MobileMark® 2014

An Overview of MobileMark® 2014

December 2014

Revision 1.0

Revision History:

      December 2014 – 1.0: Initial Release

About BAPCo

Business Applications Performance Corporation (BAPCo®) is a non-profit consortium with a charter to develop and distribute a set of objective performance benchmarks for personal computers based on popular computer applications and industry standard operating systems.

For more information about BAPCo® or a complete list of the current membership, see our website at http://www.bapco.com.

Table of Contents

1 Introduction

MobileMark® 2014 is the latest version of the premier performance qualified battery life benchmark for notebook and 2-in-1 computers. It features real world applications, updated workloads, support for Microsoft Windows® 7 32/64-bit, Microsoft Windows® 8 32/64 bit, and Microsoft Windows® 8.1 32/64-bit, and a new automatic system configuration manager.

MobileMark 2014 gives commercial and government IT decision makers, retailers, media, channel buyers, consultants, component designers, hardware designers, and manufacturers an objective, easy-to-use tool to evaluate the performance-qualified battery life of notebook and 2-in-1 computers across the wide range of activities that a user may encounter.

MobileMark 2014 is designed for those who want to:
- Evaluate and compare notebook and 2-in-1 computers for purchase consideration on the basis of performance-qualified battery life.
- Provide useful information to their audiences to assist in the evaluation and purchase of notebook and 2-in-1 computers.
- Evaluate notebook and 2-in-1 computers to better optimize the battery life and performance of the system.

Unlike benchmarks that only measure battery life, MobileMark 2014 measures battery life and performance simultaneously, showing how well a system design addresses the inherent tradeoffs between performance and power management.

Unlike synthetic benchmarks, which artificially drive components to peak capacity or attempt to deduce performance using a static simulation of application behavior, MobileMark 2014 uses real applications, real user workloads, and real data sets to accurately measure how overall system performance impacts user experience.

MobileMark 2014 builds upon BAPCo's 20-year history of building benchmarks to evaluate platform technologies. Benchmarks designed by BAPCo are the result of cooperative development between companies representing the breadth of the computing industry. They harness a consortium of knowledge to better reflect today's and tomorrow's emerging business trends.

This document describes the methodologies employed in the development of MobileMark 2014. For detailed instructions on how to install and run MobileMark products, please refer to the documentation provided on the installation media and the BAPCo web site (www.bapco.com).

2   BAPCo Development Process

BAPCo creates benchmarks in accordance with the BAPCo Development Process, a set of milestones and checkpoints collaboratively developed and agreed upon by the BAPCo membership.

Early in the process, prevailing business notebook and 2-in-1 computer usage models are identified and grouped into scenarios according to their fit within a workflow. Applications are selected for each usage model on the basis of market research and technical feasibility.

BAPCo members then join together with expert application users in development sessions to collaboratively develop a workload specification for each scenario, defining each user/computer interaction which is to be simulated by the benchmark.

The goal of the development sessions is to produce representative business application workloads for the benchmark. Each application workload consists of three elements: the input data set, the tasks performed on the input data set, and the generated output. An example of generated output would be an image generated through an iterative process of steps to create a desired appearance. These three elements of the workload are chosen to represent the workflow of a user skilled in each given application.

After the workload specifications are created at the development sessions, BAPCo developers implement the workloads according to those specifications while satisfying benchmarking constraints to ensure the stability of the benchmark, the consistency of results, and the feasibility of implementation and distribution of the benchmark.

2.1 Milestone Overview

The BAPCo development process is divided into six major phases (Initialization, Design and Planning, Implementation, Validation, Characterization and Launch). Each phase consists of a series of milestones, some of which may be worked on concurrently.

The membership must vote to close each milestone. Once all the milestones within a phase are complete, the membership must vote to exit the current phase and enter the next phase. BAPCo members work in a collaborative process where decisions regarding products are sometimes made by majority vote rather than unanimously.

The following is the list of the development phases and the corresponding milestones. Some of these milestones are explained in greater detail in the following sections, as noted in this list.

1. Initialization Phase
    a. Milestone 0 – Committee kickoff
    b. Milestone 1 – Benchmark market and customer analysis
    c. Milestone 2 – Product positioning and customer value proposition
    d. Milestone 3 – Preliminary marketing requirements document
    e. Milestone 4 – Final marketing requirements document
2. Design and Planning Phase
    a. Milestone 5 – Preliminary engineering requirements document
    b. Milestone 6 – Usage model selection (see section 2.2)
    c. Milestone 7 – Application selection (see section 2.3)
    d. Milestone 8 – Define member resource commitments
    e. Milestone 9 – Define development infrastructure
    f. Milestone 10 – Define scoring methodology (see section 2.4)
    g. Milestone 11 – Define application/API licensing requirements
    h. Milestone 12 – Plan and execute workload development sessions (see section 2.5, 2.6)
    i. Milestone 13 – Define product release criteria
    j. Milestone 14 – Select product name
3. Implementation Phase
    a. Milestone 15 – Create implementation schedule

b. Milestone 16 – Determine calibration system (see section 2.7)

c. Milestone 17 – Software development (see section 2.8)

4. Validation Phase

    a. Milestone 18 – Validation testing

    b. Milestone 19 – Define risk management plan

5. Characterization Phase

    a. Milestone 20 – Characterization testing

    b. Milestone 21 – Beta testing

6. Launch Phase

    a. Milestone 22 – Product pricing

    b. Milestone 23 – Pre-launch materials

    c. Milestone 24 – Release to manufacture vote and sign-off

    d. Milestone 25 – Duplicate and distribute media

    e. Milestone 26 – Post-launch materials

2.2 Usage Model/Scenario Selection

In Milestone 6 of the BAPCo Development Process, usage models are chosen for inclusion in a benchmark. For MobileMark 2014, BAPCo chose a wide variety of usage models in which the user experience is influenced by system performance.

BAPCo then groups related usage models into the following two scenario groups (for detailed descriptions of each scenario, please see section 2.6):

Office Productivity

The Office Productivity scenario models productivity usage including word processing, spreadsheet data manipulation, email creation/management and web browsing.

Media Creation

The Media Creation scenario models using digital photos and digital video to create, preview, and render advertisements for a fictional businesses.

2.3 Application Selection

In milestone 7, after the usage models have been collected into scenarios, applications are chosen for the scenarios on the basis of market research and technical feasibility.

Sufficient lead time is needed after the applications are selected for BAPCo to develop workloads, integrate the applications into the benchmark, and perform validation of the benchmark. Therefore, some of the application versions are not the newest available at the time of the launch of MobileMark 2014.

For MobileMark 2014, BAPCo has identified the following representative applications for the two usage scenarios.

Application	Version	Document Type
Adobe® Acrobat®	Pro XI	Portable document files, image files
Google® Chrome®	32.0	Web pages
Microsoft® Excel®	2013	Spreadsheets
Microsoft® OneNote®	2013	Free-form information gathering and multi-user collaboration
Microsoft® Outlook®	2013	E-mails
Microsoft® PowerPoint®	2013	Presentation files, image files, video files
Microsoft® Word®	2013	Word processing documents
WinZip®	Pro 17.5	Zip files, assorted document files

Table 1: Office Productivity Applications

Application	Version	Document Type
Adobe® Photoshop®	CS6 Extended	Image files
Adobe® Premiere®	Pro CS6 (On 64-bit Operating Systems)	Video files
Adobe® Premiere®	Pro CS4 (On 32-bit Operating Systems)	Video files
Trimble® SketchUp™	Pro 2013	CAD files, image files

Table 2: Media Creation Applications

2.4 Scoring Methodology

In milestone 10, BAPCo decides the types of results that will be produced by a benchmark and the scoring methodology that determines how those results are calculated.

BAPCo determines the scoring methodology before determining the content of the workloads, which helps ensure that a methodology is chosen for its ability to generate results that correspond to user experience, not for the results it produces on a pre-determined set of workloads.

In MobileMark 2014, each scenario produces a battery life rating and will also produce a performance rating. In each scenario the performance is given as a secondary result—a "performance qualification rating"—to the primary battery life rating.

MobileMark 2014 results are generated on a per-scenario basis, with each scenario requiring a separate run of the benchmark. Each scenario produces a battery life rating, and produces a performance qualification rating.

Battery Life Rating
The battery life rating for a scenario is calculated by measuring the actual battery duration observed (from a fully charged state to a fully depleted state), in whole minutes, while running that scenario in repetition.

Performance Qualification Rating
MobileMark 2014 measures system performance by measuring the response times of tasks on a PC using real applications and simulated user input. Task response times are used to generate a performance qualification rating for a scenario that reflects the user experience. The faster a PC responds to the

application workloads in MobileMark 2014, the higher its performance qualification rating will be. For more information on how task response times are measured, see section 2.8.

The performance qualification rating is given in relative terms, with respect to a calibration system. It is calibrated in such a way that a PC with performance equivalent to this calibration system for a given scenario will have a performance qualification rating of 1000. A system twice as fast as the calibration system on a given scenario (or, equivalently, that responds in half the time on average) will have a performance qualification rating of 2000.

Performance qualification ratings are given for each successfully completed scenario iteration and for the scenario overall.

Calculation of the Performance Qualification Rating

A performance qualification rating p_i is first calculated for each successfully completed iteration of a scenario (excluding any conditioning iteration and partial iterations):

$$p_i = 100 \left(\frac{s_c}{s_t} \right)$$

Where:

s_c is the sum of the response times of tasks in that iteration as performed on the calibration system, explained below.

s_t is the sum of the response times of tasks in that iteration as performed on the test system.

The calibration sum s_c is a fixed value BAPCo calculates by performing three runs of the scenario—each including a conditioning run—on a calibration system (see section 2.7). For each successfully completed iteration among all three runs (excluding any conditioning iterations and partial iterations), the sum is taken of the response times of all the tasks in that iteration. s_c is the arithmetic mean of all those iteration sums.

p_i is rounded to the nearest 1/100$^{\text{th}}$. For display purposes only, this is further rounded to the nearest integer.

Then the performance qualification rating for the scenario p_s is calculated by taking the arithmetic mean of the performance qualification ratings p_i for all successfully completed scenario iterations $i = 1, \dots, n$:

$$p_s = \frac{1}{n}\sum_{i=1}^{n} p_i$$

p_s is rounded to the nearest integer.

Results Estimation

During testing, once one full scenario iteration has completed, projected battery life and performance qualification ratings are calculated and displayed in the heads-up display. These values are rough estimates, reflecting the system behavior in the scenario iterations completed up to that point, and subject to the accuracy of the test system's battery level reporting mechanism.

The battery life rating estimate b_e is calculated using the following formula:

$$b_e = \frac{t(c_t)}{c_t - c_n}$$

Where:

t is the elapsed time since the test was started, in minutes.

c_t is the level of charge the battery reports having the capability to hold when fully charged, in any units as reported by the battery (typically amp-hours or watt-hours).

c_n is the level of charge reported by the battery presently, in the same units as c_t.

The performance qualification rating estimate is calculated by taking the arithmetic mean of the performance qualification ratings of all completed scenario iterations.

2.5 Workload Development Sessions

Once the usage models, scenarios, application models, and scoring methodology for the benchmark are decided, BAPCo members and application experts meet to create the application workloads that will be used in the benchmark.

For MobileMark 2014, the workload development sessions consisted of one week of face-to-face meetings that included representatives from BAPCo member companies and expert application users who had professional experience with the applications chosen for the benchmark. The application experts included professionals in the fields of small business marketing, financial forecasting, graphic design, video editing, web development, and enterprise IT deployment.

In the workload development sessions, the experts take the lead, weaving the usage models supplied by BAPCo into a storyboard of user interactions with a series of application models. Each user interaction is written down in a workload specification, which is later used to automate the workloads.

At the end of the workload development sessions, BAPCo comes away with a detailed workload specification for each of the benchmark scenarios and all of the input data sets needed to reproduce the workloads created at the sessions.

Additional Workload Considerations

The following additional factors were considered at the workload development sessions:

Input Data Set
Frequently in the sessions, the experts need raw digital content to serve as input data set for a workload. Examples of such content might include a video to transcode, an email to modify, or photos to manipulate. When experts need such content, care is taken to ensure that they use something that is functionally representative of content they might use or encounter professionally.

For instance, if pictures are needed in order to create a photo slideshow, an expert might walk outside and take pictures using the same equipment

he/she uses professionally. If a song track is needed as the background music for creating a movie, an expert might purchase a stock track from his/her usual online resource. Like the user interactions, all of these source materials are captured at the development session and used later in the development of automated workloads.

2.6 Scenario Workload Descriptions

The scenario workloads created at the workload development sessions for MobileMark 2014 are described below:

Office Productivity

Read and manipulate notes from a notebook. Browse multiple web sites which include a blog, online shopping site, wiki site and social networking site. Archive a diverse set of files into a single encrypted file. Convert a PDF document into an editable word processing document. Perform and mail merge. View a complex presentation that include multimedia and export it to PDF. Combine multiple scanned pages from a complex document into an encrypted PDF document using optical character recognition (OCR). Read, create and search for emails. Create and execute a rule on email inbox. Use a spreadsheet program to do data analysis.

Media Creation

Create a panoramic image using an image editing application, combine a set of photos into one high dynamic range (HDR) image, and adjust and prepare both images for print. Transcode the video to a format suitable for web publishing using a video editing application. Add visual features to an existing architectural model, and render "sketch" style views using an architectural modeling application.

2.7 Calibration System

The calibration system is a system chosen in Milestone 16 as a reference point for all other MobileMark 2014 results. BAPCo chose the configuration below for its wide availability and its representation of a typical mainstream notebook/2-in-1 computer at the time of release of MobileMark 2014.

MobileMark 2014 has been calibrated in such a way that a notebook/2-in-1 computer with performance equivalent to this calibration system for a given workload will have a performance rating of 1000. A system twice as fast as the calibration system on a given workload (or, equivalently, that responds in half the time on average) will have a performance rating of 2000. This is true for each of the scenario performance ratings.

The calibration system for MobileMark 2014 has the following configuration:

- Lenovo* ThinkPad T440*
- Intel Core i3-4010U Processor (3MB Cache, 1.70GHz)
- Integrated Intel HD Graphics 4400
- Microsoft* Windows 8.1* x64
- 4GB PC3-12800 DDR3L SDRAM 1600MHz SODIMM
- Toshiba* 500GB Hard Disk Drive, 7200rpm
- 14.0" HD+(1600×900)LED Backlight, Antiglare, With WWAN

A fresh operating system installation is performed on the system.

For more details about the configuration of the calibration system, please contact support@bapco.com.

2.8 Benchmark Implementation

Once the workload specifications have been created, BAPCo begins the work of translating the workload specifications into an automated benchmark in milestone 17.

MobileMark 2014 is built upon scripts that do things in much the same way as a user would, using controls like buttons, text input boxes, and menus to navigate applications. See Appendix C for screenshots of the benchmark in action.

To ensure that MobileMark 2014 has deterministic behavior, BAPCo uses a framework to install applications, collect system information, run the scenario scripts, record performance measurements, calculate performance ratings, and display test results. The framework is kept lightweight, consuming a minimal amount of memory and compute resources, in order to ensure that battery life and performance measurements reflect the workload behavior and do not include overhead from the framework.

The fundamental performance unit upon which the MobileMark 2014 performance qualification rating is based is *response time*. Response time is defined as the time it takes the computer to complete a task that has been initiated by the automated script. A task can be initiated by a mouse click or a keystroke. The duration of each task is measured by the framework. Examples of tasks include launching an application, finding text in a document, copying a file, encoding a video, and performing an image manipulation.

The framework has several methods of detecting task completion, depending upon the method the application uses to signal task completion to the user. For example, the framework may wait for the application to show a completion message in the form of a pop-up window, or it may wait for a progress dialog to disappear and for control of the application to be returned to the user.

2.9 Workload Characterization

Once the scenario workloads are implemented and validated against the workload specifications created at the development sessions, BAPCo members then run the benchmark on a wide variety of systems to ensure that the benchmark produces results that are valid, representative, and reproducible.

During this process, BAPCo members share data, raise concerns, and suggest workload changes. Any workload change requires a majority vote of the committee.

MobileMark 2014 is primarily a tool for measuring battery life; therefore, it's important that the power profiles of its scenario workloads are reasonably representative of user experiences and expectations. BAPCo members work together to arrive at an estimation of a representative level of user activity, but acknowledge that individual user experience could vary from the results reported by the MobileMark 2014 scenarios.

The battery life reported by a MobileMark 2014 scenario is an approximation of the battery life a user who performs that same scenario workload would expect from the same system.

One way BAPCo members adjust the power profile of the workload is by determining an appropriate proportion of user idle time relative to active time.

User Idle Time

Users of notebook and 2-in-1 computers often leave their systems idle for a period of time between sessions of active use. MobileMark 2014 simulates this behavior in the Office Productivity and Media Creation scenarios by interspersing occasional periods of user inactivity throughout the workload.

Inclusion of these idle periods better models real-world mobile usage and allows the hardware and software power management features of the system under test to behave in a realistic manner.

MobileMark 2014 includes user idle time in the Office Productivity and Media Creation scenarios in order to reflect the benefits of power-saving optimizations that are enabled by the power policy of the system under test and permitted by the MobileMark 2014 Benchmarking Rules. In order to provide a battery life metric that more closely aligns with user expectations of battery life, BAPCo chose not to model extended periods of user idle time (e.g., system standby) in MobileMark 2014.

Examples of permitted optimizations include display dimming, hard drive standby, CPU sleep, and wireless adapter power-saving mode, assuming such features are supported by the system under test and enabled in the system BIOS and OS power scheme.

Some power-saving optimizations are not permitted because of their relevance only in usage models featuring extended periods of user idle time. Examples of unpermitted optimizations include display standby, system standby, and system hibernation.

3　Workload Characteristics

This section provides data illustrating the performance characteristics of MobileMark 2014.

3.1 Sensitivity Analysis

The series of tables below shows the sensitivity of MobileMark 2014 to different system characteristics, including the amount of system memory (RAM), number of CPU cores, type of storage device, and display resolution.

Within each study only one system component (e.g. memory) is varied. All the other system components are held constant. To best illustrate the sensitivity, one configuration is chosen as a baseline and the ratings for the other configurations are shown as the percentage difference relative to the baseline.

In order to give better control over system configuration, not all sensitivity charts use the same baseline configuration. Note that component sensitivities will vary from one configuration to the next.

Also note, that all performance data in each table is relative to the baseline configuration.

Baseline configuration:
- Notebook model: Lenovo IdeaPad Z710
- BIOS: 7FCN88WW
- CPU: Intel Core i7-4700MQ Processor (2.40 GHz, 4 cores, 8 threads, 6MB cache)
- Operating system: Windows 8.1 Enterprise 64-bit
- Operating system language: US English
- Display type: 17.3" FHD (1920x1080) LED
- System graphics: Intel HD Graphics 4600
- System RAM: 8 GB DDR3L
- Hard drive: Seagate Momentus ST9160314AS 160GB 5400RPM
- Battery: 4 cell 41 Watt Hour Li-Ion
- Wireless LAN: Lenovo BGN Wireless

For the tables below, the following components were substituted as noted:
- System Memory:
    - 2 GB DDR3L (1 DIMM)
    - 4 GB DDR3L (1 DIMM)
    - 8 GB DDR3L (1 DIMM)
- CPU - Cores adjusted in OS settings
    - Intel Core i7-4700MQ Processor (1 cores HT-on, 2.4 GHz, 6 MB)
    - Intel Core i7-4700MQ Processor (2 cores HT-on, 2.4 GHz, 6 MB)
    - Intel Core i7-4700MQ Processor (4 cores HT-on, 2.4 GHz, 6 MB)
- Hard drive
    - Seagate Momentus ST9160314AS 160GB 5400RPM
    - Seagate Momentus ST500LT012-9WS142 500GB 7200RPM
    - Intel 530 Series SSDSC2BW240A401 2.5" 240GB
- Resolution
    - All resolution tests were done with the ASUS T272HUL 27 inch external monitor. The internal display was turned off.
    - 1366x768@60Hz
    - 1920x1080@60Hz
    - 2560x1440@60Hz

3.1.1 Sensitivity to System Memory

The total system memory is changed from 2 GB to 4 GB to 8 GB.

Performance			
System Memory Sensitivity	2GB (2x1GB)	4GB (4x1GB)	8GB (8x1GB)
Office Productivity Scenario	Baseline	+0.9%	+0.9%
Media Creation Scenario	Baseline	+12.2%	+13.0%

Table 1: System Memory Sensitivity

3.1.2 Sensitivity to CPU Cores

The total number of CPU cores available to the system is changed from 1 core to 2 cores to 4 cores. This was achieved by changing the number of processors in msconfig boot advanced options.

Performance			
CPU Core Sensitivity	1 Core, 2 Threads	2 Cores , 4 Threads	4 Cores, 8 Threads
Office Productivity Scenario	Baseline	+20.8%	+21.7%
Media Creation Scenario	Baseline	+41.6%	+53.9%

Table 2: CPU Core Sensitivity

3.1.3 Sensitivity to I/O Subsystem

The primary storage device is changed, from a mechanical spinning disk to 2 Solid-State Drives (SSDs).

Performance			
I/O Subsystem Sensitivity	160GB 5400 RPM HDD	500GB 7200 RPM HDD	240 GB SSD
Office Productivity Scenario	Baseline	+3.8%	+13.0%
Media Creation Scenario	Baseline	+2.2%	+6.5%

Table 3: I/O Subsystem Sensitivity

3.1.4 Sensitivity to Display Resolution

The system display resolution is changed, from 1366x768 to 1920x1080 to 2560x1440.

Performance			
	1366x768	1920x1080	2560x1440
Office Productivity Scenario	Baseline	-0.9%	-3.8%
Media Creation Scenario	Baseline	-1.1%	-4.6%

Table 4: Display Resolution Sensitivity

3.2 Application Contribution Analysis

The following charts shows the approximate contribution of each application to the MobileMark 2014 version 1.0.0.12 scenario ratings on the calibration system.

Due to system implementation differences, this chart is accurate for the calibration system only. These contributions will vary from one system to the next.
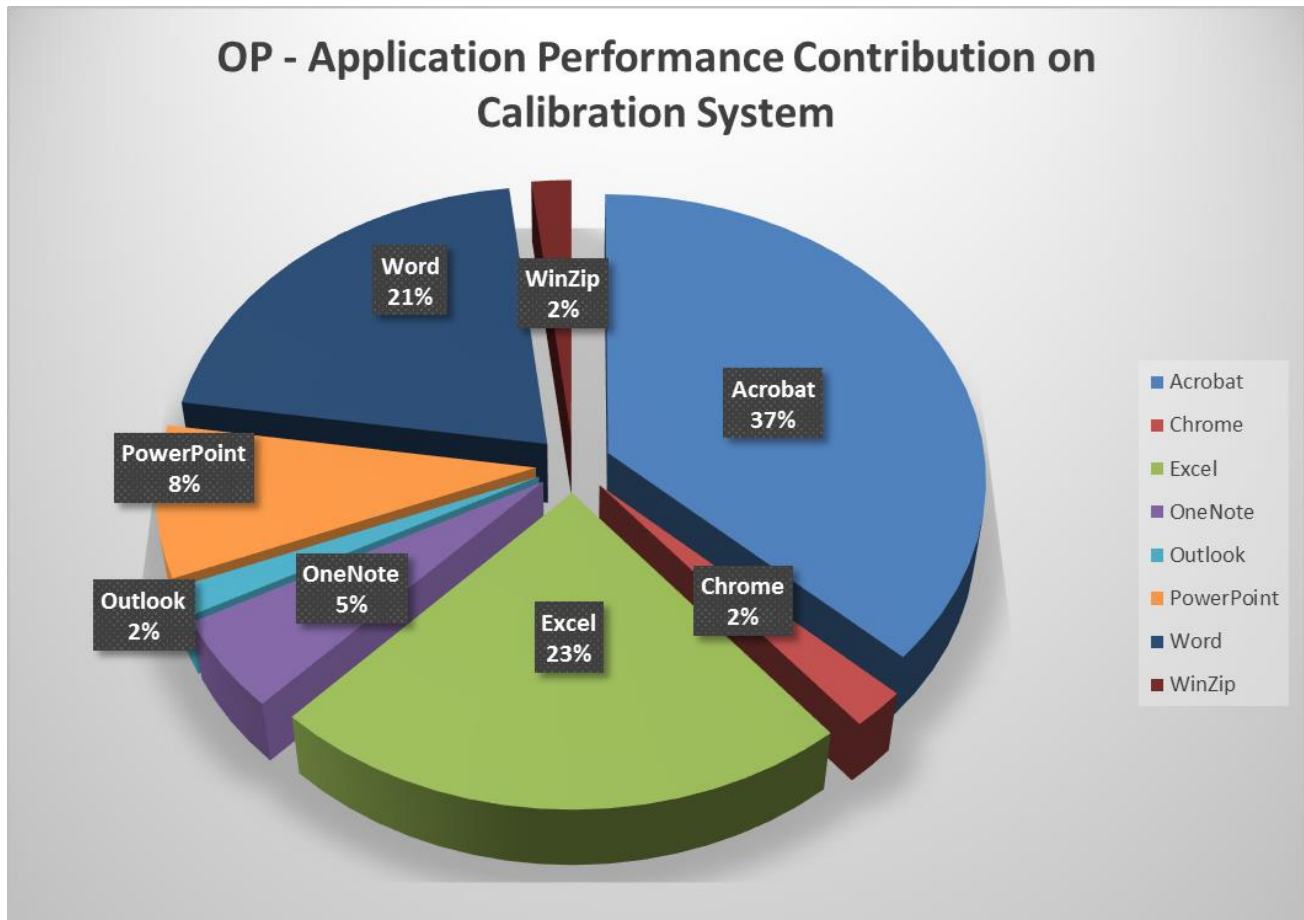
Figure 1: Application Model Contribution to Office Productivity Scenario Rating on Calibration System
Lenovo ThinkPad T440, Intel Core i3-4010U CPU (1.70GHz, 2 Cores/4 Threads, 3MB L2 cache), 4GB LPDDR3, 500GB 7200RPM Hard Drive, 14.0" LED Backlight display (1600x900), Microsoft Windows 8.1 x64.
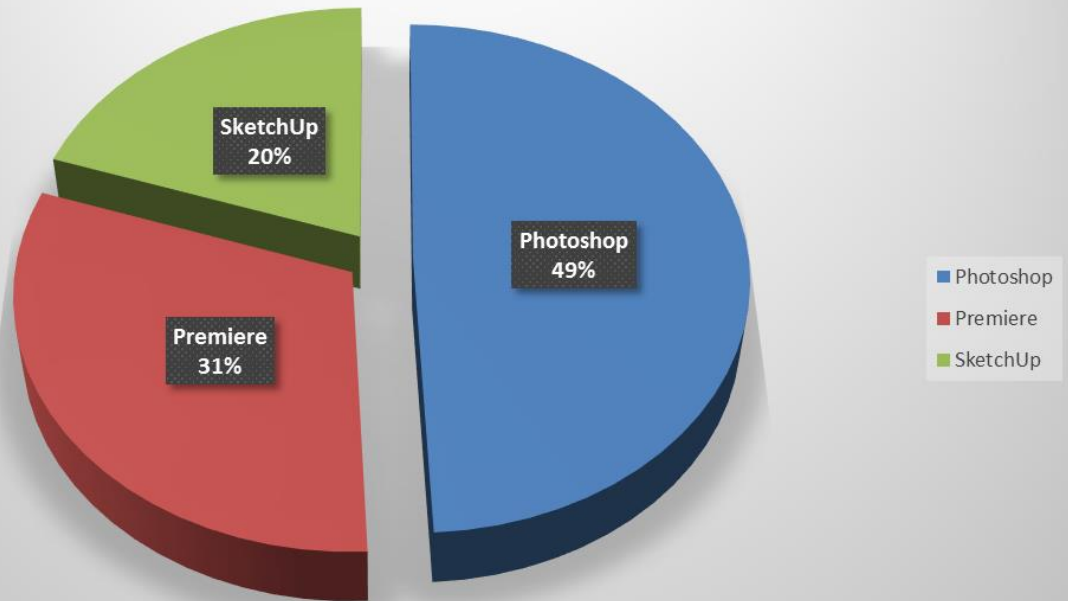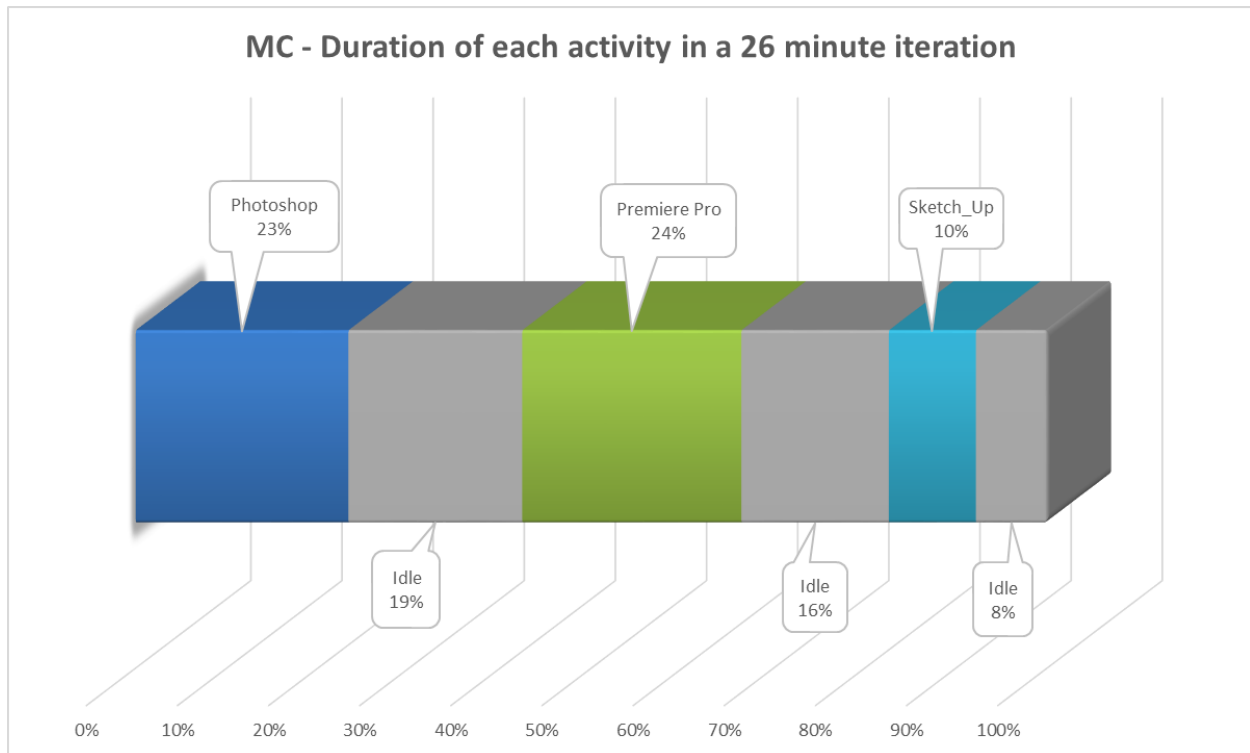
Figure 2: Application Model Contribution to Media Creation Scenario Rating on Calibration System
Lenovo ThinkPad T440, Intel Core i3-4010U CPU (1.70GHz, 2 Cores/4 Threads, 3MB L2 cache), 4GB LPDDR3, 500GB
7200RPM Hard Drive, 14.0" LED Backlight display (1600x900), Microsoft Windows 8.1 x64.

3.3 Battery Iteration analysis

The following chart shows the order each workload is run in and its relative time in an iteration for each scenario.

Due to system implementation differences, this chart is accurate for the calibration system only. These contributions will vary from one system to the next.
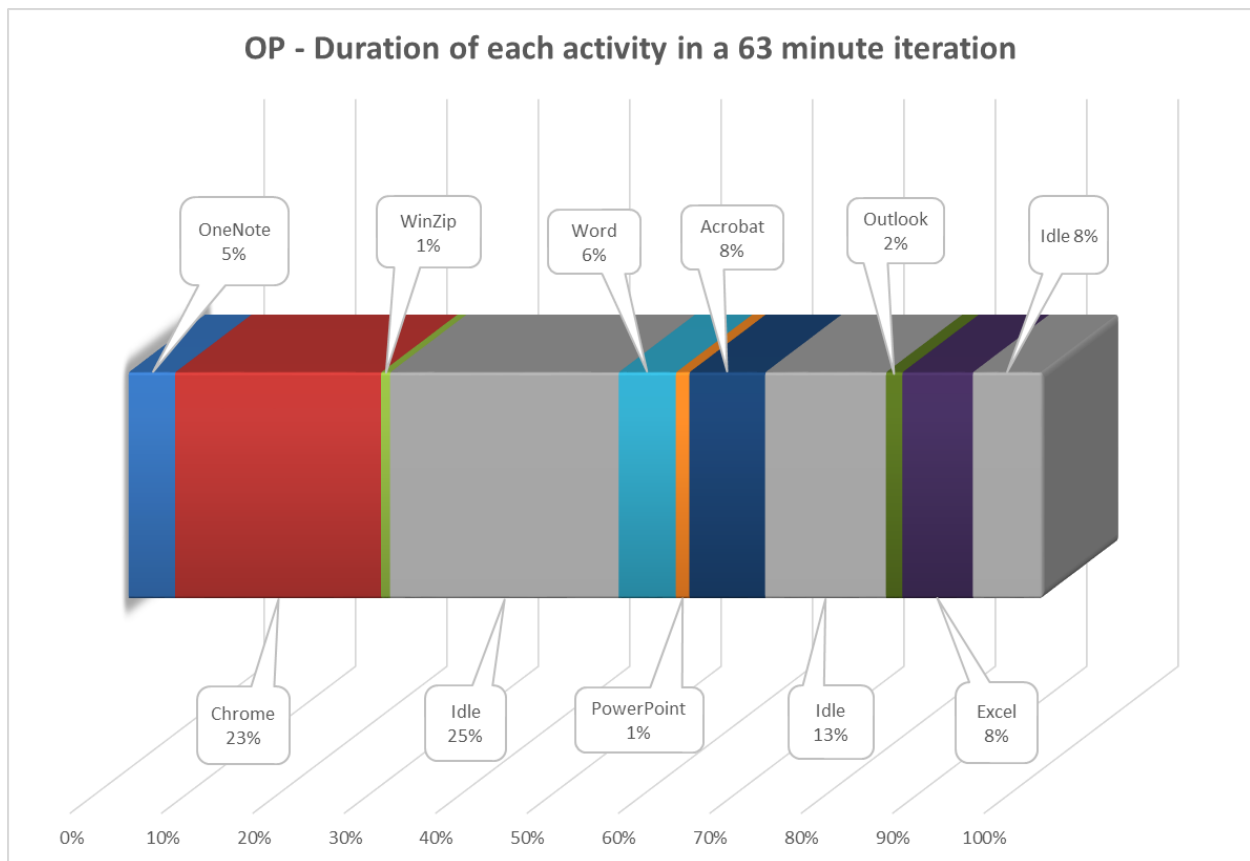
Figure 3: Office Productivity Scenario Iteration analysis on Calibration system
Lenovo ThinkPad T440, Intel Core i3-4010U CPU (1.70GHz, 2 Cores/4 Threads, 3MB L2 cache), 4GB LPDDR3, 500GB 7200RPM Hard Drive, 14.0" LED Backlight display (1600x900), Microsoft Windows 8.1 x64.

Figure 4: Media Creation Scenario Iteration analysis on Calibration system
Lenovo ThinkPad T440, Intel Core i3-4010U CPU (1.70GHz, 2 Cores/4 Threads, 3MB L2 cache), 4GB LPDDR3, 500GB
7200RPM Hard Drive, 14.0" LED Backlight display (1600x900), Microsoft Windows 8.1 x64.

APPENDIX A: System Requirements

MobileMark 2014 has the following requirements:

- **CPU**: 800 MHz AMD® or Intel® dual core processor
- **RAM**: 2 GB
- **Drive Space**: 50 GB of free space on the primary drive
- **Operating System**: Microsoft® Windows® 8.1 32/64-bit, Microsoft® Windows® 8 32/64-bit, or Microsoft® Windows® 7 32/64-bit
- **Video Resolution**: 1024 x 768
- **Graphics**: DirectX 9 compatible
- **Other**: Wireless Router

APPENDIX B: Screenshots

The screenshots below are included to illustrate the user interface and workloads included in the current version of MobileMark 2014. These screenshots may not accurately depict future releases of MobileMark 2014.

Figure 1: MobileMark Setup Screen

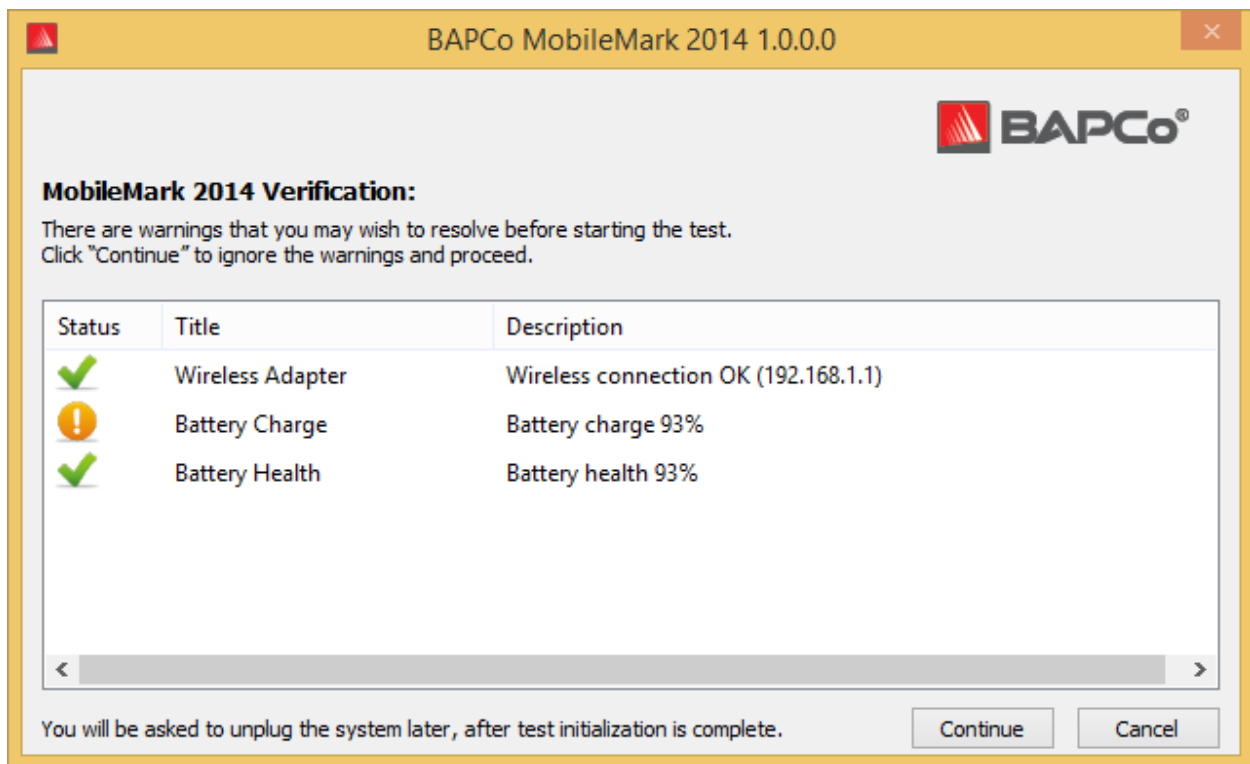*Figure 2: MobileMark Heads-Up Display*
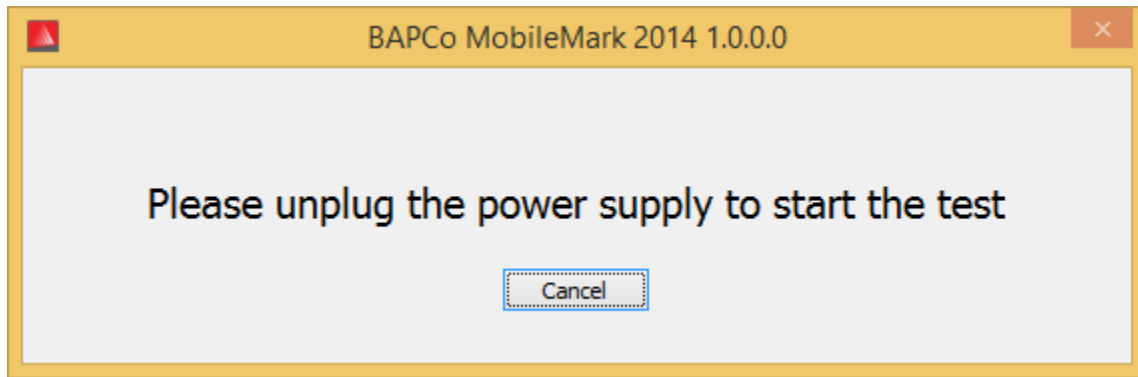
Figure 3: MobileMark Unplug Prompt
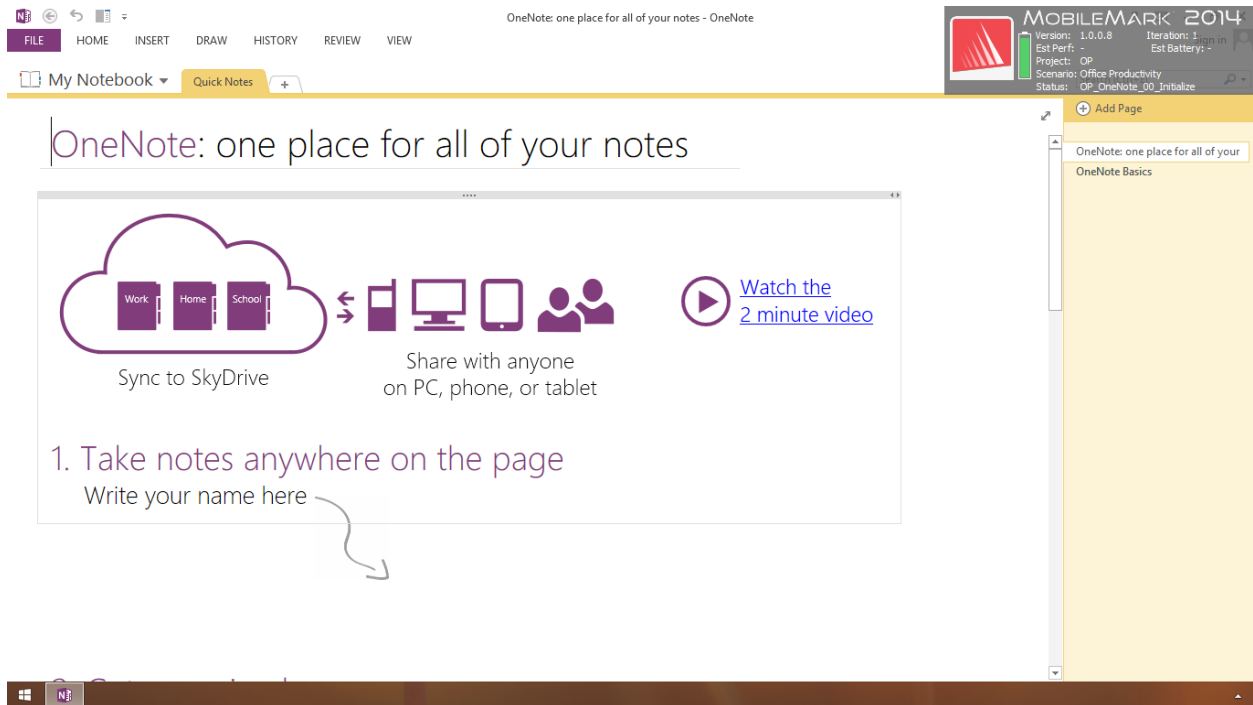

Figure 4: MobileMark Verification Windows

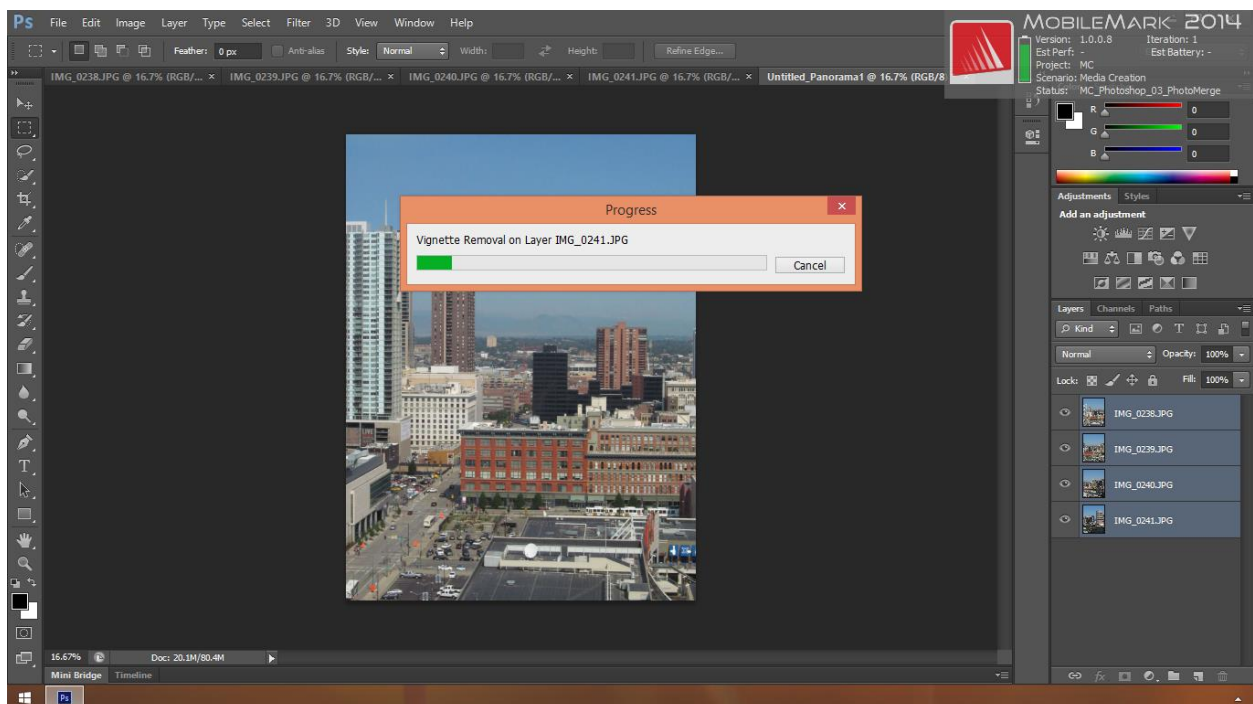*Figure 5: MobileMark Office Productivity Scenario*

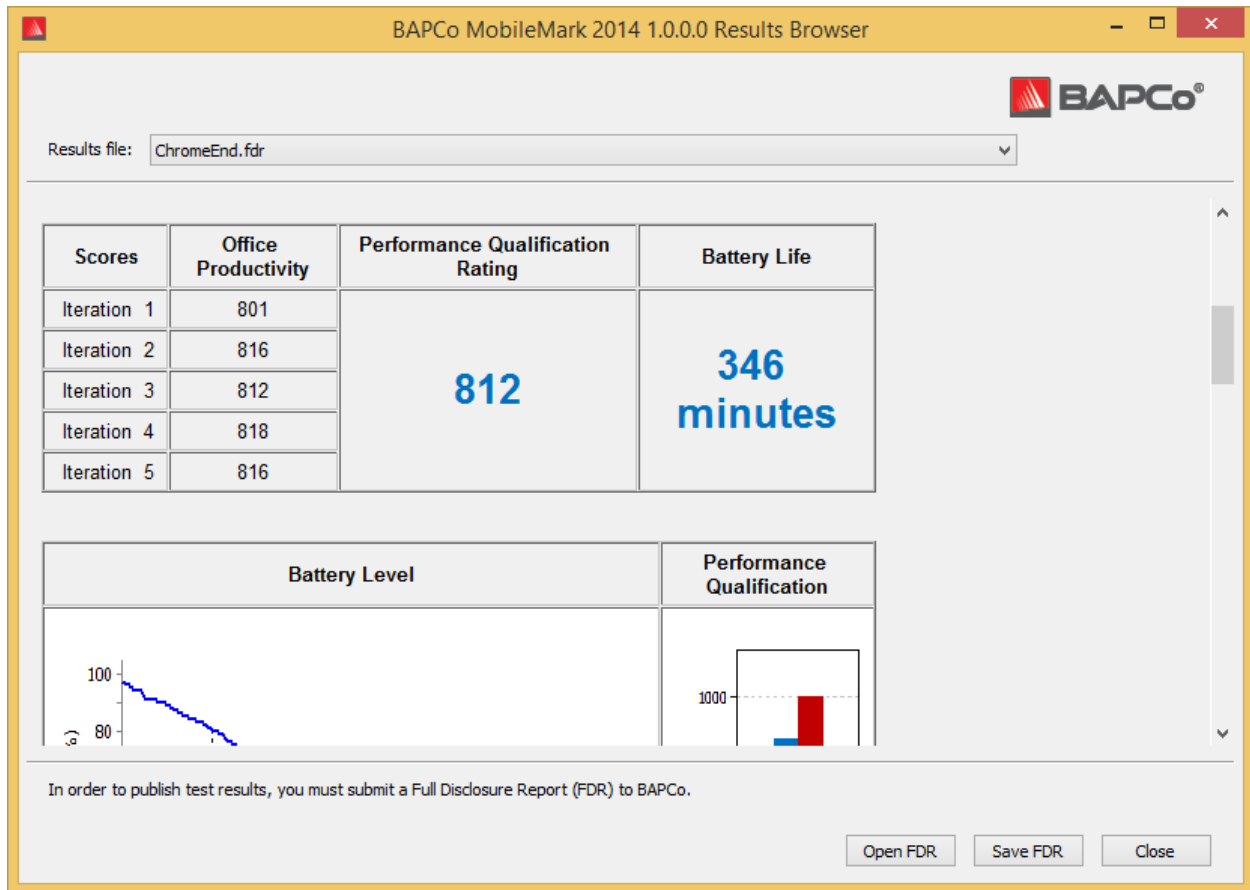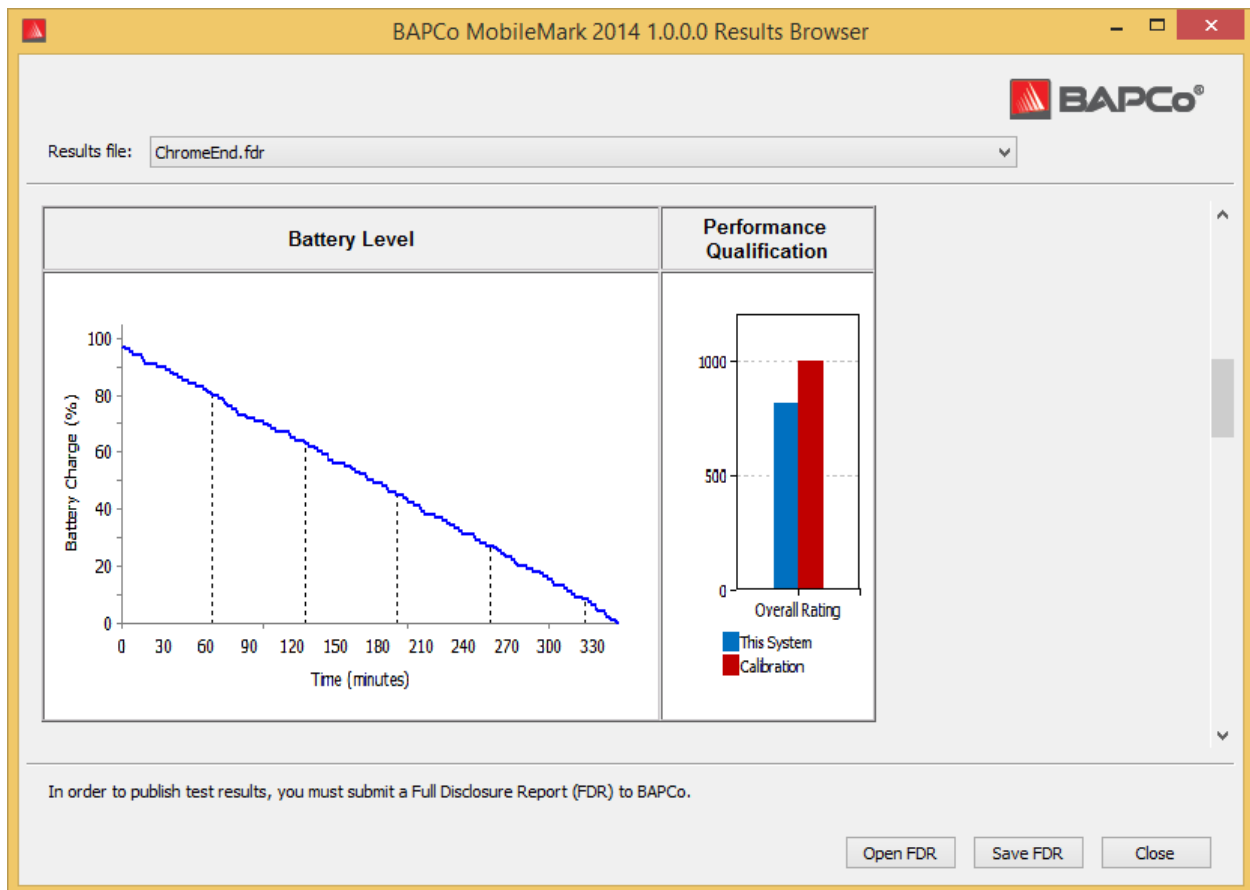*Figure 6: MobileMark Media Creation Scenario*

Figure 7: MobileMark Results Browser

Figure 8: MobileMark Results Browser (cont.)