Enterprise Technology Usability Scorecards

February 14, 2014

Executive Summary

The Usability Score Card is based on 9 specific categories of the user experience as pertaining to applications. Scores above 55 are considered to have an acceptable user experience; scores below 55 are considered in need of remediation. The lower the score the riper the application is for an overhaul from a usability perspective.

As a general grouping the Enterprise scorecards earned a mean score of 58%. From the bar of 55% half the scorecards were above the bar with a scoring of 66% and the other half below the bar with a scoring of 49%.

Two of the most important categories are the efficiency of the application and time-saving accelerators. These comprise 37% of the total score. Unfortunately, the scorecards rank 6 & 7th in these categories.

Improving the efficiency improves the ability to get more done in less time. Improving the efficiency reduces the overall man-hours wasted by poor application design.

Some of the issues raised by the scorecard:

- Lack of pre-populating of known data
- Applications do not contain necessary wizards
- Missing power tools for repeating tasks or functions
- Unclear error messaging
- Branding not updated to current standards
- Inconsistent navigation where the behavior changes depended of where you are in the application
- Screen layout not being optimized to enhance the consumption of information
- Lack of visual cues
- Visually dated applications
- Window sizes not to current standards
- Lag in loading of data.

Possible low impact changes with big benefits in UX:

Improving messaging in some instances may only mean re-writing error messaging into a user readable format, or including better help texts.

The naming of fields and buttons can improve applications.

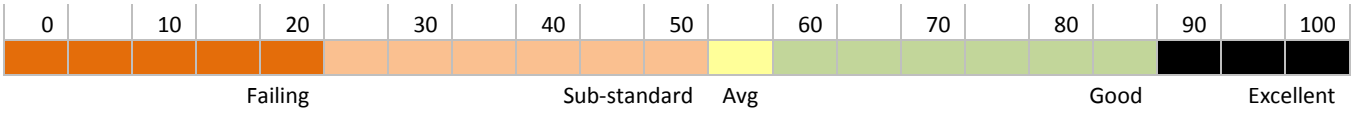
More invasive tasks include: visually lining up fields so they are easier to read and consume by users or taking advantage of screen real estate though flexible width design or newer fixed width displays.

Table of Contents

About the Scorecard

The Usability Score Card is based on 9 specific categories of the user experience as pertaining to applications. Each of the 27 facets is scored to a maximum of 5 points, some selections are weighted lower. An average score is 60.

Scores above 55 are considered to have an acceptable user experience; scores below 55 are considered in need of remediation. The lower the score the riper the application is for an overhaul from a usability perspective.



Enterprise Scorecards

In 2013 twenty scorecards were developed for application in Enterprise Technology. While the specific application were all different, utilizing a standard set of usability questions and a consistent scale to measure the answers to each question gives a unique perspective from a user centered approach to the landscape of applications in Enterprise Technology.

Each application was reviewed by multiple reviewers. As part of the process if a score for individual metrics was below average then the reviewer was requested to "defend" the score with an observation as to explain the score. These scores were then averaged within the nine categories.

As a grouping the Enterprise scorecards scored 58% which places the group as average among applications. Some individual categories within the grouping received sub-standard scores. These include Accelerators (54%), Delight (50%) and Learnability (48%). Accelerators comprise the largest part of the overall score garnering 14% of the total score.

When looking at sub-standard applications there are 6 of the 9 parameters that fall below the average as indicated in the chart below.

	Overall		Sub-standard		Good or better		Difference	Max Points
Efficiency	13.2	**57%**	11.0	**48%**	15.5	**67%**	20%	23.0
Accelerators	7.6	**54%**	6.1	**44%**	9.1	**65%**	21%	14.0
Messaging	6.6	**55%**	5.9	**49%**	7.4	**62%**	13%	12.0
Consistency	6.7	**56%**	5.3	**45%**	8.1	**68%**	23%	12.0
Visualization	7.4	**62%**	6.7	**56%**	8.1	**67%**	11%	12.0
Learnability	3.9	**48%**	3.0	**37%**	4.8	**60%**	22%	8.0
Currency	5.1	**72%**	4.5	**64%**	5.7	**81%**	18%	7.0
Speed	3.7	**62%**	3.3	**55%**	4.1	**69%**	13%	6.0
Delight	3.0	**50%**	2.5	**42%**	3.5	**59%**	17%	6.0
Totals		**58%**		**49%**		**66%**	**18%**	

Scorecards Head-to-Head

As a general grouping the Enterprise scorecards earned a mean score of 58%. From the bar of 55% half the scorecards were above the bar with a scoring of 66% and the other half below the bar with a scoring of 49%.

Using this standard metric a ranking was determined and is illustrated below.

Good:

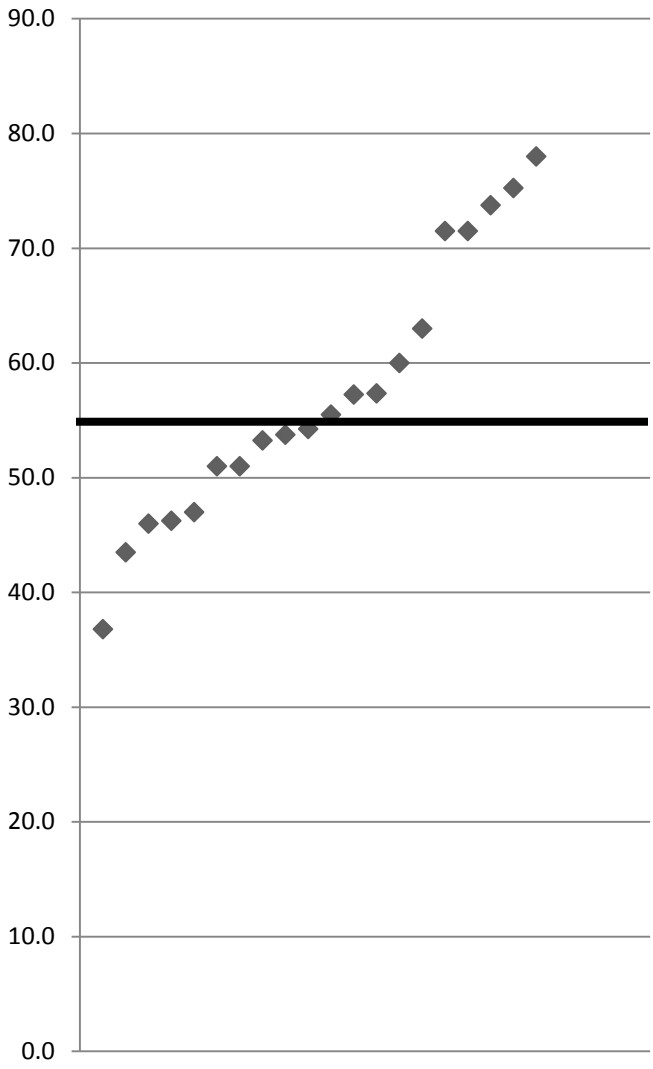
78.00	IAM Governance Sailpoint IIQ
75.25	Wageworks
73.75	Car Service Application
71.50	Thanks to You
71.50	Visitor Registration
63.00	Outside Business Activities (OBA)
60.00	Success Factors

Average:

57.33	Regulatory Tracking (RTS)
57.25	GME NYLIC
55.50	Learning Gateway

Sub Standard:

54.25	Security Liaison Application
53.75	GME NYLIM
53.22	GME International
51.00	Leads Distribution System (LDS)
51.00	WARM
47.00	Conference Center Database
46.25	Office Foreign Assets Ctrl (OFAC)
46.00	Security Liaison Tracker
43.50	Business Contingency
36.80	Remedy Self Service



From a usability perspective the items that score blow the acceptable bar are in need of a fuller review to see if there is a way to improve the overall experience. At a minimum this data should be cross-referenced with the use and impact data of these applications to determine how to address the usability issues.

What is interesting however is how the groupings are further banded above and below the line with a 5 point drop between the bottom most applications and a 8 point gain in the top most applications, indicating that the first items to investigate are the bottom 5 and in particular Remedy, scoring the lowest at 36.8.

Scorecard Categories

The usability scorecard is broken up into 9 categories. Each of the 9 categories is weighted based on importance in the overall landscape of usability components as they pertain to New York Life.

Items such as Efficiency, Accelerators and Messaging encompass the values such as speed in processing, entering/retrieving information and error correction in a business setting. Items such as maintaining current technology and utilizing the best interaction techniques are important factors but less so than other categories.

Looking at the categories in comparison to the application scorecards we see than Delight is low on the priority list, but learnability is lower. This leads to increased training costs and the potential of incorrect data being communicated into systems.

Two of the most important categories are the efficiency of the application and time-saving accelerators. These comprise 37% of the total score. Unfortunately, the scorecards rank 6 & 7th in these categories. Improving the efficiency improves the ability to get more done in less time. Improving the efficiency reduces the overall man-hours wasted by poor application design.

	Distribution	Overall	Sub-standard	Good or Better
Efficiency	23%	7	5	5
Accelerators	14%	6	7	6
Messaging	12%	5	4	7
Consistency	12%	4	6	3
Visualization	12%	3	2	4
Learnability	8%	9	9	8
Currency	7%	1	1	1
Speed	6%	2	3	2
Delight	6%	8	8	9

Scorecards by Category

Each application's score is captured in an individual scorecard and associated notes.

In looking at the 20 scorecards as a whole there are some trends that tend to lower the potential score of applications. These are broken up by category below, with an explanation of the categorization.

The categories of the scorecard can be grouped into three gradations, primary factors, secondary and tertiary.

Primary	
Efficiency	23%
Accelerators	14%
Messaging	12%
	49%

Secondary	
Consistency	12%
Visualization	12%
Learnability	8%
	32%

Tertiary	
Currency	7%
Speed	6%
Delight	6%
	19%

Primary categories focus on processing and error correction/recovery, while secondary categories focus on how understandable the interface is to the user and the user's ability to effectively consume the information the interface is communicating.

Tertiary categories focus on the experience of using the interface. If we were all robots then this category would not be a factor, but as living breathing human beings we have choices. If an interface not desirable, then people will attempt to avoid it, potentially completing tasks outside the system, such as calling instead of opening up support tickets, shifting the work from one person to another.

	Overall	Sub Standard	Good or Better
Primary	55%	**47%**	65%
Secondary	56%	**46%**	65%
Tertiary	61%	**54%**	70%

Looking at the primary and secondary metrics for the sub-standard subset we see that it produced the lowest scores.

	Primary	Secondary	Tertiary
Remedy Self Service	38.8%	30.0%	43.2%
Business Contingency	37.2%	49.2%	50.0%
Security Liaison Tracker	44.4%	40.6%	59.2%
Office Foreign Assets Ctrl (OFAC)	50.0%	42.2%	43.4%
Conference Center Database	46.4%	46.9%	48.7%
WARM	47.4%	44.5%	71.1%
Leads Distribution System (LDS)	50.0%	50.0%	55.3%
GME International	53.6%	52.3%	53.9%
GME NYLIM	48.0%	60.9%	56.6%
Security Liaison Application	52.6%	53.9%	59.2%
Learning Gateway	48.5%	60.2%	65.8%
GME NYLIC	56.6%	54.7%	63.2%
Regulatory Tracking (RTS)	59.9%	51.1%	61.4%
Success Factors	57.1%	60.9%	65.8%
Outside Business Activities (OBA)	63.9%	61.5%	63.2%
Visitor Registration	66.8%	73.4%	80.3%
Thanks to You	69.9%	71.1%	76.3%
Car Service Application	75.0%	72.7%	72.4%
Wageworks	73.5%	75.0%	80.3%
IAM Governance Sailpoint IIQ	81.6%	75.0%	73.7%

Efficiency is an application's ability to complete tasks quickly and effectively. Key metrics under efficiency include a beneficial balance of the effort provided to the value achieved. It includes the system's ability to know and utilize known information such as user profile information and previously enter data.

At its best the system can offer time saving solutions based on the data it has and avoid the user entering in data previously obtained or derived by calculation.

Issues raised by the scorecard for efficiency:

- Sorting and filtering issues
- Lack of pre-populating of known data

Accounting for 23% of the total score, efficiency is a primary factor where Enterprise applications garnered 56% of the available points and only 48% when looking at only sub-standard applications.

Accelerators are elements of an interface that increase the overall productivity through thoughtful design, power tools for habitual users and functionality that reduces the user's cognitive load.

Well designed systems make it easy for users to perform repeating actions, where ill designed systems will transfer the limitations of the interface or technology on to the user.

Issues raised by the scorecard for accelerators:

- Applications do not contain necessary wizards
- Missing power tools for repeating tasks or functions
- Lack of pre-fill when needed

Accounting for 14 % of the total score, accelerators are a primary factor where Enterprise applications garnered 54% of the available points and dismal 44% when looking at only sub-standard applications.

Messaging is a system's way of communicating issues with the user. An application's ability to communicate data errors, input errors, and workflow affects the user's ability to process work and complete tasks.

The pinnacle of exceptional messaging is when systems are constructed to either avoid errors, aid in the remediation process, or self correct while informing the user. An example is spell check where common mistakes are simply fixed and misspellings requiring interpretation are given a visual cue (wavy underline) to note the error.

Issues raised by the scorecard for messaging:

- Missing or unhelpful ALT tags
- Unclear error messaging

Accounting for 12% of the total score, messaging is a primary factor where Enterprise applications garnered 55% of the available points but only 49% when looking at only sub-standard applications.

Consistency is the application's ability to consistent with New York Life Standards, general internet standards and standards or conventions within a particular application.. Examples include: proper branding, functions behaving as expected for web applications, and the application's specific ways of completing tasks.

Applications that look and act consistently with similar applications and functions have a higher rate of overall acceptance, reuse, and lower support calls related to interface quirks.

Issues raised by the scorecard for consistency:

- Branding not updated to current standards
- Inconsistent navigation with internet norms, such as navigation used to launch both functions content
- Inconsistent navigation where the behavior changes depended of where you are in the application
- Links not appearing as links or not behaving as expected

Accounting for 12 % of the total score, consistency is a secondary factor where Enterprise applications garnered 56% of the available points and but only 45% when looking at only sub-standard applications.

Visualization revolves around how items are organized on the screen. This includes element spacing and visual alignment. It also comprises how complex flows are communicated to the user.

Proper visualization helps the user consume the interface and the information that is provided in a way that is easily understood.

Issues raised by the scorecard for visualization:

- Lack of a guided experience or flow to reduce complication
- Screen layout not being optimized to enhance the consumption of information

Accounting for 12 % of the total score, visualization is a secondary factor where Enterprise applications garnered a high 62% of the available points and 56% when looking at only sub-standard applications.

Learnability is a system's ability to train the user to utilize functionality to complete tasks. Some options include in-context help, training materials and wizards.

While not all systems need user aids, many systems do this covertly by standardizing on known conventions and metaphors such as spreadsheet tables to enter monetary data and common navigation practices such as tabs.

Learnability and consistency were noted as the largest gaps between applications that scored well and those that do not. Collectively there was a 22% difference between the two groups.

Issues raised by the scorecard for learnability:

- Lack of visual cues
- Lack of in-context help when warranted
- Use of jargon and overuse of acronyms
- Instances where training is required and is not provided or not provided adequately

Accounting for 8% of the total score, learnability is a secondary factor where Enterprise applications garnered only 48% of the available points and a dismal 37% when looking at only sub-standard applications.

Currency is the concept of browser compatibility and screen size, and is the most straight forward of the categories. Outdated code and technology and limited display reduce the overall possible score. The score for currency across all the application was the highest of any category.

Issues raised by the scorecard for currency:

- Visually dated applications
- Window sizes not to current standards

Accounting for 7 % of the total score, is a tertiary factor where Enterprise applications garnered surprising 72% of the available points and 64% when looking at only sub-standard applications.

Speed refers to both the system's ability to present data at an acceptable rate and the user's perception that information is delivered to them quickly. As an example, a system can provide a response to a query instantaneously, but if that information is not what the user needs or expects a system will appear slow.

Issue raised by the scorecard for speed:

- Lag in loading of data.

Accounting for 6% of the total score, speed is a tertiary factor where Enterprise applications garnered a high 62% of the available points and 55% when looking at only sub-standard applications.

Delight awards points for applications that are visually pleasing and instances where applications are taking advantage of newer interaction techniques. Delight produced the lowest category score in all groupings.

Issues raised by the scorecard for delight:

- Treatment of applications as simple tool to process instead of true applications
- Old interaction techniques

Accounting for 6% of the total score, delight is a tertiary factor where Enterprise applications garnered only 50% of the available points and a dismal 42% when looking at only sub-standard applications.

Raising Scores and Next Steps

The scores of the applications included are fluid, as new interaction techniques, standards and user expectations are created scores will tend to lower based on these factors. This however does not mean that scores cannot be improved as applications are redesigned, remediated or through general support.

Improving messaging in some instances may only mean re-writing error messaging into a user readable format, or including better help texts.

The naming of fields and buttons is another area where small changes can improve applications.

More invasive tasks include visually lining up fields so they are easier to read and consume by users or taking advantage of screen real estate though flexible width design or newer fixed width displays. Simply adopting current branding can also refresh old interfaces, some applications use branding that is 10 years only or 2-3 branding cycles behind the current standard.

Minimally, reviewing individual scorecards and performing heuristic evaluations, which are comprised of a more in-depth analysis based on general usability principles can guide where actions can be prioritized based on individual systems to include in future development.