Technical Report of the Usability Testing of the National Use-of-Force Data Collection Portal

February 14, 2017

Since the shooting of Michael Brown in Ferguson, Missouri, in 2014, law enforcement uses of force have called public attention to the need for more information on these types of encounters. To provide a better understanding of the incidents of use of force by law enforcement, the Uniform Crime Reporting (UCR) Program is proposing a new data collection to include information on incidents where a use of force by a law enforcement officer as defined by the Law Enforcement Officers Killed and Assaulted (LEOKA) Program has led to the death or serious bodily injury of a person, as well as when a law enforcement officer discharges a firearm at or in the direction of a person. The definition of *serious bodily injury* will be based, in part, on 18 United States Code (USC) Section 2246 (4), to mean "bodily injury that involves a substantial risk of death, unconsciousness, protracted and obvious disfigurement, or protracted loss or impairment of the function of a bodily member, organ, or mental faculty."

Goal of the Proposed Data Collection on Law Enforcement Use of Force

The goal of the FBI's data collection on law enforcement officer use of force is to produce a national picture of the trends and characteristics of use of force by a law enforcement officer, as defined by the LEOKA Program, to the FBI. The collection and reporting would include use of force that results in the death or serious bodily injury of a person, as well as when a law enforcement officer discharges a firearm at or in the direction of a person. The data collected by the UCR Program would include information on circumstances surrounding the incident itself, the subjects, and the officers. The data collection would focus on information that is readily known and obtainable by law enforcement with the initial investigation following an incident rather than any assessment of whether the officer acted lawfully or within the bounds of department policies. Publications and releases from the data collection will provide for the enumeration of fatalities, nonfatal encounters that result in serious bodily injury, and firearm discharges by law enforcement. In addition, targeted analyses could potentially identify those law enforcement agencies with "best practices" in comparison with their peers as an option for further study.

This FBI data collection will facilitate important conversations with communities regarding law enforcement actions in relation to decisions to use force and works in concert with recommendations from the President's Task Force on 21st Century Policing. Given a growing desire among law enforcement organizations to increase their own transparency and embrace principles of procedural justice, this collection will expand the measure to a broader scope of incidents of use of force to include nonfatal instances as well.

Overview of Comprehensive Testing Plan

The FBI acknowledges that managing the scope of this collection and providing good guidance will be a challenge. In order to manage this effort, the FBI completed a Comprehensive Testing Plan to outline a series of activities to help inform the FBI on the decisions impacting scope, content, and participation levels. This document was forwarded to the OMB on September 2, 2016, and sets forth the expected activities to occur both before data collection commences and at the onset of data collection with a pilot study.

The pre-testing activities occurring before data collection consists of three primary efforts all of which build upon each other for planning the pilot study. The first is a cognitive testing effort to further research some concepts connected to data elements included in the National Use-of-Force Data Collection at the request of the law enforcement community. The second activity during pre-testing is a canvass of state UCR Program managers and state CJIS System Officers to gather information on programmatic and technical capabilities of the states in anticipation of the launch of the National Use-of-Force Data Collection. The final pre-testing activity is a small-scale assessment of the usability of the data collection application.

The pre-testing activities will provide critical information that will allow for the FBI to finalize plans to conduct a pilot study over the course of the first six months of data collection using a targeted group of law enforcement agencies. The goal of the pilot study is to assess the interpretation of questions used in the National Use-of-Force Data Collection and any guidance or instructions included in the data collection. This assessment will be based on a comparison of the original law enforcement record to the submitted responses to the questionnaire for pilot agencies. In addition to the record comparison, an on-site review of records for a sample of agencies will be conducted to assess the extent of nonresponse for in-scope incidents for participating agencies. The pilot study provides the best path to assess the data collection in the context of complex law enforcement decisions.

Background Research for Usability of the Portal

The FBI's approach to questionnaire design and usability testing is a multistage, iterative approach commonly associated with agile development. Final usability testing was conducted as a part of Operational Evaluation testing in January and February 2017. In addition to Operational Evaluation testing, FBI conducted thirteen separate demonstrations of the application in front of 111 individuals. These demonstrations included the same users on certain occasions, which allowed the development team the opportunity to receive feedback on adjustments and modifications.

A demonstration was conducted after each development sprint for a total of six. The audience for these demonstrations were the Product Owner and at least one representative each from Contracts, Configuration Management, Testing, Crime Data Modernization, the Law Enforcement Enterprise Portal (LEEP) Program Office, and Security each representing particular expertise with web application development. In addition, audience members represented the sworn and civilian law enforcement community and provided feedback on the intuitive nature and ease of use of the application. The smallest number of witnesses for a sprint demonstration was 11, and the most was 22. Each sprint

demonstration consisted of a walk-through of the user stories that had been agreed to for that particular sprint, a demonstration of the applications abilities with particular emphasis on the new capabilities related to the stories, a discussion regarding whether or not the new capabilities fulfilled the agreed-to stories, and an agreement as to what stories were completed and what stories (if any) still needed completed to address the usability of the application.

Of the 13 total demonstrations, there were two internal demonstrations (one after the third sprint and one after the sixth) conducted for the project stakeholders. These demonstrations were conducted for various unit and section chiefs throughout the CJIS Division as well as the Deputy Assistant Directors of the Operational Programs Branch and the Information Services Branch. The format in both cases was a PowerPoint slideshow defining the system and relaying the approved stories followed by a live demonstration of the application. The rest of the time was spent taking feedback from the attendees for possible incorporation into later iterations of the product. Average attendance was 22.

Recently, the FBI was able to conduct a series of five demonstrations with external stakeholders. A demonstration was conducted for the FBI Inspection Division at FBI Headquarters to get them familiar with the upcoming application. The demonstration was witnessed by an audience of executives to include four FBI Special Agents and subject matter experts at FBI Headquarters. There was an active give-and-take session throughout the demonstration, and several action items were captured to investigate for possible inclusion in future releases of the application. A similar demonstration was later held at FBI Headquarters for representatives from the Drug Enforcement Agency, Bureau of Alcohol, Tobacco, Firearms, and Explosives, and the United States Marshals Service. Finally, three demonstrations were held via Skype with state UCR Program managers and state criminal justice information services managers. In all five demonstrations, attendees had the opportunity to provide feedback and ask questions about the way the application was designed to function.

Participant Selection

Volunteers were solicited from the CJIS Division and included FBI Special Agents assigned to the CJIS Division, the FBI Police force assigned to the CJIS Division, and various record clerks within the CJIS Division (as record clerks will most assuredly be the individuals inputting incident information, especially in larger law enforcement organizations).

Methodology

The Operational Evaluation consisted of three testing sessions, each one hour in length. Two sessions were conducted in one of classrooms at the FBI CJIS Division facility on the same day, while a third testing session was conducted approximately one month later. The first session consisted of nine participants who each were assigned two scenarios to follow. *Scenario A* consists of placing a use-of-force incident into the system for review, and *Scenario B* consists of creating and submitting a Zero Report. A Zero Report is simply a record submitted by an agency stating that they did not have any use-of-force incidents that month. It is expected to be the most common of all types of submissions.

The second session consisted of fourteen participants who were assigned two scenarios each. Scenarios

were a mixture of the earlier *Scenarios A* and *B*, as well as new *Scenarios C* and *D*. *Scenario C* consists of going into the system, locating the incidents created in *Scenario A*, and approving those incidents for submission. *Scenario D* was the review of a Zero Report. The nine participants for the third session completed all four tasks. Each volunteer was asked to complete a System Usability Scale (SUS) questionnaire[1]. The SUS was developed by John Brooke to gauge the usability of a variety of systems— including web applications such as the National Use-of-Force Data Collection Portal.

At each computer, a set of instructions about the data collection questions and screenshots that demonstrate the process of creating an incident, creating a zero report, reviewing an incident, and reviewing a zero report was provided. This was to replicate a similar environment where a participant in the data collection may be asked to begin the data collection process without any prior training and would rely only upon the instructional materials available from the FBI.

Evaluation/Analysis

The sessions will be evaluated via the SUS.

The SUS will consist of ten responses:

1. I think that I would like to use this system frequently.

2. I found the system unnecessarily complex.

3. I thought the system was easy to use.

4. I think that I would need the support of a technical person to be able to use this system.

5. I found the various functions in this system were well integrated.

6. I thought there was too much inconsistency in this system.

7. I would imagine that most people would learn to use this system very quickly.

8. I found the system very cumbersome to use.

9. I felt very confident using the system.

10. I needed to learn a lot of things before I could get going with this system.

---

1 Information available at <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> (accessed on December 15, 2016).

The SUS will use the following response format:

Strongly Disagree 1	2	3	4	Strongly Agree 5
○	○	○	○	○

The SUS will be scored as follows:

- For odd items: subtract one from the user response.
- For even-numbered items: subtract the user responses from 5
- This scales all values from 0 to 4 (with four being the most positive response).
- Add up the converted responses for each user and multiply that total by 2.5. This converts the range of possible values from 0 to 100 instead of from 0 to 40.

The FBI will then look at the measures of central tendency of each question, as well as the overall total of each user. Based upon an analysis of the scores and other qualitative information from the participants, the FBI should be able to determine what areas require more effort and which ones can be considered "usable." Any score over 68 is considered to be above average. Beyond the SUS, the participants were also asked for general comments and impressions and to provide a name and contact information if they were willing to discuss their results with the application development team.

Results

Based upon the 28 participants providing complete responses in the Operational Evaluation, the average score was 79.464. This score indicates that participants generally found the application to be easy to use. The ratings of four participants resulted in individual scores below 68.

Table 1 provides a breakdown of median values for each question. For the positive statements about the application (odd-numbered questions), 50 percent of participants either agreed or strongly agreed. For the negative statements (even-numbered questions), 50 percent of participants either disagreed or strongly disagreed.

Table 1. Item analysis of the System Usability Scale

	Question	N Valid	Missing	Median[a]
1.	I think that I would like to use this system frequently.	32	0	4
2.	I found the system unnecessarily complex.	32	0	2
3.	I thought the system was easy to use.	32	0	4
4.	I think that I would need the support of a technical person to be able to use this system.	32	0	2
5.	I found the various functions in this system were well integrated.	32	0	4
6.	I thought there was too much inconsistency in this system.	32	0	1.5
7.	I would imagine that most people would learn to use this system very quickly.	32	0	4.5
8.	I found the system very cumbersome to use.	29	3	1
9.	I felt very confident using the system.	29	3	4
10.	I needed to learn a lot of things before I could get going with this system.	28	4	2

[a]1 = Strongly Disagree and 5 = Strongly Agree

Additional Findings from Observation of Usability Testing

In addition to the SUS findings, it was possible to observe each participants interaction with the application. One particular issue that was discovered in usability testing is that the participant often did not know "what to do next" after submitting an incident to his or her supervisor. The application was designed to present a pop-up message telling the participant that the incident was successfully submitted. However, it stayed on the last screen once the participant clicked, "okay" on the pop-up message. In general, participants found this confusing and do not seem to understand that the incident had been successfully submitted.

Participants in the Operational Evaluation were asked to provide any additional comments on the usability of the application. The majority of the responses referenced their confusion with the pop-up message after submission.

Discussion and Conclusion

Overall, the results of the usability testing were favorable, and no major problems were identified with the use of the application. The one issue that was discovered with the pop-up message after submission has been addressed and is undergoing testing as part of the second build release of the application. Once the application is available for users, the FBI UCR Program and the application development team will continue to monitor questions received through its Help Desk to address any undetected issues with the application.