

Usability Comparison between the MDA (MDAinternet.com) and the ODG- RTW (return-to-work.com)

Prepared by Erika Noll Webb, Quintus Design LLC
For more information, please contact
ewebb@quintusdesign.com 303.449.5020

User Interface Usability Comparison between the MDA (MDAinternet.com) and the ODG-RTW (return-to-work.com)

Reed Group Ltd. was interested in understanding how their Medical Disability Advisor (MDA) product (MDAinternet.com) compared to a competitor's product, the Official Disability Guidelines (ODG-RTW) web site (return-to-work.com) in relation to general system usability. Usability of a system has more to do with how a system works for a user than with how it looks. Usability is a measurement of the quality of a user's experience when he or she interacts with a system such as either of these. The quality of the interaction is comprised of a set of different factors including the ease of learning, ease and efficiency of use, memorability, errors (both frequency and severity) and, finally, the overall subjective satisfaction of using a system.

A system that is more usable reduces costs in a variety of manners, including money saved in reduced training. Increased efficiency translates to higher worker productivity. Fewer errors and less severe errors also are reflected in a company's bottom line.

For the purposes of this comparison, a set of typical questions was developed for the reviewer to attempt to answer with each product. These "tasks" take into account the information that a user of this type of system would need to be able to find in order to set return-to-work expectations for a few employee absences. The tasks used for each product's evaluation were:

1. What topic(s) address(es) "broken thighbone"?
2. For a sprained knee, what is the suggested duration of physical therapy services?
3. How soon would it be reasonably safe for an otherwise healthy stonemason's apprentice (who lifts heavy rocks and stones all day long) to return to work? He has had a smooth recovery from an open or traditional appendectomy with an abdominal incision. He wants to return to work as soon as possible.
4. Following laparoscopic surgery for appendicitis, when would a healthy customer service agent whose job involves answering the telephone in a call center typically return to work?
5. Following surgery for a ruptured appendix, when would it clearly become unusual for an otherwise healthy accountant (who works at a desk) to have not returned to work?
6. To what diagnosis does ICD-9 code 383.9 refer?
7. To what diagnosis does CPT code 57061 refer?

General Principles and Methods

In order to execute the tasks according to measurable events, a set of usability principles, or "rules of thumb," were considered. The set used for this review was taken from Nielsen (http://www.useit.com/papers/heuristic/heuristic_list.html). The list was chosen for its applicability to the products compared.

- **Match between system and the real world**

The system should speak the user's language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Systems should follow real-world conventions, making information appear in a natural and logical order.

- **Consistency and standards**

A user should not have to wonder whether different words, situations, or actions mean the same thing. Following platform conventions minimizes confusion.

- **Error prevention**

User-oriented designs eliminate error-prone conditions.

- **Recognition rather than recall**

Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

- **Flexibility and efficiency of use**

Accelerators—unseen by the novice user—may often speed up the interaction for the expert user with the result that the system caters to both inexperienced and experienced users. One result of accelerators is that the system tailors itself to each user's most frequent actions.

- **Aesthetic and minimalist design**

Dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

- **Help and documentation**

Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Based on these principles, each enactment of the tasks was given a usability "score". The score was based on the following scale, the higher scores signifying a more usable product:

- 4 = No usability problem**
- 3 = Cosmetic problem only**
- 2 = Minor usability problem**
- 1 = Major usability problem**
- 0 = Usability catastrophe**

The factors that influence the score include frequency and impact. Minor problems that occur frequently are given a more severe score than those that are minor and infrequent. Likewise, minor issues that cause significant increases in the time to complete the task can

impact the overall effectiveness of the product and would also lead to a more severe score on this scale. Finally, problems that only have a slight impact on completion of the tasks are not rated as severely as those that potentially cause the user to make errors in the task.

Test conditions

A single tester reviewed both products using a computer with the XP operating system and Internet Explorer 6.0.2900.2180 with SP2. Times required to complete the tasks were tracked by stopwatch.

Testing Organization

Quintus Design is a consulting firm that offers services to measure and improve the quality of a user's experience when interacting with technology or a system. Core competencies include:

- * Usability
- * Surveys
- * Accessibility
- * Focus Groups
- * Training

Quintus has experience with a broad spectrum of high-tech products and services and can quickly provide relevant feedback to lower support costs, improve user efficiency, speed development time, and align with government requirements. All the consultants at Quintus have extensive training and educational backgrounds in the behavioral sciences and collectively over 50 years experience applying that expertise in high technology businesses. The result is a strategy for product design that balances a rigorous scientific approach with the pragmatic needs of the business environment. Quintus consultants have earned several awards and patents for their innovative designs.

Tester Unfamiliarity

The reviewer, Erika Noll Webb, had no prior exposure to the usability or user interface of either the Medical Disability Advisor or Official Disability Guidelines products at the outset of this project.

Results

For each task, several measurements were taken. In addition to the 0-4 scale score described above (the lower the number the less useable the interface), the total time to complete the task was measured for the standard-use case path. In addition, measurements included the number of clicks to new pages required to complete each task.

Please refer to Appendix A for details about the usability score assignments per task.

Time to complete

Figure 1 shows the mean time to complete each of the 7 tasks. For the MDA product, the mean time to complete each task was 44.3 seconds (range of 15-158 seconds). Completing the same tasks in the ODG-RTW product took twice as long, with a mean time to complete of 96.3 seconds (range of 31-156 seconds).

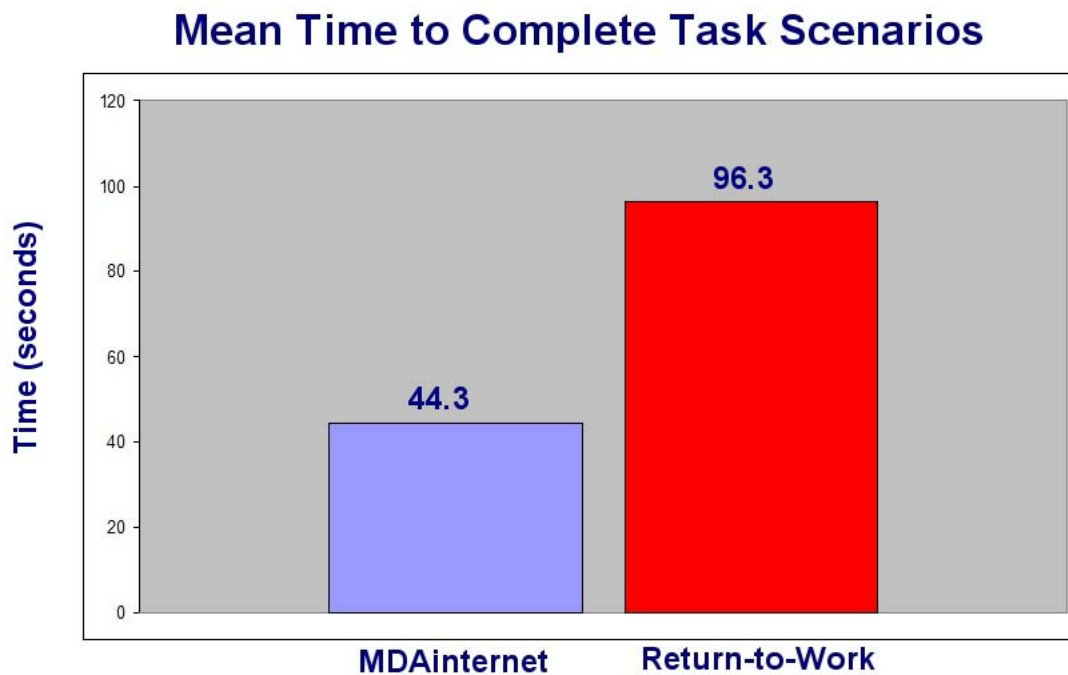


Figure 1. Mean time to complete tasks in seconds.

Number of Clicks to Complete Task

Another measurement of the ease and efficiency of a product with a task is the number of clicks that it takes a user to get to the information that he or she needs. In the comparison between the MDA product and the ODG-RTW product, the difference in clicks to find the needed information in the ODG-RTW product was striking (see Figure 2). For the MDA product, the mean number of clicks to complete a task was 4 (range of 3-6). For the ODG-RTW product, the mean number of clicks was 5.9 (range of 5-10).

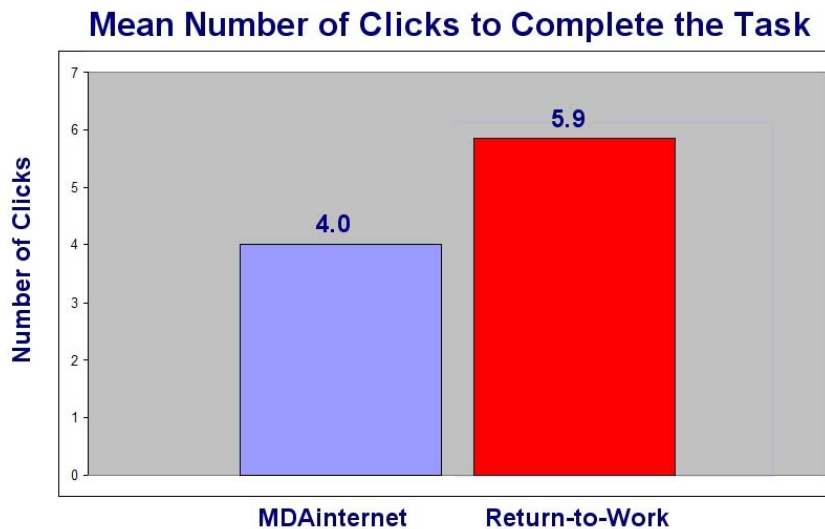


Figure 2. Mean number of clicks required to complete the task.

Usability Score

Using the scoring system described in the introduction, the reviewer scored the attempt to answer each question according to the specified principles. The scoring took into consideration each of these principles to create a single, overarching score to best describe the usability of the product.

In certain cases, the ODG-RTW product did not appear to include the necessary information. Assuming that the questions focused on the type of information that is fundamental to managing disability cases, this shortfall in the system impacts the usability because the user of the system would be unable to complete the tasks with this system and would be forced to search for the information outside of this product. Consequently the time to complete the tasks would increase. In Figure 3, the MDA product had minor usability issues with a mean usability score of 3.83 (range of 3-4), while the ODG-RTW product had a mean usability score of 0.67 (range of 0-4).

Usability Score Based on Heuristic Evaluation

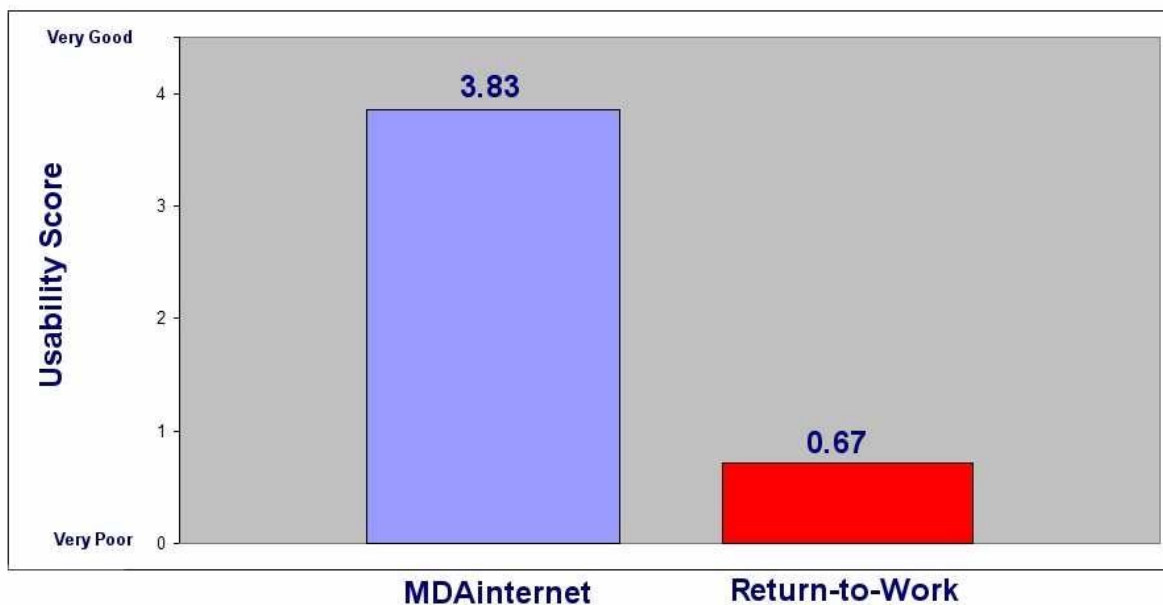


Figure 3. The mean usability score between the MDA (MDAinternet.com) and ODG-RTW (return-to-work.com) systems, where "4" indicates no usability problem and "0" indicates a catastrophic usability problem. There were few problems with the MDA product and there were more significant usability problems with the ODG-RTW product.

Appendix A

The following table describes factors that influenced the overall usability score for each task. Where a system does not achieve a perfect "4," we state the principle on the basis of which the overall score given was reduced. The final score is additionally influenced by the length of time to complete the task, the number of clicks to complete the task and whether or not the task could be completed successfully. The scores are not derived from a formula, but from expert reviewer(s) assessment of the severity of the problems encountered. The score was based on the following scale:

- 4 = No usability problem**
- 3 = Cosmetic problem only**
- 2 = Minor usability problem**
- 1 = Major usability problem**
- 0 = Usability catastrophe**

Table 1. Table 1 shows the results of the task evaluations for the MDA (MDAinternet.com) product.

Tasks	Usability Score	Principles Used in Scoring	Comments
1. What topic(s) address(es) "broken thighbone"	3	Match between system and the real world	Broken "thighbone" didn't return any hits, but broken "thigh" did. Some of the other search results are not relevant, but within broken thigh, the search made it obvious that the thighbone is the femur.
2. For a sprained knee, what is the suggested duration of physical therapy services?	4		The task was fairly easy to perform. As a side note, the answers obtained from the MDA and the ODG-RTW for this task differed substantially.
3. How soon would it be reasonably safe for an otherwise healthy stonemason's apprentice (who lifts heavy rocks and stones all day long) to return to work? He has had a smooth recovery from an open or traditional appendectomy with an abdominal incision.	4		No problems encountered.
4. Following laparoscopic surgery for appendicitis, when would a healthy customer service agent whose job involves answering the telephone in a call center typically return to work?	4		No problems encountered.

Tasks	Usability Score	Principles Used in Scoring	Comments
5. Following surgery for a ruptured appendix, when would it clearly become unusual for an otherwise healthy accountant (who works at a desk) to have not returned to work?	4		No problems encountered.
6. What is ICD-9 code 383.9?	4		No problems encountered.
7. What is CPT code 57061?	4		No problems encountered.

Table 2. Table 2 describes the ODG-RTW (return-to-work.com) product.

Tasks	Usability Score	Principles Used in Scoring	Comments
1. What topic(s) address(es) "broken thighbone"	1	Flexibility and efficiency of use Match between system and the real world Error prevention Recognition rather than recall	There was no single location for all of the topics that addressed "broken thighbone." Creating a list of all the topics that covered broken thighbone required a great deal of scanning and interpolation. The list itself had to be created outside of the product because there didn't appear to be a single efficient way to make a list within the product.
2. For a sprained knee, what is the suggested duration of physical therapy services?	4		The task was fairly easy to perform. As a side note, the answers obtained from the MDA and the ODG-RTW for this task differed substantially.
3. How soon would it be reasonably safe for an otherwise healthy stonemason's apprentice (who lifts heavy rocks and stones all day long) to return to work? He has had a smooth recovery from an open or traditional appendectomy with an abdominal incision. He wants to return to work as soon as possible.	1	Flexibility and efficiency of use Match between system and the real world Error prevention Recognition rather than recall	This task was never successfully completed. The "best practice" durations do not provide this level of granularity.
4. Following laparoscopic surgery for appendicitis, when would a healthy customer service agent whose job involves answering the telephone in a call center typically return to work?	1	Flexibility and efficiency of use Match between system and the real world Error prevention Recognition rather than recall	This task was never successfully completed. There was no clearly stated typical case, only a "best practice."

Tasks	Usability Score	Principles Used in Scoring	Comments
5. Following surgery for a ruptured appendix, when would it clearly become unusual for an otherwise healthy accountant (who works at a desk) to have not returned to work?	1	Flexibility and efficiency of use Match between system and the real world Error prevention Recognition rather than recall	This task was never successfully completed. Search time was increased due to a mismatch in terms between the task and the description in the product.
6. What is ICD-9 code 383.9?	3	Flexibility and efficiency of use	The task was completed but the process required an unnecessary number of steps.
7. What is CPT code 57061?	0	Flexibility and efficiency of use Match between system and the real world Error prevention Recognition rather than recall	This code was not included in the CPT codes in this product, so the task could not be completed. This lacuna may be an anomaly.