**使用性測試 (Usability test)**

<https://reurl.cc/4l2Vj>

**前言**

當你的團隊千辛萬苦完成了軟體、app 或網頁，滿心期待開放給大家使用，結果發現使用者一直不按下「結帳」、或者使用者用了反應太難用等等問題，當下只能穿著衣服改衣服，原來忙碌的團隊更加忙碌，新的專案因此受到耽擱。但如果再發佈前，就能先掌握這些問題，免除發佈後修修改改的狀況呢？**使用性測試 (Usability test)**，不但可以運用在產品發佈後的測試，也可以在研發時就先測試找出問題，但該怎麼做？

在這邊文章你可以暸解到：

1. 何謂使用性測試

2. 使用者測試的內涵與精神

3. 該如何施行使用性測試

**使用性與使用性測試**

**據ISO 9241的標準定義，使用性 (Usability) 為產品本身在特定的情境(context)下為特定的使用者使用，其所具有之有效性 (effectiveness) 、效率 (efficiency) 與滿足 (satisfaction)。簡言之，當你使用一項產品或服務，該產品 / 服務 是否有滿足使用者的目標 (goal)，使用情況是否為有效使用及有效率的，即為使用性 (Usability test)。**

使用性測試 (Usability test) 即是測試上述三個面向內容，檢視產品對於使用者的表現。然而，在不同產品上使用性的實際定義與測試，也有所不同。如若是在電商網站上，使用者在使用性上，可能為「使用者是否能找到想要的商品」、「結帳流程簡便快速」等；而若是生產力工具，如筆記軟體，對使用者而言，「易於擷取畫面」或「多種類的標示工具」，就可能會是使用性中的內容。總言之，使用性會因使用者使用的產品不同，體現在實際操作的任務 (task)，但使用性的定義仍相同。

**使用性測試流程**

一個簡單、完整的測試可以大致分為四個步驟，分別為「定義與設定」、「受測者招募」、「測試與觀察」、「結果分析」。依據規模與時程，測試中各步驟可隨之調整。

***定義與設定***

在測試前，先決定好本次測試內容的目標。兩種類別的使用性測試，「形構型測試」(formative)與「總結型測試」(summative)，兩者差別在於施測的時機。形構型測試施測於產品完成前，依據施測的結果可以提供給設計者，作為產品設計或改善的參考；相反地，總結型測試施測於產品完成之後，除了獲得使用者反饋外，也可將其他相類似的產品予以測試，作為產品間互相比較的一途。

　　在前篇文章有介紹任務 (task)，使用者在產品中執行任務進而滿足目標。舉例來說，若使用者使用音樂串流app，「找尋想要的音樂」，可視為其中之一個任務；而產品有提供甚麼任務，是設計者需考量的部分。在使用性測試中，即是以任務為測試單位；一個產品提供的任務眾多，在測試時需決定哪些任務為測試標的。

***受測者招募***

受測者將直接影響受測後的結果，因此在挑選上需依產品目標使用族群為基準。此基準參雜眾多面向，可從基本背景資料 (性別、年齡、教育程度等) 或 使用者能力下手，基準的設定越詳盡越能確保招募到符合的受測者。至於受測者數目多少最好？Thomas (2013)等人認為，6–8位即足夠，大部分使用性問題在前六位即可找到 ; 然而，如果受測者是有類別的，則每個類別至少要有4個。Nielsen (2000) 則認為，5位受測者即可找出大部分的問題，但總體數目可以視有無類別而增減。

***測試與觀察***

測試的環境，可盡量挑選干擾較小之房間或環境，避免影響受測進行。施測進行前，應該先對受測者做整個測試的過程、目的的說明，並解釋測試後的數據、個人資料的處理方式。而紀錄受測者與產品互動的過程，如若以app測試，則可將測試時的app畫面予以錄影 ; 受測者的表情、手勢亦同，在最後與受測者討論時，可拿出來加以討論。測試所使用的指標，也就是測試時執行任務時所用的測量標準，常用的如下 (Thomas, 2013 ; Jeff, 2016)：

1.任務完成 (task success) — 使用者是否有完成任務。透過有完成的受測佔比，即是該任務完成率。

2.任務時間 (task time) — 使用者完成任務所花費的時間。對大部分的產品來說，花費時間越短對使用者是正面的；然而對於一些產品並非如此，如遊戲、教育軟體等，時間越短的意義並非正面。

3.任務錯誤 (error) — 使用者執行任務時的非預期錯誤，如錯過、按錯按鈕，錯誤有時甚會導致任務的失敗。怎樣的情形算錯誤呢? 只要是會導致任務在執行上，效率低、嚴重的損失、任務失敗，就可以算做任務錯誤。在處理錯誤時，必須了解導致錯誤的前因後果，以利改善。

4.效率 (efficiency) — 效率綜合了多個面向，對大部分的產品來說，效率代表使用者在越短的時間內完成任務，錯誤的發生也少。效率的計算，可 以 任務完成率 / 任務平均花費時間 之百分比來表達；如任務B任務完成率為65，平均完成時間1.5分鐘，則任務B的效率為65/1.5，43%。

除了以上的計算方式，使用者若在任務中耗費的步驟越少，亦能提升效率；如在電商中，購買商品所需要的步驟越少，則總體時間可減少、錯誤的發生也可避免。

5.學習性 (learnability) — 初學者從對產品不熟到上手，即為使用者學習歷程，歷程的難易快慢為學習性；學習性差對於剛使用產品的使用者，可能因此障礙而放棄使用。學習性常用的計算方式，可以前述任務時間、錯誤、完成等來做為計算；將受測者安排一定時間的多次測試，來檢視指標變化。

如以任務時間來看，將同一受測者執行同任務，第1次測試完成時間為50秒、隔一個月再測為40、再一個月為35；待受測者表現時間趨於穩定，即是該任務最大的程度表現。藉由改善產品並再檢測指標的變化，再來看是否在學習性上有提升。

在測試正式開始時，情境(Scenario)建立可讓使用者理解背景，讓任務的執行更加合理，情境的篇幅精簡扼要即可。閱讀完後接著為任務描述，任務應切合產品與情境的脈絡，不可指示使用者如何操作。以訂票網站來舉例，情境與任務指示如下：

＊情境建立

　　「你與朋友為了下週一旅行籌備已久，你們決定要搭乘台鐵前往目的地。預計在下午三點左右從台北出發前往高雄，為了能方便聊天，你們不希望座位分開來坐，並且能越快到達高雄越好。」

＊任務

　　「請預訂下禮拜一下午三點左右，台北至高雄的車票共四張」

　　然而在執行過程中，受測者可能會遇到疑惑障礙而無法操作，此時可決定是否該提供協助。對於受測者因碰到疑惑障礙而無法繼續任務，又可被稱為使用性問題 (usability issue) ; 相較於上述以量化數值測量的指標，使用性問題為質性描述使用者執行上的問題。

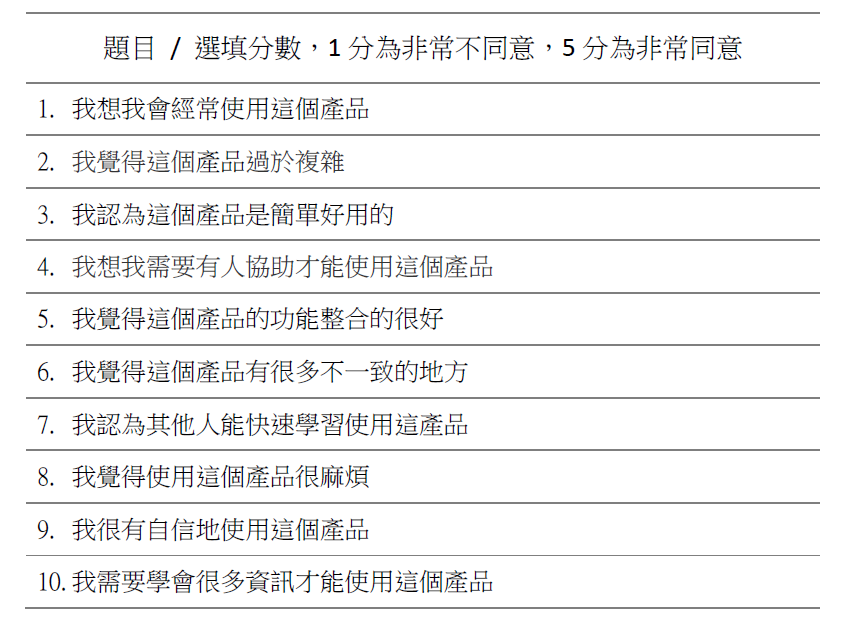
***結果分析***

　　完成測試後，測量數值的意義因產品的不同而有不同。如果是以電子商務性質來看，讓消費者無障礙的購買商品，相信是電商網站的目標 ; 此時「任務完成」、「效率」、「任務錯誤」三個指摽的意義重要性不言而喻。要檢視哪些指標，端看產品本身的目標與性質而定 ; 有了指標的數值後，可以再進行產品調整再測試一次，並將前後版本的數值對照來看是否有進步。

　　在使用性問題的處理上，首先需定義問題從開始到結束為何，了解是什麼原因導致問題的發生，而問題的性質又是甚麼，是屬於導引、內容或功能? 接著以問題發生的頻率與嚴重程度，此兩指標來決定使用性問題的優先處理順序。

　　在使用性的定義之中有提到，滿足 (satisfaction) 也是其中之一的要素。使用者的滿足，來自於使用產品完的總體印象，因此在測量滿足通常為使用後再行施測。要注意的是，產品的比較需要其性質需一致，不同的產品來比較是無意義。(Brooke,1996)

SUS (System Usability Scale) 量表是目前較為人知及運用，測量使用者滿足五點量表，題目共10題：



　　題目的分數計算，奇數題為正向題，將各題分數減一為最後分數，如若為3分計算結果為2分；偶數題為負向題，以5分減去該題分數為最後分數，如若為4分計算結果為1分。各題分數計算完畢後，將所有分數加總乘以2.5，即得到總分。Jeff (2011) 指出，SUS的平均分數為68分，因此可以將此分數作為產品及格分數。

**結語**

　　使用性測試提供設計者評估產品的途經，然而如果不知產品本身的功能與測試的目的，使用性測試的作用僅只是徒然。此外，使用性測試中，符合商業目標亦為重要；在執行測試前，應先諮詢其他利害關係人的目標需求，一併考量於測試之中。

參考資料：

1. Albert, W., & Tullis, T. (2013). Measuring the user experience: collecting, analyzing, and presenting usability metrics. Newnes.

2. Sauro, J., & Lewis, J. R. (2016). Quantifying the user experience: Practical statistics for user research. Morgan Kaufmann.

3. Sauro, J. (2011). Sustisfied? little-known system usability scale facts. UX Magazine, 10(3), 2011–3.

4. Affairs, A. S. (2014, May 15). Running a Usability Test. Retrieved January 29, 2018, from https://www.usability.gov/how-to-and-tools/methods/running-usability-tests.html

5. Nielsen, J. (2000). Why you only need to test with 5 users.