methodologies for engineering applications.

Test Director:

Adam Jacoff

Intelligent Systems Division
National Institute of Standards and Technology
Gaithersburg, MD 20899

Contributors:

Elena Messina, Hui-Min Huang
Ann Virts, Anthony Downs
Richard Norcross, Raymond Sheh

Email: RobotTestMethods@nist.gov Tel: 301-975-4235
http://www.nist.gov/el/isd/ks/response_robot_test_methods.cfm

Table of Contents

Acknowledgments:
The Department of Homeland Security (DHS) sponsored the development of these standard test methods under an Interagency Agreement with the National Institute of Standards and Technology (NIST). Other organizations have sponsored robot test and evaluation events to capture robot capabilities data within standard and draft standard test methods, including: Joint Improvised Explosive Device Defeat Organization (JIEDDO), Army Research Laboratory (ARL), and NIST's Office of Law Enforcement Standards (OLES).

Disclaimer:
Certain commercial equipment, instruments, or materials are identified in this document. Such identification does not imply recommendation or endorsement by the Department of Homeland Security or the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.

Policy:
The International System of Units (SI) is used throughout this document. Conversions from SI units to U.S. Customary units are made where possible but approximate equivalents are used to specify materials which are readily available in the domestic market or to avoid excessive fabrication costs of test apparatuses while maintaining repeatability and reproducibility of the test method results.

Guide for Evaluating, Purchasing, and Training with Response Robots Using DHS-NIST-ASTM International Standard Test Methods

Background

Emergency responders literally risk life and limb interacting with known hazards to protect the public. They typically wear only conventional personal protective equipment while manually dealing with a variety of extreme hazards for which remotely operated robots should be well suited. Examples include disabling or dismantling improvised explosive devices (pipes, packages, vehicle); searching for survivors in collapsed or compromised structures; investigating illicit border tunnels; establishing situational awareness during police actions; monitoring large scale industrial or transportation accidents; or assessing potential terrorist attacks using chemical, biological, or radiological sources. Responders want to “start remote and stay remote” when dealing with such hazards and need capable robotic systems that can be remotely operated from safe stand-off distances to provide situational awareness, negotiate complex environments, perform dexterous manipulation of objects, and many other tasks necessary to mitigate hazardous situations. Many responder organizations already own robots but have had difficulty deploying them effectively. New robots are promising advanced capabilities and easier operator interfaces, but it is hard for responders to sift through the marketing. Responders need quantitative ways to measure whether any given robot is capable and reliable enough to perform specific missions. They also need ways to train and measure operator proficiency to isolate deficiencies in equipment and/or improve very perishable operator skills.

Since 2001, a series of Presidential Policy Directives on National Preparedness have prompted increased funding for new and better technologies for emergency responders, including purchasing of response robots. The most recent 2011 Directive outlines the need for strengthening the security and resilience of the United States through systematic preparation for threats including acts of terrorism, pandemics, significant accidents, and catastrophic natural disasters. The Directive emphasizes three national preparedness principles: 1) an all-hazards approach, 2) a focus on capabilities, and 3) outcomes with rigorous assessments to measure and track progress in building and sustaining capabilities over time. This project applies all three principles specifically for response robots.

In 2005, the U.S. Department of Homeland Security, Science and Technology Directorate (DHS S&T) engaged in a multi-year partnership with the National Institute of Standards and Technology (NIST) to develop a comprehensive suite of standard test methods to quantify key capabilities of robots for emergency response and other hazardous applications. The resulting suite of ***DHS-NIST-ASTM International Standard Test Methods for Response Robots*** **measures robot maneuvering, mobility, manipulation, sensing, endurance, radio communication, durability, reliability, logistics, and safety** for remotely operated ground vehicles, aquatic vehicles, and small unmanned aerial systems in FAA Group I under 2 kg (4.4 lbs). The objective is to facilitate quantitative comparisons of different robot configurations based on statistically significant robot capabilities data captured within standard test methods to understand deployment capabilities, guide purchasing decisions, and support operator training with measures of proficiency.

This suite of test methods is being standardized through the ASTM International Standards Committee on Homeland Security Applications; Operational Equipment; Robots (E54.08.01) which includes equal representation of robot developers, emergency responders, and civilian/military test administrators. Robot developers benefit by using the standard test methods as tangible representations of operational requirements to understand mission needs, inspire innovations, make trade-off decisions, measure incremental improvements, and highlight break-through capabilities. Responders and soldiers benefit by using robot capabilities data captured within the standard test methods to guide purchasing decisions, support training, and measure operator proficiency. Fifteen standards have been adopted internationally and dozens more are being validated with associated apparatuses, procedures, and performance metrics. This suite of test methods addresses a range of robot sizes and capabilities including throwable robots for reconnaissance tasks, mobile manipulator robots for package size and vehicle-borne improvised explosive devices, and rapidly deployable aerial and aquatic systems.

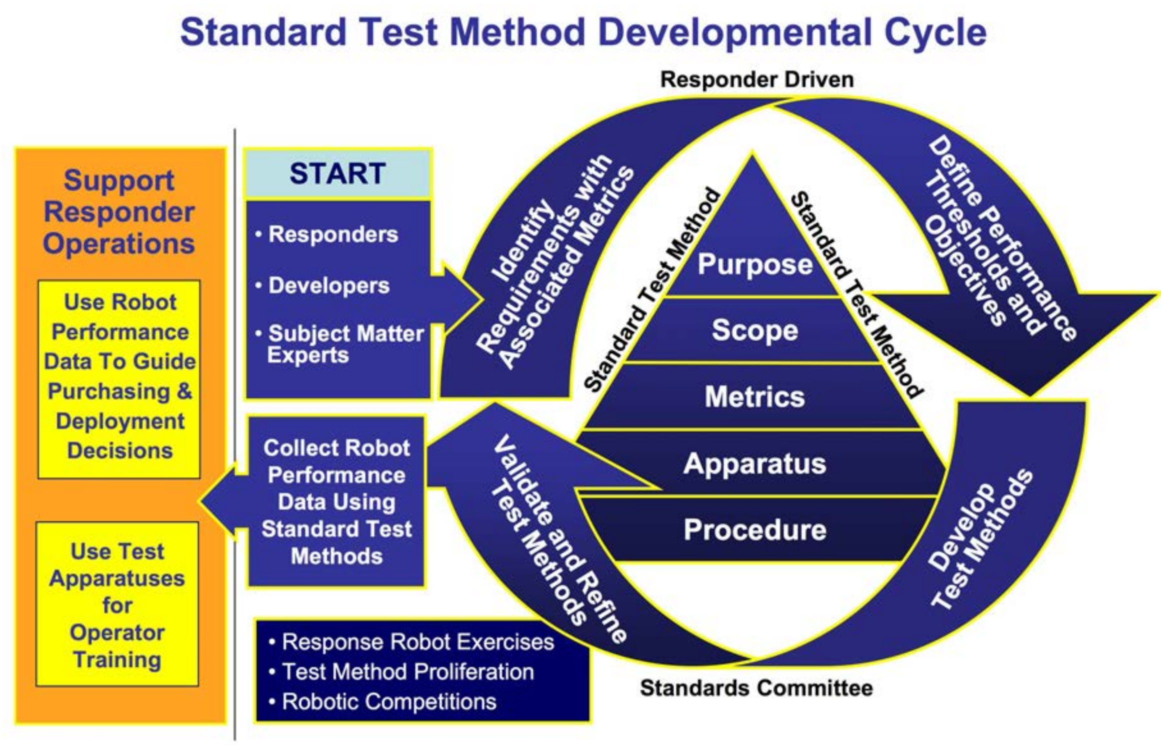


FIGURE 1: The development cycle for DHS-NIST-ASTM International Standard Test Methods for Response Robots is a responder driven process for generating, validating, and standardizing test methods.

Objective

This document describes the suite of ***DHS-NIST-ASTM International Standard Test Methods for Response Robots.*** These standard test methods measure baseline robot/operator capabilities necessary to perform operational tasks defined by emergency responders, soldiers, and their respective organizations. No single test method describes a robot's overall capabilities. But any user can select a set of standard test methods, 20 or more, to represent capabilities necessary to perform intended missions. If a robot/operator cannot complete the specified set of standard test methods, they will not likely be able to perform the operational tasks during deployments. Conversely, if a robot/operator can practice and demonstrate success (to statistical significance) across the specified set of representative test methods, it is much more likely that the robot/operator will be able to perform the associated operational tasks during deployments, even with the increased complexity of unknown environments.

Repeated testing within standard test methods followed by operational scenarios with embedded test apparatuses can provide training with inherent measures of proficiency prior to deployment. These standard test methods are designed with increasing levels of apparatus difficulty to ensure that all robots, and even novice operators, can be measured performing basic tasks. The test apparatuses provide incrementally more difficult settings to challenge even the most advanced robot capabilities. The key to developing a good standard test method is ensuring that the figurative measuring stick is long enough to capture performance at both ends of the available robotic capabilities spectrum, and that it separates performance results in between. These standard test methods also enable controlled introduction of environmental complexity such as darkness, terrains, confined spaces, etc. The overall suite is expanding to answer new mission requirements every year, and some test methods have already been updated to widen their scope for testing autonomous systems.

FIGURE 2: The suite of standard test apparatuses provide a breadth first approach toward testing. They rely on inexpensive materials that can be sourced around the world. They can be embedded into operational scenarios to provide measures of operator proficiency.

This document describes how to use these standard test methods to evaluate robots, specify and defend purchasing decisions, and train operators with measures of proficiency. Appendices reference related documents with complete listings of test method descriptions and captured robot capabilities data so they can be updated more often. The test method descriptions contain brief discussions of the purpose, metric, apparatus, and procedure. More comprehensive discussions are available within the standards published through ASTM International. Test methods in the prototyping stage are not included. The robot capabilities data includes only data sets captured during comprehensive testing events using 20 or more test methods so the trade-offs in competing capabilities are clear. Robot data sets are presented within the class of similarly sized and equipped robots defined by particular test event sponsors to convey the overall capabilities of each robot with general class of robots. Additional robot data collections are ongoing and plans are being made to make available all the robot data sets on the DHS website for *System Assessment and Validation for Emergency Response (SAVER)* so they may be updated regularly with the latest test results and disseminated.

What is a Standard Test Method?

The difference between a "standard test method" and a "standard equipment specification" is that standard test methods focus on measuring capabilities while imposing no design constraints or other specifications on the robotic system. This approach doesn't inhibit innovation while robot developers work toward implementing and hardening solutions to sometimes competing requirements. Standard test methods are essentially just agreed upon ways to test robotic capabilities. So we are not developing a specification for a "standard robot" of any kind, like an equipment standard. Rather, we are developing a standard way to test remotely operated robotic systems.

Our consensus standards development process is being conducted within the ASTM International Standards Committee on Homeland Security Applications; Operational Equipment; Robots (E54.08.01) which includes equal representation of robot developers, emergency responders, and civilian/military test administrators. Standard test methods include detailed descriptions of the following:

- **Apparatus:** A repeatable and reproducible representation of tasks that users expect the robot to perform reliably. Apparatuses are typically inexpensive and made of readily available materials so they can proliferate widely.
- **Procedure**: A script for the test administrator and the robot operator to follow.
- **Metric**: A quantitative way to measure the performance of the robot. Capability objectives and lower thresholds of acceptability may be specified using this metric.

How Do These Standard Test Methods Work?

There are only a few simple rules:

- A specific robot configuration (developmental or purchasable) must be clearly described and then subjected to all applicable test methods to measure its particular combination of capabilities. This enables easy analysis of performance trade-offs for different robot models, or different configurations of the same robot model.
- The robot must always be controlled from a remote operator station, out of sight and sound of the test apparatus but within radio comms range (except for the radio comms test methods).
- "Expert" operators designated by the robot developer are used to capture the best possible performance of the robot for comparison purposes. No robot developer should promise any more than this level of capability. And everybody else can try to train up to some percentage relative to that "expert" operator proficiency.
- Autonomous or assistive capabilities are encouraged and should improve remote operator efficiency, reduce workload, and/or improve survivability of the robot down-range (e.g. self righting, centering between obstacles, retro-traverse back into radio comms range, etc.). In the case of systems with some autonomy, the overall system performance is still captured in the same way within the test apparatuses. However, the percentage of time the operator interacts with the system is also captured as a rough measure of workload. More advanced measures are possible as more systems attempt to deploy such capabilities. We expect to be able to define sets of standard apparatuses as repeatable mission complexities and establish "levels of autonomy" based on the NIST's Autonomy Levels for Unmanned Systems (ALFUS) scale.

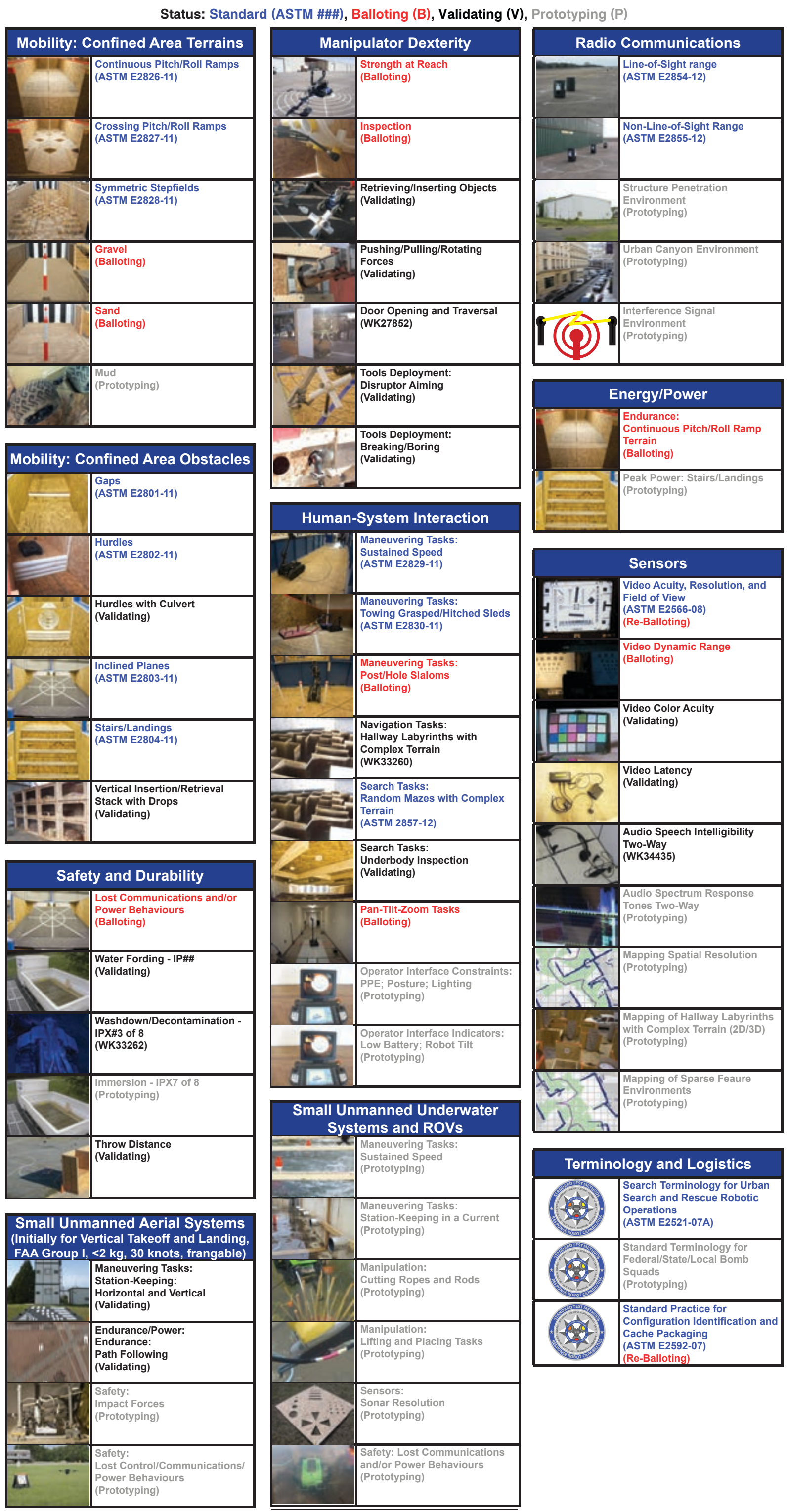


FIGURE 3: The suite of DHS-NIST-ASTM International Standard Test Methods for Response Robots provides a breadth first approach to testing in order to capture statistically significant performance in a rapid and repeatable way. This allows testing more often to ensure that system changes quantifiably improve overall performance.

Who Benefits From These Standard Test Methods?

For robot developers, standard test methods provide robot developers tangible representations of operational requirements to help understand mission needs, make trade-off decisions, inspire innovation, measure incremental improvements, highlight break-through capabilities, and harden new approaches. The test apparatuses can be used to practice and refine systems during development to help debug issues, to identify necessary improvements, and then to convey system capabilities to interested users.

For program managers, standard test methods clearly articulate program goals in terms of desired combinations of robot capabilities. They can encourage innovation and measure outcomes, which can be remotely monitored. Program deliverables can be tied to demonstration of capabilities within specified combinations of standard test methods (to statistical significance). Final evaluations can be conducted using embedded test apparatuses in operational scenarios.

For responders and soldiers, and their respective organizations, standard test methods provide objective and repeatable robot capabilities data. Users can trust the data captured at any participating standard test facility, no matter when or where in the world the testing was conducted. This helps inform and guide purchasing decisions by clearly indicating the range of available robot capabilities in any given test method, and the particular combinations of capabilities available in certain robot configurations. Responders and soldiers should no longer specify a series of “requirements” to guide a robot purchase, because all too often those requirements are competing with each other in the context of technical practicality, reliability, cost, etc. Over and over again this process has led either to disappointment, excessive cost, or both. Rather, responders and soldiers should make purchasing decisions by specifying available combinations of robotic “capabilities” as demonstrated within a suite of standard test methods. This process recognizes that robot developers have already made trade-off decisions in trying to implement functional and affordable systems while considering technical practicality, reliability, cost, etc.

How Are These Standard Test Methods Developed?

The process used to develop standard test methods begins with specific robot capability requirements defined by emergency responders and soldiers that could make their deployments more effective, efficient, or safe. Each requirement must have an associated metric as a way to measure the capability. Sometimes the metric is as simple as elapsed time. But we try not to make every test trail a race. Rather, we try to establish a task-based testing paradigm emphasizing statistically significant repetitions with time per task as a secondary measure of efficiency. Users can specify capability objectives and lower thresholds of performance below which will not be acceptable to communicate a range of acceptable performance that robot developers can use to make trade-of decisions. Where such robot requirements already exist, such as for some bomb squad applications, they may be used directly. Some responder communities, such as FEMA urban search and rescue teams, were solicited during the course of this project and have provided over 100 such requirements for 13 different robot categories.

The requirements are prioritized by responders and prototype test apparatuses are generated to isolate, repeatably test, and measure robot performance. Response robot evaluation exercises are hosted in responder training facilities to allow responders to validate the test methods and learn about emerging robotic capabilities. International robot competitions featuring the prototype test apparatuses are used to inspire innovation, leverage robot traffic (over 100 missions per competition), to refine apparatus designs. Robot competitions also support proliferation of the standard test methods for practice by encouraging benchmark comparisons for qualification. Once the apparatus is validated by responders and test administrators, it is balloted through the ASTM Standards Committee on Homeland Security Applications; Operational Equipment; Robots (E54.08.01).

FIGURE 4: The approach toward developing DHS-NIST-ASTM International Standard Test Methods for Response Robots includes iterative validation events on either side of the standards process shown down the center panel. The right panel shows outreach to robotic manufacturers and researchers in the form of practice events and robot competitions to inspire and guide robot development while gaining feedback on prototype test methods. The left panel shows outreach to responders and soldiers in the form of Response Robot Evaluation Exercises hosted at responder training facilities. This is where we validate the test methods and introduce embedded test apparatuses into operational scenarios.

Definition of a Response Robot (Ground, Aerial, Aquatic)

The suite of ***DHS-NIST-ASTM International Standard Test Methods for Response Robots*** was developed to measure the range of capabilities of response robots independent of robot size, and broadly applicable to a variety of missions intended by responders and soldiers. The working definition of a “response robot” is a remotely deployed device intended to perform operational tasks at operational tempos from safe operational stand-offs. It should provide remote situational awareness, a means to project operator intent through the equipped capabilities, and improve overall effectiveness of the mission while reducing risk to the operator. Key features include:

- Rapidly deployed
- Remotely operated from an appropriate standoff
- Mobile in complex environments
- Sufficiently hardened against harsh environments
- Reliable and field serviceable
- Durable or cost effectively disposable
- Equipped with operational safeguards

Response robots include ground vehicles up to about 500 kg (1100 lbs) or so; small unmanned aerial systems (sUAS) within the FAA Group I defined as under 2 kg (4.4 lbs), under 30 knots maximum forward airspeed, and harmless upon impact; aquatic systems including remotely operated vehicles for swift water rescue, inspection of bridges, ports, ship hulls, etc.

FIGURE 5: Over 100 robots have been tested to varying degrees of completeness across the roster of standard test methods. These are some examples of response robots that show the range of sizes of ground robots, some examples of vertical take-off and landing sUAS, and some small aquatic ROVs that are indicative of the robots targeted within this project.

Summary of This Standard Testing Approach

The suite of standard test methods provides a rapid, quantitative, and comprehensive evaluation of remotely operated robots. Individual tests typically take less than an hour, except for certain endurance tests. So given typically 20-30 applicable test methods, a reliable robot can get through all the testing in less than a week. It is a purposefully breadth-first approach since robot capabilities data is short-lived as robot technologies and implementations change and mature. The resulting statistically significant capabilities data defines the overall characteristics of a given robot, and places that robot within context across its class of related robots.

Robot configurations are typically subjected to 20-30 test methods chosen by a sponsor or procurer to capture baseline capabilities necessary for intended missions. The chosen combination of test methods help determine capability trade-offs, reliability, etc. “Expert” operators provided by the developer perform the tests for the standards process to capture the best possible performance for comparison. Robot developers should not promise any better performance. The operator controls the robot from a remote location, typically out of sight and sound of the test apparatus but within radio communications range (except for the radio comms test methods), to maintain total reliance on their system interface at all times. Test trials include between 10-30 repetitions to achieve statistical significance (80% reliability with 80% confidence). Interactions with incapacitated robots during test trials are allowed to reset the robot to the start point or to make minor repairs with no spare parts. Up to three interactions are allowed within a thirty repetition trial to maintain statistical significance. Every interaction is documented in Field Maintenance & Repair forms to identify indications of issues, remedies implemented, and tools used throughout the testing process. Testing events typically take less than one week and include some operational tasks with embedded standard apparatuses to leverage and extend the challenges imposed. Testing may take place at any robot test facility housing the suite of standard test methods, or at Response Robot Evaluation Exercises typically administered by NIST. Robot testing may also be conducted at events where the robots and user communities are typically assembled such as conferences or regional training events.

New test method requirements can come from any source: responder, robot developer, procurement sponsor, program manager, or other source. For a recent procurement process, NIST fabricated new draft standard test methods for a class of ultra lightweight reconnaissance robots under 10 kg (22 lbs). The sponsor had a specific requirement for durability with an emphasis on throwing the robot during deployment, possibly over a wall, onto a roof, or simply past some obstacles. NIST prototyped the Throw Distance test method to measure the down-range distance a robot could be thrown over a 2.4m (8ft) wall. After each throw, reconnaissance tasks ensured that the robot remained functional prior to the next throw. These tasks included driving the nearest circular line on the ground (control/latency), identifying hazmat labels on a barrel placed at the center of the circle (camera pointing and visual acuity), and listening to audible random numbers played within the barrel (audio acuity and 2-way communications, if equipped). As with the entire suite of standard test methods, other operational targets can be used for robots equipped to detect explosives, radiation sources, hazardous chemicals, etc. But this test was sufficient for the roster of robots tested. The test method validation process started immediately within the prototype apparatus. The developers all learned about their systems as they reluctantly began to throw (or haltingly toss) their robots over the wall. The engineers on the teams considered how to soften impact to survive 10 or more sequential repetitions. Some changed their wheel designs, sprocket designs, and/or materials. One developer used a more sophisticated “flight” behavior to maintain heading and orientation -- a real innovation.

FIGURE 6: A) The draft standard test apparatus Throw Distance includes a 2.4m (8ft) tall wall to throw over, an adjoining remote control station to limit sight lines down-range, and landing locations with 4m (13ft) diameter circular lines for the robot to follow. B and C) Robots may be thrown over the wall in any manor with a two-step approach while staying on the 2.4m (8ft) OSB panel on the ground. D) Hazmat targets are placed at the center of the circle for the robot to identify to demonstrate functionality after each throw. Colored discs on the ground mark the landing locations for each trial which can add up to 30 repetitions.

Ultimately this process worked for robot developers, procurement sponsors, and the end users as the final robots delivered were clearly more capable and reliable than the initial set tested. Without this process and the design iterations and revisions it inspired, they almost certainly would have failed at some point in the hands of the end users in the field.

Outcomes: Response Robot Capabilities Compendium and Collaborating Test Facilities

There are several intangible outcomes from this process as well. For example, the standard test methods clarify communications and expectations between responders and robot developers. The physical test apparatuses and agreed upon metrics help convey mission requirements to robot developers while refining expectations of capabilities for responders. And, of course, they help measure the results.

However, the main outcome from this suite of standard test methods is a growing Response Robot Capabilities Compendium, which is a database of test results and associated bar charts describing the various baseline capabilities of tested robot configurations -- almost like robot DNA where no two are exactly similar. Robot data generated by any participating standard test facility can be included in the capabilities compendium and compared no matter where or when the testing occurred (e.g. U.S., Germany, Japan, etc.). Bar graphs for each individual test method show the capability of each robot configuration relative to the class of robots within that test method. Some missions may require “best-in-class” performance for a particular capability, while allowing average performance in other capabilities. In general, the capabilities compendium and the bar charts help inform responders, soldiers, and their respective organizations about the trade-offs of capabilities currently available, and begin to align expectations regarding what the robots can do during deployments.

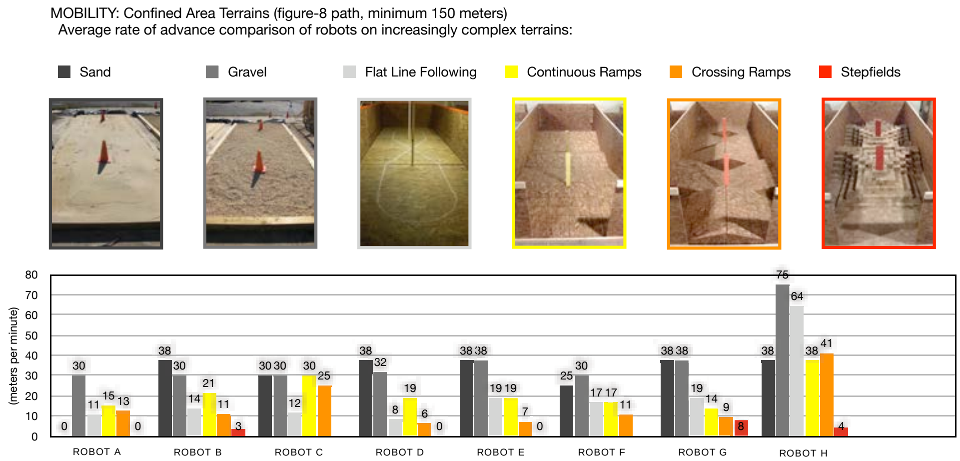


FIGURE 7: Robot data in the form of bar charts provide easy ways to compare different robot models within a range of applicable test methods. Any end user can decide which test methods are important for their intended missions, and focus on the robots that demonstrate the right combination of capabilities to take a closer look. The variety of mobility terrains are shown here as an example, but every sub-suite of standard test methods produces similar bar charts across all robots tested to statistical significance.

Currently, all robot capabilities data has been captured either at NIST or at a NIST hosted Response Robot Evaluation Exercise. The data was collected primarily to support the standardization process, but it has already proven useful for guiding several robot procurements. In 2013, NIST opened a new robot test facility on its main campus in Gaithersburg, MD. The nearly 1,000 square meter (10,000 square foot) facility will support continued development of the standards and maintain calibration experiments with a growing roster of collaborating test facilities in the U.S. and internationally. Many other locations host particular subsets of the standard test methods to support robotic development, program management, or procurement needs.:

- Southwest Research Institute, San Antonio, TX (opened 2010)
- International Rescue Systems Institute, Kobe, Japan (opened 2011)
- Bundeswehr, Koblenz, Germany (opened 2012)
- University of Massachusetts, Lowell, MA (opened 2013)
- Curtin University, Perth, Australia (expected 2014)
- SPAWAR, San Diego, CA (expected 2015)

FIGURE 8: NIST’s new Robot Test Facility on the main campus in Gaithersburg, MD contains all the test methods and prototypes being validated.

Evaluating Response Robots

This suite of standard test methods captures baseline capabilities of remotely operated robots intended for complex environments with variable terrains, assorted obstacles, confined areas, lighting changes, etc. Selected suites of standard test methods can represent intended missions. Usually 20-30 test methods are selected for a comprehensive evaluation, even if only 5-10 of them are considered high priority indicators of mission success. Each test method typically takes less than an hour to complete, so the overall test time is between 20-30 hours, less than a week. This usually leaves time to conduct some operational scenarios with embedded standard tasks as described later in this document.

There are just a few simple rules:

- Robots must be described in detail to capture the major sub-systems that could affect performance: wheels vs. tracks; tether vs. battery (how many and which chemistry); radio comms (power, frequencies, and antenna gains); external sensors; payloads; manipulators; etc. That robot configuration should then be subjected to all applicable test methods to measure inherent trade-offs in capabilities (e.g. addition of the manipulator inhibited stair climbing). This enables comparisons of different configurations of the same robot model or different robot models. Any variation in the robot configuration should be retested across the entire suite of test methods to capture a new set of capabilities.

- Robots are always remotely operated, out of sight and sound of the test apparatuses but within communications range (except for the radio communications test methods), so the operator must rely on the system for all their situational awareness. Practice within the test apparatus prior to testing is welcomed. The operator may refine particular techniques with “eyes on” during practice. But testing is always remotely operated.

- “Expert” operators designated by the robot developer are used to capture data for the standards process. Since all incentives are aligned, this can be considered the best possible performance of the robot and so good for comparisons. No robot developer should promise any more than the capability demonstrated. And everyone else can try to train up to some percentage of that “expert” operator proficiency.

- Autonomous or assistive capabilities are encouraged and should improve remote operator effectiveness, reduce workload, and/or improve survivability of the robot down-range (e.g. self righting, centering between obstacles, retro-traverse back to within radio comms, etc.). In these cases the overall system performance is captured in the same way but the percentage of time the operator interacts with the system is also captured as a rough measure of operator workload. More advanced measures are possible as more systems begin to deploy such capabilities. We expect to be able to define sets of standard apparatuses to define certain repeatable mission complexities and establish “levels of autonomy” based on the NIST’s Autonomy Levels for Unmanned Systems (ALFUS) scale.

Identifying the Robot Configuration To Be Tested

The first step is to fully identify the particular robot configuration to be tested according to the procedure described in the *Standard Practice for Configuration Identification and Cache Packaging (ASTM E2592-07)*. The robot developer, sponsor, or procurer may select the particular robot configuration to be tested for a variety of reasons and the choice is totally up to them. That configuration should get assigned a make, model, and unique configuration name to define the choices of battery type, radio system, sensors, payloads, manipulator, etc. This is to ensure that the configuration can be replicated for any subsequent testing if necessary, and so it can be specified for purchasing. The robot should be weighed and photographed in the 20 cm (8 in) metered backdrop shown below, including close-up images of all configuration components and the operator interface. Images of the tool set and loaded shipping containers should be captured along with the shipping weights. Videos of routine robot maintenance tasks should also be captured, including a battery change, track change, payload removal/attachment, etc. The tool set should also be photographed at the start of the test week, but if any additional tools are used during Field Maintenance Resets throughout the week, they need to be added to the image so that it is clear what tools are necessary for deployment.

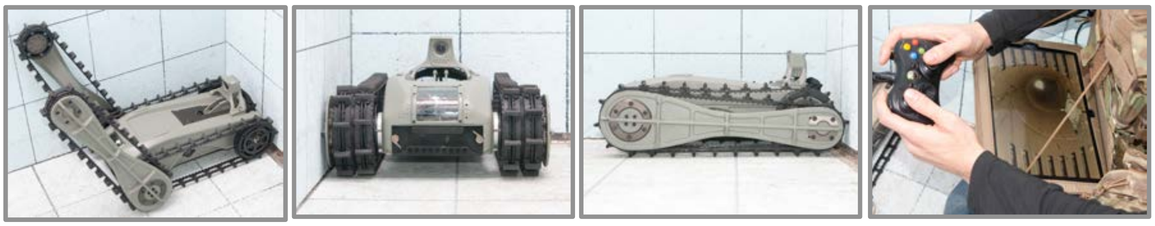


FIGURE 9) Robot configuration identification involves taking pictures of the robot in front of the metered backdrop with 20 cm (8 in) squares. Detail images should capture the sensors, radio, payload locations, operator control unit, etc.

FIGURE 10: A) The robot packaging container sizes and weights should also be documented along with the tool set. Some responders have clear guidelines for how long they should be operational in the field without re-supply. So the number of batteries to pack, for example, is a big issue.

Selecting Test Apparatus Sizes and Targets for Expected Missions

These standard test methods provide baseline capability evaluations prior to more operational training scenarios. The apparatuses associated with these test methods challenge specific robot capabilities in repeatable ways to facilitate direct comparisons of different robot models and particular configurations of similar robot models. Many of the test apparatuses use terrains, targets, and tasks that are intentionally abstract to facilitate standardization which requires capture of repeatable test results within a test facility and reproducible results across different test facilities (internationally too). The test apparatuses are generally fabricated using readily available materials to remain inexpensive and proliferate widely. Robot developers can build them to support innovation, refinement, and hardening. Responders can build them for robot evaluations or proficiency training. Many test apparatuses are constructed with oriented strand board (OSB) plywood to provide a common friction surface similar to dust covered concrete. The specific terrains, targets, and tasks used can be modified or replaced with more operationally representative examples while using the same apparatuses and procedures to further support training, practice, and comparison of specific system capabilities. Operational scenarios with embedded standard apparatuses can measure and compare performance within degraded operational environments relative to that measured within the more controlled standard apparatuses.

The suite of standard test apparatuses are specified in three different sizes to reference various confined spaces in which robots must operate. The smallest size apparatus has a minimum of 0.6 m (2 ft) lateral clearance. The mid size apparatus has a minimum of 1.2 m (4 ft) lateral clearance. The large size apparatus has a minimum of 2.4 m (8 ft) lateral clearance. For the terrain apparatuses, for example, with their internal figure-8 path for robots to negotiate has a minimum clearance between the central pylons and the perimeter wall. For the mobility obstacle apparatuses such as the stairs, the minimum clearance is between the two side walls. In each size obstacle apparatus, the launch and landing areas on either side of the obstacles can be made square with removable interior walls hung between the perimeter walls. With the interior walls installed, the launch and landing areas become square requiring robots to turn in place similar to interior stairwell landings. Without these interior walls installed, the mobility obstacles are more like outdoor pass-through obstacles.

When choosing which size standard apparatus is right for your intended mission, consideration should not necessarily be tied to the size of the robot. Rather, the selected apparatus size should reference the expected clearances within the intended mission environments. For example, a robot expected to board an airplane on a runway should test in the small size apparatuses to capture the most applicable test results. The goal is to determine if the robot can ascend narrow stairs in all weather conditions, turn in place within the rather small entry area of the plane, traverse down the aisle between the seats, and ultimately reach to the side and manipulate objects in the overhead bins, on the seats, or under the seats. The small size apparatuses are for any narrow access environment including buses, trains, dwellings, etc. The mid-size apparatus with 1.2 m (4 ft) lateral clearance can be used if expected deployments involve indoor lobbies of public buildings, parking structures, etc., although doorways and spaces between parked cars are not often that large. The large size apparatuses with 2.4 m (8 ft) lateral clearance are for parking lots, road sides, and outdoor areas in general.

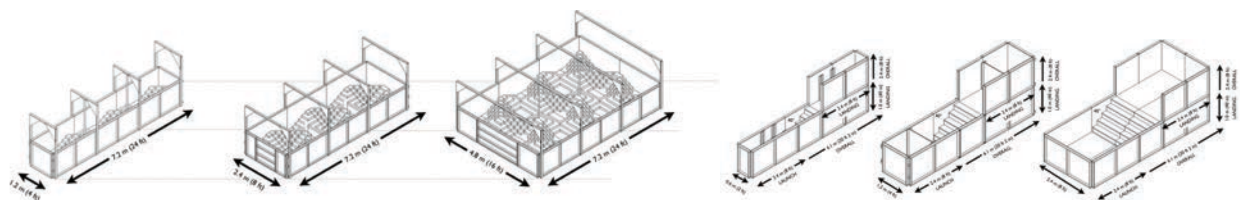


Figure 12: Three sizes of apparatuses reference the various intended mission environments rather than simply the size of the robot. Examples shown are Mobility Terrains: Stepfields and Mobility Obstacles: Stairs

Visual targets are used within the test apparatuses to evaluate the visual and color acuity of robots under test in conditions with the lights on (> 300 lux, which is equivalent to an office environment) and with the lights off (< 0.1 lux, which is almost unnavigable without additional light). Visual targets consist of Landolt “C” visual acuity charts, also known as Tumbling C’s, and standard hazardous material labels and placards. Landolt Tumbling C’s are essentially line resolution tests that can be read through the remote operator station and announced by a robot operator to the test administrator. The test administer then verifies the reading before scoring the result on the form. A correct reading of a particular line of 5-10 Tumbling C’s produces a numeric measurement of the visual acuity that can be referenced to average human vision. The visual acuity test method uses Tumbling C charts to identify the robot’s far field and near field visual acuity. Concentric C versions shown below are also used typically embedded into other test apparatuses of scenarios as visual targets to provide an indication of the robot’s visual acuity.

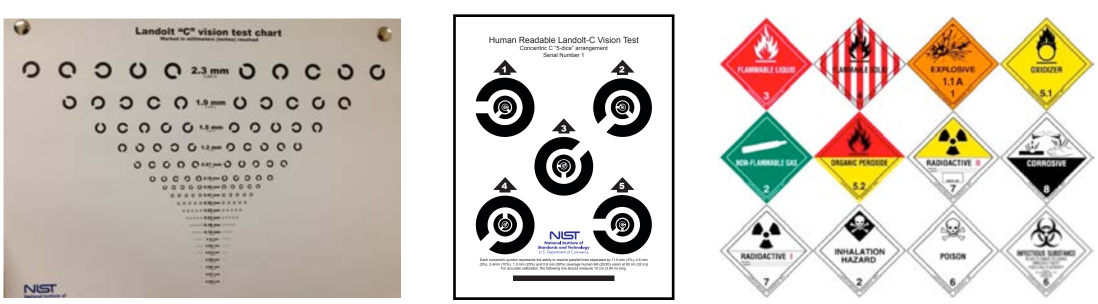


Figure 11: Visual targets are used within the test apparatuses to evaluate the visual and color acuity of robots under test in lighted and dark conditions. A) Landolt C chart for near field vision (far field chart looks similar, but larger). B) Concentric Landolt C charts for embedding into test apparatuses and operational scenarios are generated by NIST so can be scaled to any size. They appear in the 5-dice pattern to emphasize at least some statistical significance. C) Hazardous materials labels, which are 10 cm (4 in) on a side.

Hazardous materials labels provide a variety of standard visual targets that introduce modest complexity for visual identification tasks and operational relevance for some users. The labels contain four attributes including color, icon, text, and number. The text and numbers are sized for average human acuity. Identification of any three of four attributes is considered successful identification of the target.

More operationally relevant objects can be used to provide targets for reconnaissance tasks. For example, simulated pipe bombs, simulated artillery shells, timer devices, power sources, cell phones, detonation cords, trip wires, etc. Non-visual targets can also be used to test the capabilities of onboard sensors. For example, we have placed trace chemical, radiological, and explosive sources within the test apparatuses to identify proximity at initial detection and then localization accuracy of identified sources.

Testing to Statistical Significance

Robot capability data collections are conducted using the test apparatuses and associated test procedures to capture robot and remote operator performance across a statistically significant number of repetitions. Robots are tested to completion of certain tasks with "expert" operators designated by the developer to capture the best possible capability for a given robot in a given apparatus. The number of repetitions for each test method is determined by ASTM (or the test Sponsor) using statistical principles while considering test administration practicalities for longer tests, such as the Endurance test method. The elapsed time of each test is typically not included as a standard metric to de-emphasize speed in favor of task completeness, although the test duration is captured secondarily as an efficiency measure or the trial may be time limited. Timing measures are typically reported as an average time to perform each repetition, or as an average time to perform a particular sub-task within a test method that can produce varying levels of completeness. This enables even novice operators to quantitatively establish their proficiency as a percentage of “expert” performance within the same test method. The test method forms use graphical representations of the data to convey an understanding of the test results and facilitate comparisons across different robot configurations.

Test trials typically consists of 30 repetitions to demonstrate statistical significance to at least 80% reliability with 80% confidence. Statistically significant trials are shown by a green check mark on the upper right corner of the form. During the first trial within a particular apparatus setting, the Test Administrator may stipulate that the robot was dominating the apparatus at that setting after demonstrating the first 10 successful repetitions with no failures. However, if there are any failed repetitions, a second set of 10 repetitions is required. For a trial to be noted as statistically significant, no more than 1 failure in 20 repetitions, or 3 failures in 30 repetitions are allowed. This enables setting the apparatus to some known capability and quickly moving toward more aggressive apparatus settings to determine the limit of the robot’s capabilities. If more than one test trial is attempted for any reason, all but the initial test trial must be tested to 30 repetitions for a given apparatus setting to ensure statistical significance across test trails.

Resetting the Robot During Test Trials

During test trials robots may become stuck or incapacitated. The operator has the option to call a Field Maintenance Reset which allows the operator to leave the operator station, reset the robot to the start position within the test apparatus, and perform routine maintenance for up to 10 minutes (or other limit set by the sponsor). The goal is to allow some interaction with the robot in order to continue the trial to completion. The tool set captured in the cache packaging tools picture is allowed with the robot at the start point. No spare parts are allowed except tape, cable ties, etc. A Maintenance/Repair form will be filled out including the test method, indication of failure, remedy used, tools used, and overall time to execute the field maintenance. The maintenance interaction may be captured on video as well to be used later for training or other purposes. This policy is intended to mimick a field maintenance procedure, so the robot is considered to be down range with some limited number of tools and personnel. Any person or team of people may interact with the robot at the start point but the robot may not be removed from the start point. The actual list of field maintenance tools necessary to keep the robot operational will be evident after the week of testing is complete. Likely points of failure on the robot will also be clear.

Abstaining from Test Methods

Each robot configuration should be tested in all applicable test methods and may attempt each test as many times as necessary to attain a satisfactory result. Robots may abstain from a particular test method when considered not applicable or choose not to release the resulting data from a specific test trial when considered not successful. This encourages robot developers to attempt test methods and learn about their systems without consequence. In either instance, the page will be marked as “ABSTAINED” to indicate that the test method was available at test time and the developer acknowledges the omission of performance data.

DID NOT COMPLETE TEST
ABSTAINED
OR WILL NOT RELEASE DATA

FIGURE 13: This marking placed over a test method form indicates that the test method was available at test time and the developer acknowledges the omission of performance data.

Although some robot implementations may not be designed or equipped for particular test methods, (e.g. robots without manipulators in the manipulator test methods) this testing methodology makes no assumptions regarding capabilities. Specifics of particular robot configurations should be considered when the robot has abstained from a given test method. If the test method is considered critical to the operational needs of the sponsor or user, the test should be considered failed until the robot can demonstrate satisfactory performance at a later date.

If a robot returns to the test facility to quantify improvements in performance for a particular robot configuration, the robot will be subjected to a subset of tests representing each of the test method suites to ensure compatibility with the previous configuration. An example roster of make-up tests include: Endurance, Radio Comms, Decreasing Slalom, Crossing Ramps, Inclined Plane, Pan-Tilt-Zoom Tasks.

Documenting Test Trials

Every standard test method includes a form to capture all the associated information related to a particular test trial. Beyond the basic information about where and when the test trial occurred, essential information about the particular configuration being testing, and the settings involved with the test apparatus and environmental conditions, the test form documents what happens on every repetition during the trial along with the timings involved.

The forms are graphically presented to convey performance at a glance where possible. They display success clearly, and identify where and when anomalies occur. Some anomalies like reseting a stuck robot, or reboots of the operator interface, are noted and timestamped so they can be quickly referenced in the quad-screen video described below. If a Field Maintenance Reset occurs, an additional form captures the operator’s indication of a problem, the remedy implemented, the tools used, and the timings involved. This separate set of field maintenance issue reports are collected across all test trials within the testing event, potentially up to a week of operation. Robot developers can learn about reliability or maintenance issues of particular components or sub-systems. For example, if the right track comes off several times in different test apparatuses, the engineers will want to understand the particular causes to make refinements.

Test forms such as the one shown below indicate successful repetitions with blue check marks and anomalies with red Xs and associated timestamps. Any administrative pauses, to fix an apparatus or camera for example, are indicated with an orange check mark and timestamp. Administrative pauses are not reflected in the elapsed time of the trial. The fictional form shown below indicates a robot with statistically significant performance and reasonably efficient repetitions when functioning properly. But the robot had field maintenance issues which would be documented to inform the robot developers so they can refine their systems.

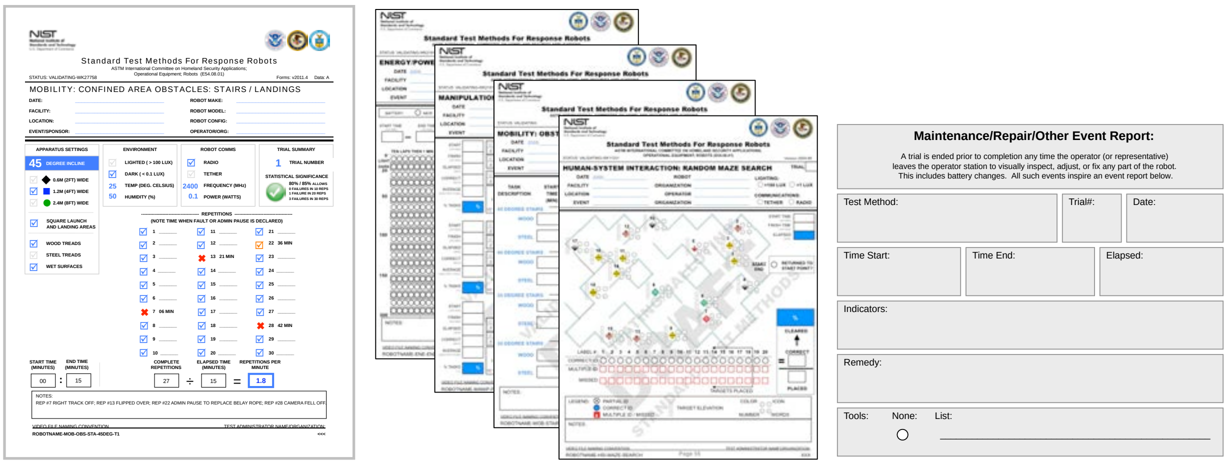


FIGURE 14: A) A sample test form showing successful repetitions in blue check marks and anomalies in red Xs within a 30 repetition trial. B) Examples of other data collection forms designed to convey capabilities and anomalies graphically. C) Each Field Maintenance Reset inspires a separate form to capture the indications of the issue, the remedy implemented, the tools used if any, and the total time needed for the repair.

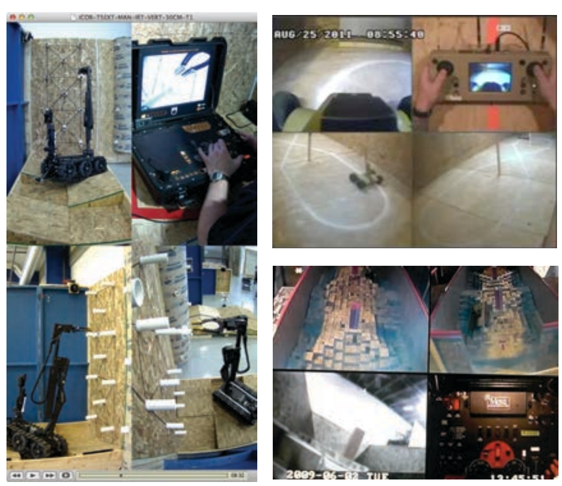


FIGURE 15: Quad screen video with four simultaneous video feeds show the robot and operator actions along with data such as robot position tracking when appropriate.

Capturing Quad-Screen Video

The main goal of this project is to capture and convey robot performance. Although the forms are reasonably graphical in an attempt to describe test trial results, there is no substitute for watching the robot succeed or fail within a standard test apparatus. There are lessons in both outcomes for experts and novices alike. So quad-screen video is another product of the testing process. Four simultaneous video feeds capture the robot from two or more angles, what the operator is seeing, and what the operator is attempting to do. Additional measurement systems can also be included when appropriate. The quad-screen video capture system inherently time-syncs the four video images so there is no question about the cause and effect of any particular operator interaction. There is no video editing necessary after the trial, which can be tedious and expensive. The product is immediately useful and is often reviewed by expert operators to see tendencies in their operation. Novices can benefit by stepping through mistakes made during training because their impressions from the operator interface perspective is typically incorrect. Whether the trail was a success or failure, contains highlights or bloopers, it is always very informative.

Reviewing Performance Data

The robot performance data captured with the test methods is typically subjected to five sequential reviews to ensure that it accurately reflected the performance captured at test time. The first review is performed by the Test Administrator and robot operator together, just after the test trial is completed. At that time they discuss the outcome of the trial, review the information on the form for accuracy, and decide if additional trials are necessary to capture the best possible performance. Both the Test Administrator and the robot operator initial the test form to show they concur with the results.

The second and third reviews are typically internal to the team of test administrators conducting the event. This ensures that the data gets transferred correctly from the handwritten form into the electronic form (online electronic forms, or eForms, are coming soon). Test Administrators as a team discuss anomalies that may have occurred during testing, identify areas of confusion, and calibrate exception handing procedures. The eForms file naming convention includes a the robot make, model, configuration, test method, apparatus setting, trial number, and date stamp such as v2011.3. The fourth review is done by the robot developer using the eForms. Any issues or modifications are listed and discussed by telephone or video conference with the Test Director. The fifth and final review is conducted by the Test Director. This is typically done with all the quad video queued up to view quickly as an overall product quality check (and system refinement if necessary) before introducing the eForms and video into the Response Robot Capabilities Compendium.

Guiding Robot Purchases

Selecting a Suite of Representative Test Methods

The first step in specifying and defending a robot purchase, especially one requiring a sole source justification, is to select a prioritized list of standard test methods that address intended mission capabilities. The prioritized test methods can be 5-10 within the context of 20-30 test methods overall to evaluate the entire system.

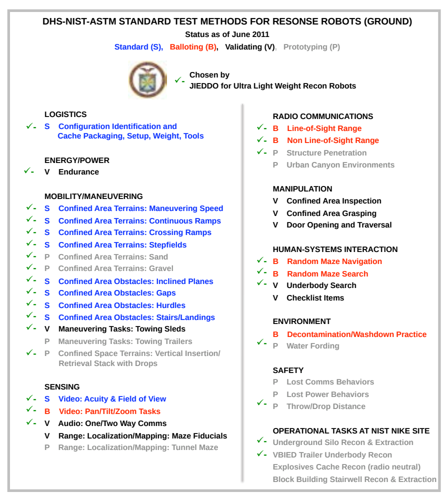
The next step is to specify performance objectives in each test method along with a lower capability thresholds below which is not acceptable. The combination of capability objectives and lower thresholds provides a design space for engineers to make inevitable trade-offs for cost or reliability. This flexibility also works well when trying to specify a robot purchase while balancing each robot's specific trade-offs.

For example, a lightweight reconnaissance robot may be roughly described by the following prioritized set of standard test method. These would be considered essential for the intended missions:

- weight (robot + operator control unit) -- *need to carry them far down range*
- visual acuity (near field, far field) -- *need to monitor far field signs of trouble and read near field*
- underbody inspection (20 cm / 8 in) -- *need to inspect or take shelter under vehicles*
- stair climbing (45° incline, steel treads, wet); -- *need to ascend to upper elevations of dwellings*
- radio comms range (line-of-sight, non-line-of-sight); -- *emphasis on non-line-of-sight when inside*

There are other important capabilities that must be monitored to ensure overall system functionality:

- endurance (distance, average speed).
- reliability (field maintenance and resets);
- water fording (10 cm / 4 in);
- throw distance (2.4 m / 8 ft)
- and a dozen others are recommended to define the overall characteristics of a given robot.



DHS-NIST-ASTM STANDARD TEST METHODS FOR RESONSE ROBOTS (GROUND)
Status as of June 2011
Standard (S), Balloting (B), Validating (V), Prototyping (P)

✓- Chosen by JIEDDO for Ultra Light Weight Recon Robots

LOGISTICS
✓- S Configuration Identification and Cache Packaging, Setup, Weight, Tools

ENERGY/POWER
✓- V Endurance

MOBILITY/MANEUVERING
✓- S Confined Area Terrains: Maneuvering Speed
✓- S Confined Area Terrains: Continuous Ramps
✓- S Confined Area Terrains: Crossing Ramps
✓- S Confined Area Terrains: Stepfields
✓- P Confined Area Terrains: Sand
✓- P Confined Area Terrains: Gravel
✓- S Confined Area Obstacles: Inclined Planes
✓- S Confined Area Obstacles: Gaps
✓- S Confined Area Obstacles: Hurdles
✓- S Confined Area Obstacles: Stairs/Landings
✓- V Maneuvering Tasks: Towing Sleds
P Maneuvering Tasks: Towing Trailers
✓- P Confined Space Terrains: Vertical Insertion/ Retrieval Stack with Drops

SENSING
✓- S Video: Acuity & Field of View
✓- B Video: Pan/Tilt/Zoom Tasks
✓- V Audio: One/Two Way Comms
V Range: Localization/Mapping: Maze Fiducials
P Range: Localization/Mapping: Tunnel Maze

RADIO COMMUNICATIONS
✓- B Line-of-Sight Range
✓- B Non Line-of-Sight Range
✓- P Structure Penetration
P Urban Canyon Environments

MANIPULATION
V Confined Area Inspection
V Confined Area Grasping
V Door Opening and Traversal

HUMAN-SYSTEMS INTERACTION
✓- B Random Maze Navigation
✓- B Random Maze Search
✓- V Underbody Search
V Checklist Items

ENVIRONMENT
B Decontamination/Washdown Practice
✓- P Water Fording

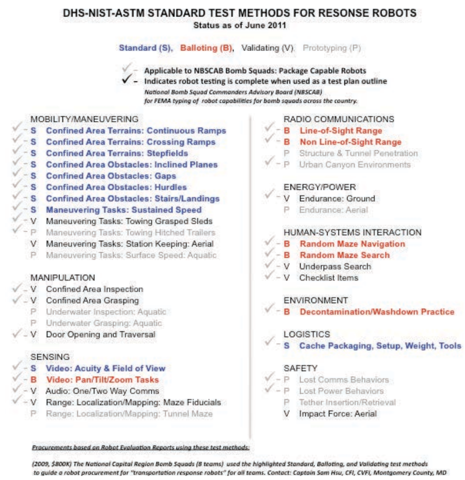
SAFETY
P Lost Comms Behaviors
P Lost Power Behaviors
✓- P Throw/Drop Distance

OPERATIONAL TASKS AT NIST NIKE SITE
✓- Underground Silo Recon & Extraction
✓- VBIED Trailer Underbody Recon
Explosives Cache Recon (radio neutral)
Block Building Stairwell Recon & Extraction

FIGURE 16: A) The roster of robots considered by JIEDDO to be ultra light weight reconnaissance robots. B) The list of 27 test methods selected by JIEDDO to specify their robot procurement. They initially purchased roughly $32M worth of robots to support field testing in Afghanistan.

METROTECH

Suspicious Package Missions Including Transportation (Buses/Trains)



DHS-NIST-ASTM STANDARD TEST METHODS FOR RESONSE ROBOTS
Status as of June 2011

Standard (S), Balloting (B), Validating (V), Prototyping (P)

✓ – Applicable to NBSCAB Bomb Squads: Package Capable Robots
✓ – Indicates robot testing is complete when used as a test plan outline
National Bomb Squad Commanders Advisory Board (NBSCAB) for FEMA typing of robot capabilities for bomb squads across the country.

MOBILITY/MANEUVERING
- ✓ – S Confined Area Terrains: Continuous Ramps
- ✓ – S Confined Area Terrains: Crossing Ramps
- ✓ – S Confined Area Terrains: Stepfields
- ✓ – S Confined Area Obstacles: Inclined Planes
- ✓ – S Confined Area Obstacles: Gaps
- ✓ – S Confined Area Obstacles: Hurdles
- ✓ – S Confined Area Obstacles: Stairs/Landings
- ✓ – S Maneuvering Tasks: Sustained Speed
- ✓ – V Maneuvering Tasks: Towing Grasped Sleds
- ✓ – P Maneuvering Tasks: Towing Hitched Trailers
- V Maneuvering Tasks: Station Keeping: Aerial
- P Maneuvering Tasks: Surface Speed: Aquatic

MANIPULATION
- ✓ – V Confined Area Inspection
- ✓ – V Confined Area Grasping
- P Underwater Inspection: Aquatic
- P Underwater Grasping: Aquatic
- ✓ – V Door Opening and Traversal

SENSING
- ✓ – S Video: Acuity & Field of View
- ✓ – B Video: Pan/Tilt/Zoom Tasks
- ✓ – V Audio: One/Two Way Comms
- ✓ – V Range: Localization/Mapping: Maze Fiducials
- P Range: Localization/Mapping: Tunnel Maze

RADIO COMMUNICATIONS
- ✓ – B Line-of-Sight Range
- ✓ – B Non Line-of-Sight Range
- P Structure & Tunnel Penetration
- ✓ – P Urban Canyon Environments

ENERGY/POWER
- ✓ – V Endurance: Ground
- P Endurance: Aerial

HUMAN-SYSTEMS INTERACTION
- ✓ – B Random Maze Navigation
- ✓ – B Random Maze Search
- ✓ – V Underpass Search
- ✓ – V Checklist Items

ENVIRONMENT
- ✓ – B Decontamination/Washdown Practice

LOGISTICS
- ✓ – S Cache Packaging, Setup, Weight, Tools

SAFETY
- ✓ – P Lost Comms Behaviors
- ✓ – P Lost Power Behaviors
- P Tether Insertion/Retrieval
- V Impact Force: Aerial

Procurements based on Robot Evaluation Reports using these test methods:

(2009, $800K) The National Capital Region Bomb Squads (8 teams) used the highlighted Standard, Balloting, and Validating test methods to guide a robot procurement for "transportation response robots" for all teams. Contact: Captain Sam Hsu, CFI, CVFI, Montgomery County, MD

FIGURE 17: A) The roster of robots the National Capital Region Bomb Squad Working Group (Metrotech) of 13 federal, state, and local teams considered to deal with suspicious packages on transportation systems. B) The list of 29 test methods selected used to specify their purchase.

Specifying Robots Based On Actual Capabilities

This is where the Response Robot Capabilities Compendium, the database containing all the robot data, enables active filtering on results for chosen test methods. This filtering activity or the associated bar charts can help inform and shape initial expectations regarding the trade-offs in capabilities that will ultimately be necessary. In either case, the user must begin to specify the desired performance levels for each test method while seeing exactly how many robots are being filtered out by those decisions. In the Compendium database, the capability values available to choose for filtering are actual values achieved by tested robots. So if you see several robots bunched up near a reasonably desirable capability level, it is a good policy to adopt that capability level (if you can live with it in the context of your intended mission) and see where the other test methods begin to separate robots. There are so many conflicting design constraints in these robots that the first 5 or so prioritized test methods will likely make the differences clear and decisions easier.

This process can be accomplished using only the bar charts, of course, but it can be a bit harder to play out too many different combinations of capabilities. However, if you're simply trying to down-select to three robots to conduct operational scenarios, the bar charts are quite easy to use. They typically present all related test methods on a single chart with colored bars indicating different levels of performance. So it is quite easy to see which robots are the most mobile, which are the fastest, which abstained the hardest terrain (for whatever reason!). The bar charts also provide immediate feedback if your expectations were initially too lofty regarding the overall class of robot capabilities.

Acceptance Testing Upon Robot Delivery

Once a purchase is complete, taking delivery is the next step and your acceptance testing plan is already in place. You can simply have the "expert" operator provided by the robot developer repeat the set of test methods specified for the purchase (possibly in conjunction with your training even). Based on how well the developer supplied "expert" operator performed the original tests, since those were the numbers you used to specify and defend your purchase, the delivered robot should score at least 80% of the original capability to ensure their quality control. The percentage can vary as appropriate given the particular circumstances of the purchase, especially if the configuration purchased is a bit different than the original configuration tested.

Comparing System Costs

The process described in this document focuses on using standard test methods to quantify key capabilities of response robots. However, It does not address the all-important issue of system cost, which is arguably the most important number in the process. Recent procurements using this process introduced a helpful way to inject the cost of the system into consideration as one more data point to consider. They asked the robot developers to provide quotes for packages of robotic systems and other available products, equipment, and features totaling a specified price point. Some guidance could be provided regarding the types of deployment equipment you might be interested in, but otherwise this worked well to get a tangible comparison of costs and capabilities into context.

For example, if the specified package price were set to $100,000, some robot developers might be able to deliver two mobility bases, one manipulator payload, a fiber-optic tether, and shipping boxes. Another robot developer might only be able to deliver the basic robot chassis along with a minimalist operator interface and low power radio. Other developers will provide their own combination of robots and equipment adding up to the specified package price. It may be more beneficial to consider setting several package prices that robot developers can strategize around and re-use, such as $25K, $50K, $100K, $200K and so on.

Cost is always a driving issue, and can become overwhelming very quickly with the myriad of options and features for sale. This approach at least enables side-by-side consideration of system cost across different robot developers and robot configurations, and puts the value of the package into context when considered along with the measured capabilities. (Do you really need the expensive radio comms addition, or does the basic radio suffice allowing you to afford the manipulator option that performed well?)

Other considerations also play key roles in any prospective robot procurement. Do you already own a robot? If so, you likely have a cache of parts on hand for maintenance and failures and spent a substantial amount of time learning how to replace components yourself. Do you like your robot's operator interface? Either way you may not want to move to an unfamiliar operator interface figuring you would have to learn the new system, maintain the old system, and keep trained up on both! These are all considerations when thinking about purchasing a new robot. This process can at least help identify and understand the benefits of a good interface and how it can improve remote operator performance. It can also show how some "robotics" in they system, how some basic onboard intelligence within the robot, can really improve the effectiveness of a remote operator. At least this process begins to align expectations and inform decision making before purchasing a new robot, so should lessen the potential for disappointment and increase the probability that a new robot will integrate into your cache well enough to improve your effectiveness

Training Operators with Measures of Proficiency

Beyond purchases, standard test methods can help measure operator proficiency and support training for deployments. During the standardization process, each robot is tested using an "expert" operator provided by the robot developer to capture the best possible performance. Using that data, any individual using the same robot configuration can measure themselves relative to that expert capability. Given that robot operation is a very perishable skill, maintaining and improving skills is essential to be effective.

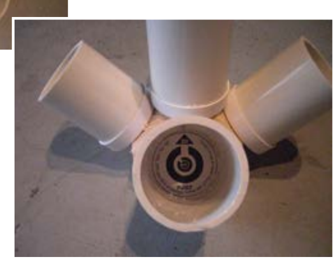
Repeated testing in standard test apparatuses and associated operational tasks can provide essential training and practice opportunities prior to any deployment. If a robot and remote operator cannot complete the specified set of standard test methods that represent an operational task, it is not likely the robot and remote operator could perform that task during deployment. Conversely, if a robot and remote operator can practice and reliably demonstrate statistically significant performance across the set of representative test methods, there is a much greater likelihood that the robot and remote operator can perform the task during deployment.

Although the repeatability and reproducibility of operational tasks is limited, embedding standard test apparatuses into the scenarios can help measure robot and remote operator capabilities. Each apparatus essentially contains its own statistical significance in a way, by providing repeatable tasks within the apparatus itself. Individual user organizations or small groups can provide structured training programs and operator proficiency testing to ensure their robot operators are at some specified level of capability. For a perishable skill such as robot operation, this could be a periodic process.

Training Sets of Standard Test Methods

Standard test methods can measure baseline capabilities specifically for operators. This is a newly emerging focus for this project. Such measures of operator proficiency can be used to inspire improvement and track progress. For example, robot developers typically provide training for operators when delivering a new robot. The first several hours of training can be structured to use standard apparatuses to introduce the novice operator to system functions, providing ongoing practice tasks lasting long after the training is over. Several standard test methods provide simple yet essential tasks that could provide a structured curriculum. Robot developers can use a curriculum like this to remotely monitor progress of novice operators, identify and address early frustrations, isolate particular issues in the operator interface, and recognize excellence in new operators or improved interfaces. At least robot developers can begin to measure the learning curve associated with their system. A sample curriculum could be:

- Control the camera to look around:
  Sensors: Pan/Tilt/Zoom Tasks
- Control the audio system to listen and talk:
  Sensors: Audio Acuity (2-Way)
- Drive to follow a path:
  Maneuvering: Line Following
- Drive to avoid obstacles:
  Maneuvering: Decreasing Slalom
- Drive on modest terrain:
  Mobility: Crossing Ramp Terrain
- Drive and control camera at the same time:



Maneuvering: Maze Navigation with Complex Terrain

- Inspection tasks (small robots without a manipulator):
  Human-Robot Interaction: Underbody Inspection

- Inspection tasks (robots with a manipulator):
  Manipulation: Dexterity: Inspection: Surrounding Ground Locations
  Manipulation: Dexterity: Inspection: Elevated Surfaces

- Grasping tasks (robots with a manipulator):
  Manipulation: Strength: Surrounding Ground Locations
  Manipulation: Dexterity: Retrieval/Insertion Tasks

Other options: advanced mobility terrains and obstacles like stairs, towing tasks, tool deployments .

Ship in one container

Set up in 4 man-hours

Modest lab footprint for multiple test methods

(12 test methods shown)

Clear/easy measures of performance

(see blue capped targets as successful tasks)

FIGURE 18: A sample set of training test methods in a box can contain introductory tasks or rather advanced tasks targetted for specific responder communities. For example, package capable bomb squad robots or C-VBIED tasks.

For responders who already own a robot, more focused training sets could address specific responder missions and encourage a productive competition among and across teams to improve overall capabilities. For example, bomb techs all using the same robot could compete using the training suites outlined below. These training suites break down basic missions into 10 different test methods. Each test method is time limited to 10 minutes and are conducted sequentially in any order (presumably easiest to hardest for the given operator). In 100 minutes of mission time, with 10 different task scores, you would have a pretty clear understanding of the your proficiency. You could then, of course, compare that to the “expert” operator provided by the robot developer working on the same make/model robot.

ULTRA-LIGHTWEIGHT RECONNAISSANCE ROBOTS

- Pan-Tilt-Zoom Tasks
- Slalom
- Maze Navigation
- Underbody Inspection
- Crossing Ramps
- Symmetric Stepfields
- Vertical Insertion/Retrieval Stack with Drops
- Hurdles with Culvert
- Stairs (Descending)
- Towing Grasped/Hitched Sleds

PACKAGE CAPABLE ROBOTS

- Pan-Tilt-Zoom Tasks
- Slalom
- Maze Navigation
- Crossing Ramps with Grasped Payload
- Inspection (Wall, Elevated Shelves)
- Insertion/Retrieval (Wall, Elevated Shelves)
- Disruptor Aiming (Nodal, Cylindrical)
- Stairs
- Grasp, Traverse, and Place on Inclined Plane
- Towing Grasped/Hitched Sleds

C-VBIED CAPABLE ROBOTS

- Pan-Tilt-Zoom Tasks
- Slalom
- Crossing Ramps with Grasped Payload
- Inspection (Vehicle Cab)
- Disruptor Aiming (Nodal, Horizontal Cylindrical)
- Lifting (Vehicle Cargo Bay)
- Door Opening and Traversal
- Breaking/Boring
- Strap Cutting
- Towing Hitched Trailers

Response Robot Evaluation Exercises

NIST has hosted eight Response Robot Evaluation Exercises at FEMA responder training facilities around the country including “Disaster City” in College Station, TX. At these events, more than 120 robots have practiced and collected capabilities data within the suite of test methods prior to deploying with responders into the range of operational training scenarios available on site. Participating robot developers, responders, and test administrations have validated more than 40 standard, draft standard, and prototype test methods at these events using our consensus standards development process. Several smaller testing events have also been hosted to focus on particular sub-suites of test methods. For example, in conjunction with the Bomb Squad Commanders Conference and Training Symposium in Socorro, New Mexico (2012) hundreds of bomb squad commanders were introduced to newly developed draft standard test methods for counter vehicle-borne improvised explosive devices (C-VBIED). These testing events give particular groups of responders insight into recent advances within the robotics community while capturing essential test method "repeatability" and "reproducibility" data for the standards development process. Feedback from these user communities set the agenda for the project and identify priorities.

FIGURE 19: Examples of using standard test methods as incrementally complex challenges on the way to more operationally difficult environments like the rubble pile at Disaster City. A robot should be able to complete all the standard terrains to statistical significance prior to trying the rubble. There are lessons to learn in negotiating the Stepfield terrain (shown in the middle distance above), like how to right the robot when it rolls over!

In addition to DHS, other organizations have sponsored robot data collection events using standard and draft standard test methods, which were essential to provide "repeatability" and "reproducibility" required for the standards development process and helped establish the available range of capabilities for specific classes of robots. Such sponsors included robot development programs such as the U.S. Army Research Laboratory (ARL) Robotics Collaborative Technology Alliance (RCTA); robot procurements for the U.S. Department of Defense Joint Improvised Explosive Device Defeat Organization (JIEDDO); NIST's Office of Law Enforcement Standards (OLES); and robot competitions such as MAGIC 2010 and the annual RoboCupRescue Robot League which use standard and prototype test methods as challenge tasks to inspire and measure new technical developments.

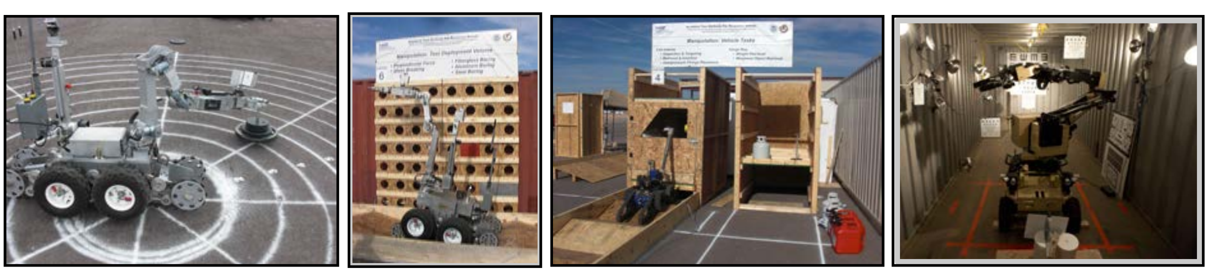


FIGURE 20: A few test methods for larger robots practicing counter vehicle-borne improvised explosive device tasks: A) heavy lifting of objects around the robot, B) tool deployment for cutting access, C) inspecting a vehicle cab and cargo bay, and the D) Pan-Tilt-Zoom Tasks.

NIST Nike Site Robot Test Facility

NIST's Nike Site Robot Test Facility is located in Gaithersburg, MD, just off the main NIST campus. The facility is a national resource for industry, universities, and government agencies. It's mission is to facilitate manufacturing and deployment of advanced robotic systems through development of performance test methods, measurement capabilities, and standards, with requisite support for emerging sensors, intelligent behaviors, open-architecture controllers, and high-fidelity simulation tools.
The site hosts three dozen robot test methods in various stages of standardization: prototyping, validating, balloting, accepted standard. The site also hosts several operational tasks for robots to perform after quantifying their capabilities in the standard test methods. The operational tasks utilize the underground silos, structures, and vehicles placed around the site to support specific robot evaluation, practice, or operator training needs for various organizations. For example, the National Capital Region Bomb Squad Working Group made up of a dozen federal, state and local bomb squads use the site for periodic robot training and have used capabilities data captured in the test methods to guide robot purchasing decisions.

FIGURE 21: NIST Nike Site Robot Test Facility in Gaithersburg, MD.

Operational Scenarios with Embedded Standard Test Methods

Operational scenarios with embedded standard test methods provide an indication of robotic capabilities at a reasonable cost, and help operators practice and compare performance. They can provide essential training and practice opportunities prior to deployment. Any operational task should be tested to statistical significance, although the repeatability and reproducibility of such tasks are necessarily limited due to the environmental complexity. The test method apparatuses contain multiple sub-tasks which act as repetitions for the purposes of establishing statistical significance. The operational tasks shown below were set up to provide measurable performance in a variety of challenging environments. They emphasize reconnaissance tasks with measures of coverage of the scenario and identification of embedded eye charts, hazmat labels, and objects of interest such as simulated pipe bombs, etc.

FIGURE 22: Examples of easily embedded standard test methods for operational scenarios and robot competitions. A) A combination visual acuity test chart including some formal measures of acuity, hazmat labels, and QR codes of various sizes to test visual acuity of autonomous robotic systems. B) Manipulator inspection tasks using planar and cylindrical surfaces. The cylinders also act as mapping fiducials. C and D: The so-called "pipe star" nodal apparatus that supports inspection, disruptor aiming, object insertion, and object retrieval tasks.

When robotic systems have limitations that preclude completion of the operational tasks shown below, for example due to radio drop-out, or too steep stairs, etc., these scenarios still provide examples of how embedded standard test methods can help get comparable data from otherwise very qualitative demonstrations. When such scenarios are not strictly administered, when they are treated more as demonstrations, they do not appear in the Response Robot Capabilities Compendium or in the bar charts.

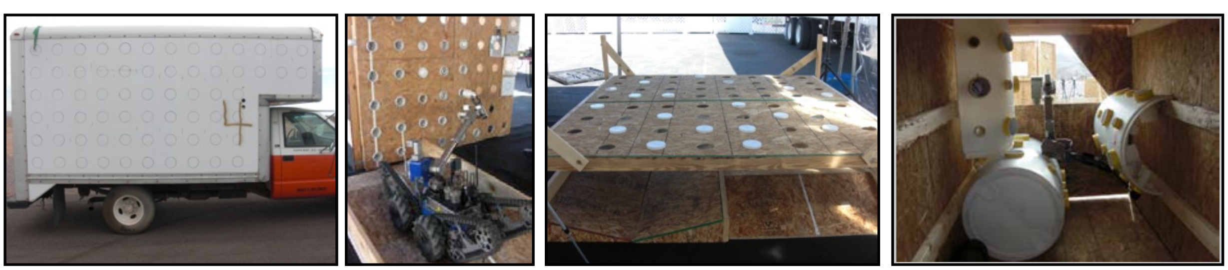


FIGURE 23: The V*ehicle-Borne Improvised Explosive Device* task included a *Personnel-Borne Improvised Explosive Device* in the vehicle cab. In this case, the major test methods were replicated so as to be repeatable. After demonstrating all the necessary competiencies the robot would be allowed to perform similar tasks on the real vehicle.

FIGURE 24: **The *Metrobus Package Removal/Disruption* task included negotiating and inspecting a commuter bus containing suspicious packages (briefcase with handle, backpack, and box) of different weights (2 stripes = 5kg/10lbs and 4 stripes = 10kg/20lbs), in varying locations (front, middle, and rear) requiring different approaches (forward, side, and diagonal respectively), and at varying elevations (under seat, on seat, overhead shelf). The robot has to enter the bus from a remote operator location (150m/500ft), remove from the bus the heaviest package possible of each type in each location, and deliver to a further remote site for disruption. If a particular type of package cannot be removed from a particular location, the robot should disrupt it in place.**

FIGURE 25: The *Underground Silo Reconnaissance* task included negotiating and inspecting a defunct underground Nike missile silo with multiple complex rooms, uneven flooring obstacles, and extensive radio attenuation. Entry could be gained by A) descending a stairway with 40 degree incline and 9 inch risers or B) vertical insertion through a manhole ladder. C) Barrels with four directional targets to identify were placed throughout the environment requiring access to confined areas to identify targets from all directions. D) A variety of flooring complexity divided the space. Two different operator stand-off positions were available depending on radio communications capability of the robot: 10m (top of the stairs), 100m non-line-of-sight.

FIGURE 26: The *Trailer Underbody Reconnaissance* task included nine tractor trailers parked side by side. Ten visual acuity targets and ten objects of interest were placed on the underside of the trailers, one under each trailer with two under one particular trailer. Three operator stand-off positions were available depending on radio communications capability of the robot: 10m line-of-site, 75m line-of-sight, and 100m non-line-of-sight.

FIGURE 27: The *Building Stairwell Reconnaissance* demonstration scenario included a cement block stairwell structure which covers the entrance to a defunct underground Nike missile silo. A) Entry was gained either through an open ground-level window, an open second-level window, or B) through an open roof skylight accessible by throwing the robot onto the roof from the ground followed by a 12 ft drop onto the stairwell. C) The structure contained steel stairs with landings between three floors (two above ground and one below ground). D) Barrels with four directional targets to identify were placed on each level requiring access to confined areas to identify targets from all directions. One table-top location on the second level also had targets to identify (shown in B). Two operator stand-off positions were available depending on radio communications capability of the robot: 10m (outside the window) and 100m non-line-of-sight.

FIGURE 28: The *Explosives/Drug Lab Reconnaissance* demonstration scenario included a two-room bungalow set up to represent an explosives or drug lab. A/B) Entry was gained through an open ground-level window. Barrels with four directional targets to identify were placed in each room requiring access to confined areas to identify targets from all directions. C) It contained table-top objects of interest, D) plastic bags on the floor, trip wires at doorways, etc. Two operator stand-off positions were available depending on radio communications capability of the robot: 10m (outside the window) and 100m non-line-of-sight.

FIGURE 29: The *Roadside IED Identification and Blow In Place* demonstration scenario included the following: A) Start point behind a burned out van, 100m from the suspected IED location seen just beyond the bungalow. B-G) Suspicious rubble piles on either side of the road to be cleared, each contained hazmat labels to identify. Successful identification of hazmat labels without a co-located IED were considered cleared. H) The IED was a simulated artillery shell in a shallow hole covered with trash bags, shown with the disruptor in position to blow in place.

FIGURE 30: Search and map an intact building containing random maze test apparatus within some rooms. Locate any signs of life (visible shape or motion, audible, thermal) and any suspicious packages, hazardous materials, placards, or labels. All objects of interest should be located on a 2-D map of the area with recognizable navigation features.

FIGURE 31: Search and map a dwelling with a basement. Locate any signs of life (visible shape or motion, audible, thermal) and any suspicious packages, hazardous materials, placards, or labels. All objects of interest should be located on a 2-D map of the area with recognizable navigation features.

FIGURE 32: Search and map a semi-collapsed structure. Locate any signs of life (visible shape or motion, audible, thermal) and any suspicious packages, hazardous materials, placards, or labels. All objects of interest should be located on a 2-D map of the area with recognizable navigation features.

FIGURE 33: Conduct wide area survey of a hazmat/passenger train derailment with operational stand-off.

FIGURE 34: Conduct confined space and rubble pile searches.

Robots Tested Comprehensively

More than 100 robots have been tested within the standard test methods in one form or another, and with varying degrees of comprehensiveness. This document focuses only on robots tested across a wide variety of standard and draft standard test methods during procurement actions, so that the trade-off assessments can be made regarding individual robot configuration capabilities.

Robots Weighing: 0 - 20 kg (1 - 44 lbs)

Eight robot configurations from seven robot developers were tested. These robots were considered within the class of ultra lightweight recon robots weighing less than approximately 10 kg (22 lbs) so they could be thrown over walls, through windows, etc., to perform rapid reconnaissance within a variety of operational environments. The robots shown below are in order of increasing weight from 0.5kg (1.1lb) to 7.9kg (17.4lb):

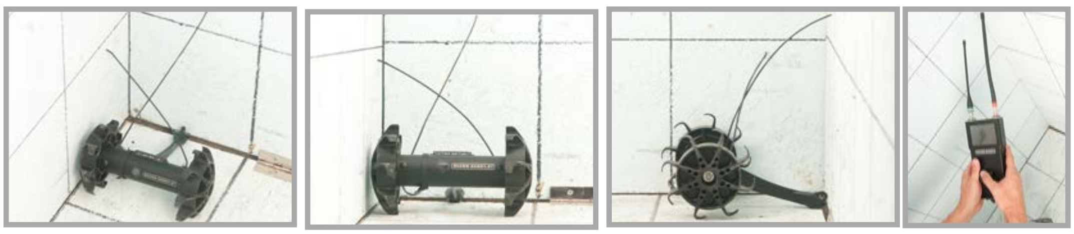


Figure 35) Recon Scout XT, Recon Robotics, Inc., USA (0.5kg / 1.1lbs)

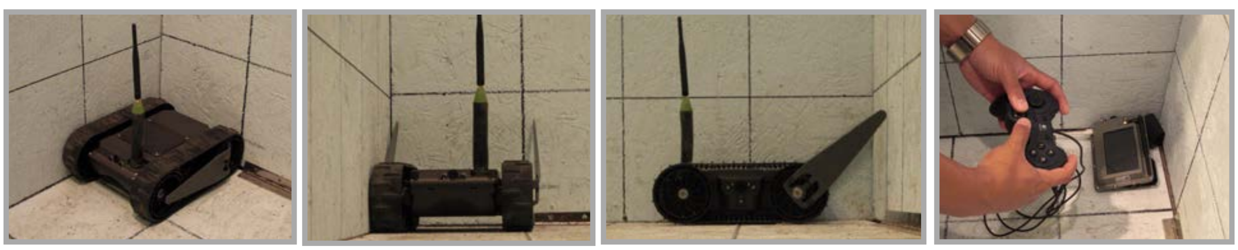


Figure 36) iRobot 110 FirstLook, iRobot Corp., USA (2.3kg / 5.0lbs)

Figure 37) Armadillo–Base–Wheels, MacroUSA Corp., USA (3.0kg / 6.6lbs)

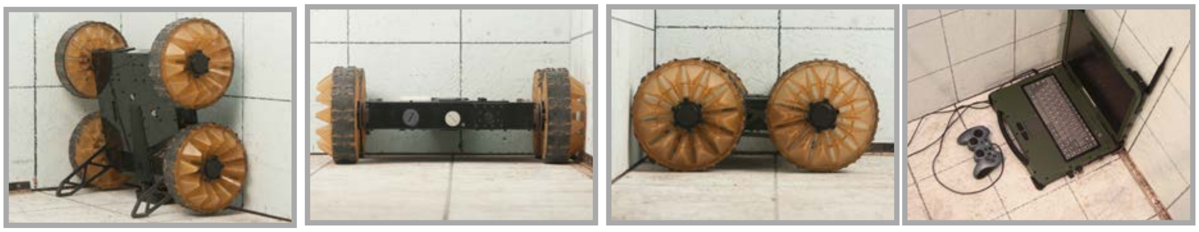


Figure 38) Sand Flea, Sandia National Laboratory and Boston Dynamics, USA (4.7kg / 10.4lbs)

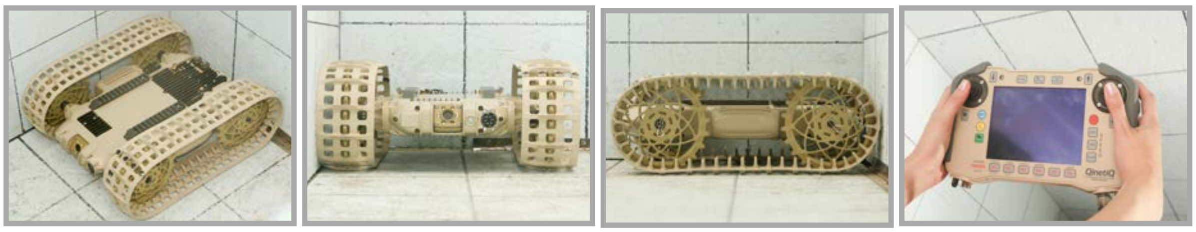


Figure 39) Dragon Runner DR-10 Base, QinetiQ North America, Inc., USA (4.7kg / 10.4lbs)

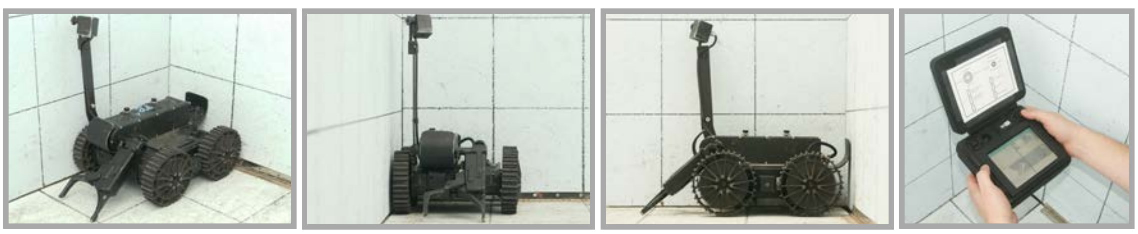


Figure 40) Armadillo–Manipulator–Wheels, MacroUSA Corp. and RE^2, Inc., USA (4.8kg /10.6lbs)

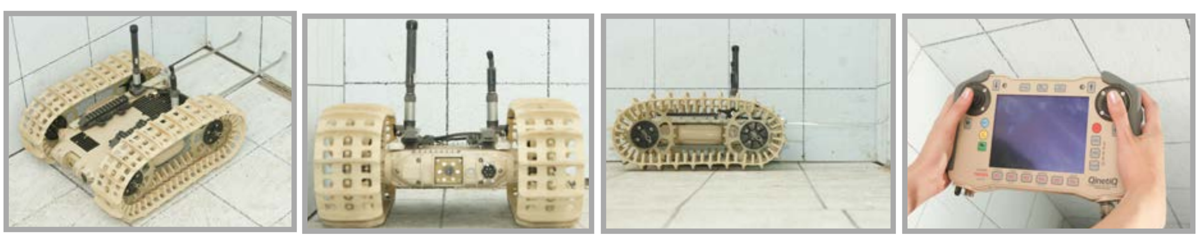


Figure 41) Dragon Runner DR-10 with Stair Climber Kit (SCK) and External Antennas, QinetiQ North America, Inc., USA (4.8kg / 10.6lbs)

Figure 42) Micro Tactical Ground Robot (MTGR), Roboteam Ltd., Israel (6.5kg / 14.3lbs)

Figure 43) Pointman LRV, Applied Research Associates, Inc., USA (7.9kg / 17.4lbs)

Robots Weighing: 20 - 100 kg (44 - 220 lbs)

Figure 44) The roster of robots considered for purchase by the National Capital Region Bomb Squad Working Group of federal, state, and local bomb squads in the Washington, D.C. area.

Resulting Capabilities Compendium

Comparison and Trade-Off Tool (software interface)

The Response Robot Capabilities Compendium contains performance data from all robots subjected to comprehensive testing within the DHS-NIST-ASTM International Standard Test Methods for Response Robots. Currently, NIST has conducted all the testing as part of the standards development process. However, recently opened test facilities in Kobe, Japan; Koblenz, Germany; and Southwest Research Institute in San Antonio, TX will soon start contributing additional robot performance data. Given the myriad combinations of robot sizes, shapes, weights, and capabilities, a software interface into the database is the best way to understand the implications of specifying certain attributes or performance thresholds. This interface allows the user to see which robots have demonstrated statistically significant performance for their highest priority capabilities necessary to perform their intended mission. They can quickly see the effects of specifying too stringent a requirement in any particular capability or attribute as the number of robots that have successfully demonstrated the specified combination are filtered. Reducing a stringent threshold for even one requirement can bring several more robots into consideration. So users quickly learn the trade-offs involved and what the state of the science can deliver with regards to the combination of attributes and capabilities they have in mind.

This tools provides an initial survey of different classes of robots demonstrating statistically significant performance in particular combinations of capabilities. The filtered list of robots that results should then be evaluated in detail by referencing the graphical test forms side by side.

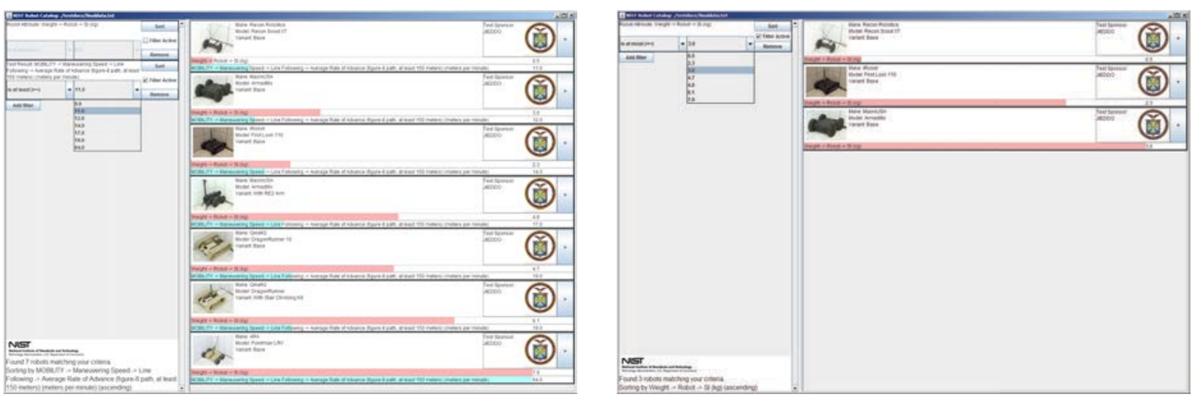


Figure 45) Screen captures of the filtering available in the Response Robot Capabilities Compendium.

Bar Charts

The graphical test forms associated with each test method provide an intuitive understanding of the robot's capabilities in order to facilitate side by side comparisons. However, there are dozens of test methods in the suite and users of the data benefit from comparisons across the entire class of robots. Bar charts shown below help identify Best-In-Class robots in specific test methods, and allow initial identification of trade-offs for particular robot configurations. Again, once a search is narrowed to several robots, a detailed study of the associated performance data forms is recommended.

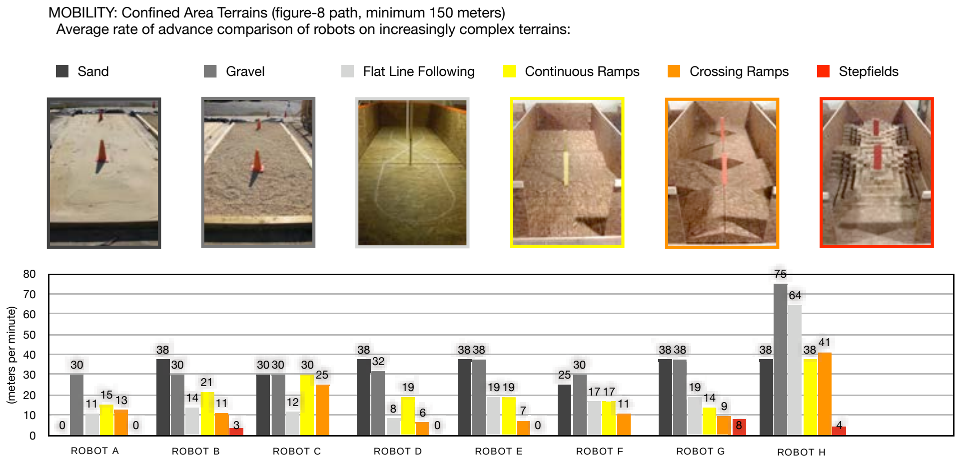


FIGURE 46: An example of bar charts generated from the performance data shows the comparison across a class of robots on increasingly complex terrains. Missing bars mean the robot abstained from the test method. Sponsors get charts with all the robot names identified. Robot developers get the same charts with only their robot name identified, providing a clear indication of their capabilities within the class.

Response Robot Capabilities Compendium: Comparison and Trade-Off Tool

Ultra Lightweight Recon Robot Class V2011.08-2012.02A

Rapid Evaluations Using DHS-NIST-ASTM International Standard Test Methods

Sofware Version: BETA Version 46+ INTERNAL USE ONLY DO NOT DISTRIBUTE

Data File: Ultra_Lightweight_Recon_Robot_Class_V2011.08-2012.02A.txt

The Department of Homeland Security sponsored the development of these standard test methods under an Interagency Agreement with the National Institute of Standards and Technology. Certain commercial equipment, instruments, or materials are identified in this document. Such identification does not imply recommendation or endorsement by the Department of Homeland Security or the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.

Robot+OCU Weight
Robot Weight
OCU Weight

	Photo	Make / Model / Configuration	Robot+OCU Weight	Robot Weight	OCU Weight	Data
Anon		MAKE: **Recon Robotics** MODEL: **Recon Scout XT** CONFIGURATION: **Base**	1.5 (kg)	0.5 (kg)	1.0 (kg)	Data Version: **v2011.08A** Facility: **NIST Nike Site** Location: **Gaithersburg, MD** Sponsor: **JIEDDO**
Anon		MAKE: **iRobot Corp.** MODEL: **FirstLook 110** CONFIGURATION: **Base**	3.1 (kg)	2.3 (kg)	0.8 (kg)	Data Version: **v2011.08B** Facility: **Disaster City 2011** Location: **College Station, TX** Sponsor: **JIEDDO**
Anon		MAKE: **MacroUSA** MODEL: **Armadillo** CONFIGURATION: **Base**	4.2 (kg)	2.8 (kg)	1.4 (kg)	Data Version: **v2011.08A** Facility: **NIST Nike Site** Location: **Gaithersburg, MD** Sponsor: **JIEDDO**
Anon		MAKE: **MacroUSA and RE2** MODEL: **Armadillo** CONFIGURATION: **Manipulator**	6.1 (kg)	4.7 (kg)	1.4 (kg)	Data Version: **v2011.08A** Facility: **NIST Nike Site** Location: **Gaithersburg, MD** Sponsor: **JIEDDO**
Anon		MAKE: **Applied Research Associates** MODEL: **Pointman-1000-005** CONFIGURATION: **Base**	10.8 (kg)	7.9 (kg)	2.9 (kg)	Data Version: **v2011.08B** Facility: **Disaster City 2011** Location: **College Station, TX** Sponsor: **JIEDDO**
Anon		MAKE: **QinetiQ North America** MODEL: **Dragon Runner DR-10** CONFIGURATION: **Base**	11.0 (kg)	4.7 (kg)	6.3 (kg)	Data Version: **v2011.08A** Facility: **Disaster City 2011** Location: **College Station, TX** Sponsor: **JIEDDO**
Anon		MAKE: **Roboteam** MODEL: **MTGR** CONFIGURATION: **Base**	11.1 (kg)	6.5 (kg)	4.6 (kg)	Data Version: **v2012.02A** Facility: **NIST Nike Site** Location: **Gaithersburg, MD** Sponsor: **JIEDDO**
Anon		MAKE: **Sandia National Labs and Boston Dynamics** MODEL: **Sand Flea** CONFIGURATION: **Base**	11.3 (kg)	4.7 (kg)	6.6 (kg)	Data Version: **v2011.08A** Facility: **NIST Nike Site** Location: **Gaithersburg, MD** Sponsor: **JIEDDO**
Anon		MAKE: **QinetiQ** MODEL: **Dragon Runner DR-10** CONFIGURATION: **Stair Climbing Kit**	12.4 (kg)	6.1 (kg)	6.3 (kg)	Data Version: **v2011.08A** Facility: **Disaster City 2011** Location: **College Station, TX** Sponsor: **JIEDDO**

No filters active.

Appendix A: Robot Capabilities Bar Charts

See Associated File:
DHS-NIST-ASTM International Standard Test Methods for Response Robots
Charts for ___________ Robot Class (v####.#)

Appendix B: Test Method Descriptions

See Associated File:
DHS-NIST-ASTM International Standard Test Methods for Response Robots
Test Method Descriptions (v####.#)

Guide for Evaluating, Purchasing, and Training with Response Robots
Using DHS-NIST-ASTM International Standard Test Methods

Guide for Evaluating, Purchasing, and Training with Response Robots
Using DHS-NIST-ASTM International Standard Test Methods