Sertifikat

No. 1596/DAK-25/S/VII/2021

Penerbit Deepublish

Memberikan penghargaan setinggi-tingginya kepada:

Dr. Tenia Wahyuningrum, S.Kom., M.T.

sebagai penulis buku yang berjudul

"Buku Referensi Mengukur Usability Perangkat Lunak"

dengan nomor ISBN **978-623-02-3154-4**

yang telah diterbitkan di Deepublish pada tahun 2021.

Yogyakarta, 30 Juli 2021
CEO Deepublish

deepublish publisher
(Penerbit Buku Pendidikan Anggota IKAPI)

An Nuur Budi Utama

C5

Buku Referensi

Mengukur Usability Perangkat Lunak

Penjaminan kualitas produk perangkat lunak merupakan faktor penting dalam bisnis. Kualitas produk perangkat lunak ditentukan oleh seberapa banyak produk memenuhi kebutuhan pengguna, seberapa banyak produk melakukan kinerjanya, dan seberapa banyak jumlah cacat dalam perangkat lunak tersebut. Oleh karena itu, perlu dilakukan pemeriksaan kualitas sebelum produk perangkat lunak dipasarkan, sebab perbaikan setelah pengiriman dinilai sangat mahal dan mempengaruhi kredibilitas perusahaan. Salah satu karakteristik kualitas perangkat lunak yang berkaitan dengan pengguna adalah *usability*.

Usability menilai sejauh mana produk atau sistem dapat digunakan oleh pengguna tertentu, untuk mencapai tujuan tertentu. Buku ini menjelaskan bagaimana cara mengukur *usability* sebuah perangkat lunak, mulai dari perencanaan pengukuran, membuat *dashboard* perencanaan, mempersiapkan ruangan pengukuran, memilih partisipan, dan prosedur pengukuran. Buku ini juga dilengkapi dengan studi kasus, sehingga pembaca dapat menerapkan teori dalam buku ke dalam praktik.

Penerbit Deepublish (CV BUDI UTAMA)
Jl. Kaliurang Km 9,3 Yogyakarta 55581
Telp/Fax : (0274) 4533427
Anggota IKAPI (076/DIY/2012)
cs@deepublish.co.id
Penerbit Deepublish
@penerbitbuku_deepublish
www.penerbitdeepublish.com

Kategori :

Buku Referensi
Mengukur Usability Perangkat Lunak
Dr. Tenia Wahyuningrum, S.Kom., M.T.

Buku Referensi

Mengukur Usability Perangkat Lunak

Dr. Tenia Wahyuningrum, S.Kom., M.T.

Buku Referensi

MENGUKUR USABILITY PERANGKAT LUNAK

deepublish | publisher

UU No 28 tahun 2014 tentang Hak Cipta

Fungsi dan sifat hak cipta Pasal 4

Hak Cipta sebagaimana dimaksud dalam Pasal 3 huruf a merupakan hak eksklusif yang terdiri atas hak moral dan hak ekonomi.

Pembatasan Pelindungan Pasal 26

Ketentuan sebagaimana dimaksud dalam Pasal 23, Pasal 24, dan Pasal 25 tidak berlaku terhadap:

i. Penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;
ii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;
iii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan
iv. Penggunaan untuk kepentingan pendidikan dan pengembangan ilmu pengetahuan yang memungkinkan suatu Ciptaan dan/atau produk Hak Terkait dapat digunakan tanpa izin Pelaku Pertunjukan, Produser Fonogram, atau Lembaga Penyiaran.

Sanksi Pelanggaran Pasal 113

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

deepublish publisher

Dr. Tenia Wahyuningrum, S.Kom., M.T.

Buku Referensi

MENGUKUR USABILITY PERANGKAT LUNAK

deepublish
glorify and develop the intellectual of human's life

deepublish | publisher

BUKU REFERENSI MENGUKUR USABILITY PERANGKAT LUNAK

Tenia Wahyuningrum

Desain Cover :
Herlambang Rahmadhani

Sumber :
https://www.shutterstock.com

Tata Letak :
Gofur Dyah Ayu

Proofreader :
Avinda Yuda Wati

Ukuran :
x, 85 hlm, Uk: 15.5x23 cm

ISBN :
No ISBN

Cetakan Pertama :
Bulan 2021

Hak Cipta 2021, Pada Penulis

Isi diluar tanggung jawab percetakan

Copyright © 2021 by Deepublish Publisher
All Right Reserved

Hak cipta dilindungi undang-undang
Dilarang keras menerjemahkan, memfotokopi, atau
memperbanyak sebagian atau seluruh isi buku ini
tanpa izin tertulis dari Penerbit.

PENERBIT DEEPUBLISH
(Grup Penerbitan CV BUDI UTAMA)
Anggota IKAPI (076/DIY/2012)

Jl.Rajawali, G. Elang 6, No 3, Drono, Sardonoharjo, Ngaglik, Sleman
Jl.Kaliurang Km.9,3 – Yogyakarta 55581
Telp/Faks: (0274) 4533427
Website: www.deepublish.co.id
www.penerbitdeepublish.com
E-mail: cs@deepublish.co.id

deepublish publisher

KATA PENGANTAR

Alhamdulillah, segala puji dan syukur penulis panjatkan ke hadirat Allah Swt. atas rahmat dan karunia-Nya, buku ini telah selesai disusun. Buku ini disusun agar dapat membantu para mahasiswa dalam mempelajari pengukuran *usability* pada perangkat lunak. Buku ini menjelaskan tentang pengertian *usability*, cara mengukur, persiapan pengukuran, memilih partisipan disertai dengan studi kasus. Penulis pun menyadari jika di dalam penyusunan buku ini mempunyai kekurangan, namun penulis meyakini sepenuhnya bahwa sekecil apapun buku ini tetap akan memberikan sebuah manfaat bagi pembaca. Akhir kata untuk penyempurnaan buku ini, maka kritik dan saran dari pembaca sangatlah berguna untuk penulis ke depan.

Tenia Wahyuningrum

deepublish | publisher

DAFTAR ISI

KATA PENGANTAR v

DAFTAR ISI vi

DAFTAR GAMBAR viii

DAFTAR TABEL x

BAB I PENDAHULUAN 1

A. Perbedaan user interface, user experience dan usability 1

B. *Usability* bagian dari kualitas perangkat lunak 5

BAB II USABILITY 10

BAB III UKURAN *USABILITY* 21

A. Perencanaan pengukuran 22

Pengukuran kinerja 23

Pengukuran kepuasan 23

Dashboard perencanaan 24

Prosedur pengukuran 25

Persiapan ruangan 26

B. Memilih Partisipan 31

Jumlah partisipan 32

Cukup lima partisipan 32

Lima Partisipan Tidak Cukup 34

Cara merekrut partisipan 35

BAB IV UKURAN KINERJA 37

A. Efektivitas 39

Completion Rate 39

Error rate 41

deepublish | publisher

B. Efisiensi 44
Time-based efficiency 44
Overall relative efficiency 45

BAB V UKURAN KEPUASAN 48
A. Skala Likert 48
B. Analisis kepuasan pengguna 49
Task Level Satisfaction 49
Test Level Satisfaction 52
System Usability Scale Questionnaire (SUS) 52
Usability Metric for User Experience (UMUX) 56

BAB VI STUDI KASUS 59
A. Pemilihan partisipan 59
Kriteria usability expert 59
Kriteria intermittent user 60
B. Metode pengukuran *usability* 61
C. Pengukuran oleh *expert* 62
Langkah 1: 62
Langkah 2: 68
Langkah 3: 69
D. Pengukuran oleh *intermittent user* 70
Langkah 1: 70
Langkah 2: 70
Langkah 3: 71
Langkah 4: 71
Pengukuran validitas 72
Pengukuran reliabilitas 73
Skor SUS 74

LAMPIRAN 77

DAFTAR PUSTAKA 82

deepublish publisher

DAFTAR GAMBAR

Gambar 1. Kaleng susu kental manis 1
Gambar 2. Kaleng soda 2
Gambar 3. Kemasan susu kental manis Frisian Flag 3
Gambar 4. Kemasan tutup anti tumpah 4
Gambar 5. User interface (UI), user experience (UX), dan usability produk 4
Gambar 6. User interface (UI) button login 6
Gambar 7. User experience pada call to action button 7
Gambar 8. Hirarki button 8
Gambar 9. Visibility of system status 9
Gambar 10. Metode evaluasi usability 16
Gambar 11. Prosedur pengukuran usability 26
Gambar 12. Contoh layout ruangan usability tipe single room 27
Gambar 13. Contoh layout classical usability room 29
Gambar 14. Partisipan dan masalah usability 33
Gambar 15. Completion Rate 41
Gambar 16. Error rate 43
Gambar 17. Metode pengukuran usability 61
Gambar 18. Progress bar saat menyalin file 62
Gambar 19. Password pada live-mail 63

deepublish | publisher

Gambar 20. Folder/map di dunia nyata dan di dalam sistem 63
Gambar 21. Tombol validasi di Ms. Excel 64
Gambar 22. Navigasi bar pada Microsoft Excel, Word, dan PowerPoint 65
Gambar 23. Autofill Microsoft Excel 65
Gambar 24. Keyboard untuk menavigasi halaman 67
Gambar 25. Pesan kesalahan pada pengisian form 68

deepublish | publisher

DAFTAR TABEL

Tabel 1. Definisi usability 10
Tabel 2. Karakteristik usability 11
Tabel 3. Pemetaan terminologi dalam usability 12
Tabel 4. Pemetaan terminologi dalam usability (lanjutan) 13
Tabel 5. Jenis kuesioner usability (Perlman, 2016) 20
Tabel 6. Dashboard perencanaan (diadopsi dari Usability Test Plan Dashboard oleh David Travis) 24
Tabel 7. Sauro Lewis CGS 55
Tabel 8. Contoh jawaban responden SUS 56
Tabel 9. Contoh jawaban responden UMUX 57
Tabel 10. Tabel kriteria severity 69
Tabel 11. Hasil rata-rata penilaian ketiga expert 70
Tabel 12. Hasil tabulasi kuesioner SUS 71
Tabel 13. Hasil tabulasi kuesioner SUS 72

deepublish | publisher

BAB I
PENDAHULUAN

"Design is not just what it looks like and feels like.
Design is how it works – Steve Jobs*"*

A. Perbedaan user interface, user experience dan usability

Apakah perbedaan *user interface*, *user experience* dan *usability*? Ketiga istilah ini seringkali digunakan para pengembang dan desainer perangkat lunak, tapi, apakah kalian tahu bedanya? Agar memudahkan pemahaman mengenai ketiga istilah tersebut, mari kita lihat Gambar 1.

Gambar 1. Kaleng susu kental manis
(sumber: https://stock.adobe.com/search?k=condensed%20milk%20can)

Gambar 1 menunjukkan kemasan kaleng *condensed milk* atau susu kental manis, lalu perhatikan bagaimana kaleng tersebut dalam keadaan terbuka dan tertutup. Bagi penggemar susu kental manis di Indonesia, tentu tidak asing dengan *brand* susu bendera atau "Frisian Flag". Sebelum tahun 2015, kemasan kaleng susu bendera cukup sulit di buka, sama seperti kemasan susu kental

deepublish | publisher

manis pada umumnya (lihat Gambar 1). Konsumen biasanya akan menggunakan ujung pisau untuk melubangi sisi atas kaleng saat akan menuangkan isinya. Bahkan di pasaran, dijual pula pisau khusus untuk membuka kaleng jenis ini. Hal ini tentu dirasa cukup merepotkan bagi para konsumen. Sejak puluhan tahun berdiri, nampaknya produsen susu kental manis “Frisian Flag” tidak banyak berinovasi pada kemasan kaleng ini. Akan tetapi, setelah 2015-an, produsen susu mengadopsi kaleng soda pada bagian atas kemasannya, sehingga memudahkan para pelanggan saat membuka dan menuangkan susu. Inovasi ini disebut dengan metode “Klik Tarik Tuang” (KTT). Gambar 2 memperlihatkan kemasan kaleng soda yang cukup mudah dibuka hanya dengan menarik *tab ring.*

Gambar 2. Kaleng soda
(sumber: https://www.vectorstock.com/royalty-free-vector/soda-can-icon-flat-design-vector-27912969)

Kaleng soda biasanya terbuat dari wadah logam yang terbuat dari alumunium atau baja berlapis timah. Kemasan kaleng ini memiliki sejarah yang cukup panjang, dimulai sejak dipasarkannya bir komersial pertama pada tahun 1935 di Virginia (Maxwell, 1993). Tidak lama kemudian, minuman bersoda juga tersedia dalam kaleng. Soda yang memiliki keasaman dan tekanan lebih tinggi, biasanya memiliki pelapis pada bagian dalam, untuk menjaga rasa produk tidak hancur oleh reaksi kimia dengan logam. Belakangan,

deepublish publisher

"StaTab" -alat pembuka kaleng- diperkenalkan oleh perusahaan aluminium Reynolds tahun 1983 dan telah menjadi standar minuman kaleng saat ini.

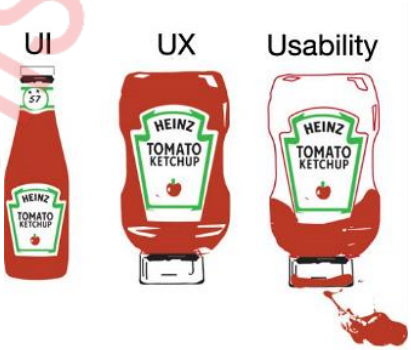
Gambar 3. Kemasan susu kental manis Frisian Flag
(sumber: iklan di Google)

Gambar 3 menunjukkan kemasan susu kental manis merek “Frisian Flag” yang telah mengalami perubahan pada desain bagian atasnya. Setelah muncul inovasi tutup kaleng dengan sekali klik, mereka berpikir kembali tentang bagaimana cara menyimpannya agar terhindar dari semut dan serangga lain yang masuk dalam kemasan. Biasanya para konsumen susu akan menutup dengan tutup gelas, dan memasukkan kaleng ke dalam mangkuk berisi air. Hal ini dirasa masih cukup sulit dan merepotkan, hingga akhirnya, kemudian muncul terobosan baru dalam kemasannya, tutup kaleng dari plastik, yang dapat ditutup dengan rapat setelah dibuka. Plastik klip merupakan desain baru kemasan yang memberikan nilai lebih pada kaleng, dan memberikan rasa nyaman dan puas pada pengguna. Dengan demikian, kaleng susu kental manis “Frisian Flag” menjadi lebih mudah dibuka, lebih mudah disimpan, anti tumpah, anti berceceran dan lebih higienis.

deepublish publisher

Gambar 4. Kemasan tutup anti tumpah
(sumber: Iklan di Google)

Inovasi yang lahir dari *user experience*, itulah *usability*. Hal ini terus menerus dilakukan, agar produk dapat digunakan semaksimal mungkin. *Usability testing* merupakan salah satu pengujian perangkat lunak non-fungsional yang melibatkan unsur interaksi manusia dalam penilaiannya. Terdapat beberapa istilah dalam desain interaksi manusia dan komputer, antara lain *user interface, user experience* dan *usability*. Gambar 5 menjelaskan perbedaan antara ketiganya.



Gambar 5. User interface (UI), user experience (UX), dan usability produk
(sumber: https://measuringu.com/evaluating-ux/)

deepublish publisher

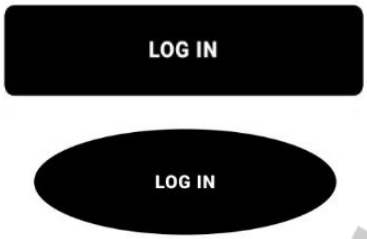
Gambar 5 menjelaskan bahwa pada awalnya, desain saus tomat merek “Heinz” menggunakan botol kaca, dengan tutup dari bahan alumunium yang dapat dilepas dari botolnya. Setelah beberapa lama dengan desain tersebut, pengguna merasa kesulitan dalam menuangkan saus tomat, terutama bila isinya tinggal sedikit. Kemudian, berdasarkan masukan dan pengalaman pengguna, produsen “Heinz” mengubah desain botolnya dengan botol plastik, yang tutupnya menyatu dengan botolnya, serta disimpan dalam bentuk terbalik (tutup berada di bawah). Hal ini kemudian memudahkan pengguna dalam memakai produk, sehingga pekerjaan menjadi lebih mudah. Dalam ilustrasi di atas, dapat dijelaskan bahwa *user interface* (UI) berhubungan dengan tampilan produk sesuai dengan fungsinya. Para desainer UI memikirkan bagaimana warna tampilan, bentuk interaksi, ukuran tombol, dsb. Sedangkan *user experience* (UX) muncul akibat dari persepsi dan respons dari pengguna sebagai reaksi dari penggunaan sebuah produk. UX menekankan pada apa yang pengguna rasakan pada saat menggunakan produk tersebut, bagaimana produk tersebut dapat membangkitkan perasaan emosional pengguna dan bagaimana sebuah produk dapat meningkatkan nilai dari tugas yang dikerjakan. *Usability*, yang merupakan bagian dari *user experience*, memastikan pengguna dapat memakai produk dengan efektif, efisien dan merasa nyaman serta puas.

B. *Usability* bagian dari kualitas perangkat lunak

Penjaminan kualitas produk perangkat lunak merupakan faktor penting dalam bisnis (El-Rayyes & Abu-Zaid, 2012). Kualitas produk perangkat lunak ditentukan oleh seberapa banyak produk memenuhi kebutuhan pengguna, seberapa banyak produk melakukan kinerjanya, dan seberapa banyak jumlah cacat dalam perangkat lunak tersebut (Sowunmi, 2015). Oleh karena itu, perlu dilakukan pemeriksaan kualitas sebelum produk perangkat lunak dipasarkan, sebab perbaikan setelah pengiriman dinilai sangat mahal dan mempengaruhi kredibilitas perusahaan (Shah, 2012).

deepublish | publisher

Salah satu karakteristik kualitas perangkat lunak yang berkaitan dengan pengguna adalah *usability*. Gambar 6 merupakan contoh dari UI pada tampilan perangkat lunak.



(a) *rounded rectangle*

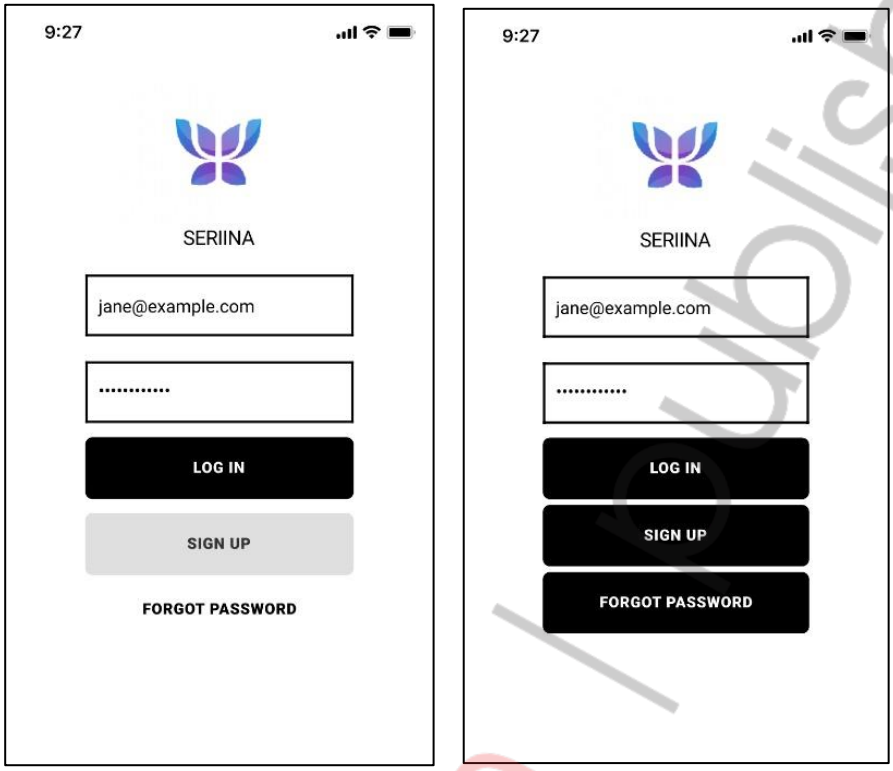
(b) *oval*

Gambar 6. User interface (UI) button login

Saat merancang sebuah tampilan antarmuka *button login* pada aplikasi atau situs web, para desainer UI memikirkan apa yang dapat dilihat pengguna pada tombol tersebut. Bagaimana bentuknya? Segiempat atau oval? Apa warnanya? Putih atau hitam? Bagaimana tombol bekerja? Desain tombol, kolom entri, dan tata letak aplikasi atau situs web adalah tempat pengguna berinteraksi dengan sistem. Jadi singkatnya desain UI adalah desain titik interaksi di mana pengguna dapat berkomunikasi dengan sistem.

Para desainer UX lebih memikirkan bagaimana sebuah *button* dapat bekerja dengan baik. Apakah tombol "*forgot password*" sebaiknya dibuat dengan sebuah *link* tulisan saja? Ataukah dibuat sama bentuknya dengan tombol "*log in*"? UX adalah pengalaman keseluruhan pengguna akhir agar aplikasi dapat bekerja lebih baik lagi.

deepublish publisher



(a) (b)

Gambar 7. User experience pada call to action button

Gambar 7 menunjukkan sebuah tampilan susunan *button* pada antarmuka aplikasi untuk meningkatkan pengalaman pengguna. Saat menggunakan aplikasi, pengguna mungkin menemukan ratusan tombol yang meminta untuk melakukan suatu tindakan-apakah itu berlangganan pemberitahuan berita instan melalui *e-mail*, untuk mendaftar aplikasi, atau untuk membeli kaus baru yang sudah lama ditunggu-tunggu. Dalam dunia pemasaran *online*, tombol-tombol ini sering disebut sebagai *call to action* (CTA) *button*. Hierarki *button* yang dirancang pada aplikasi disusun berdasarkan tingkat kepentingannya. Gambar 7 (a) memperlihatkan desain *button* yang lebih baik dari desain (b). Hal ini dapat dilihat bahwa aturan pembuatan *button* tidak perlu untuk menampilkan

deepublish | publisher

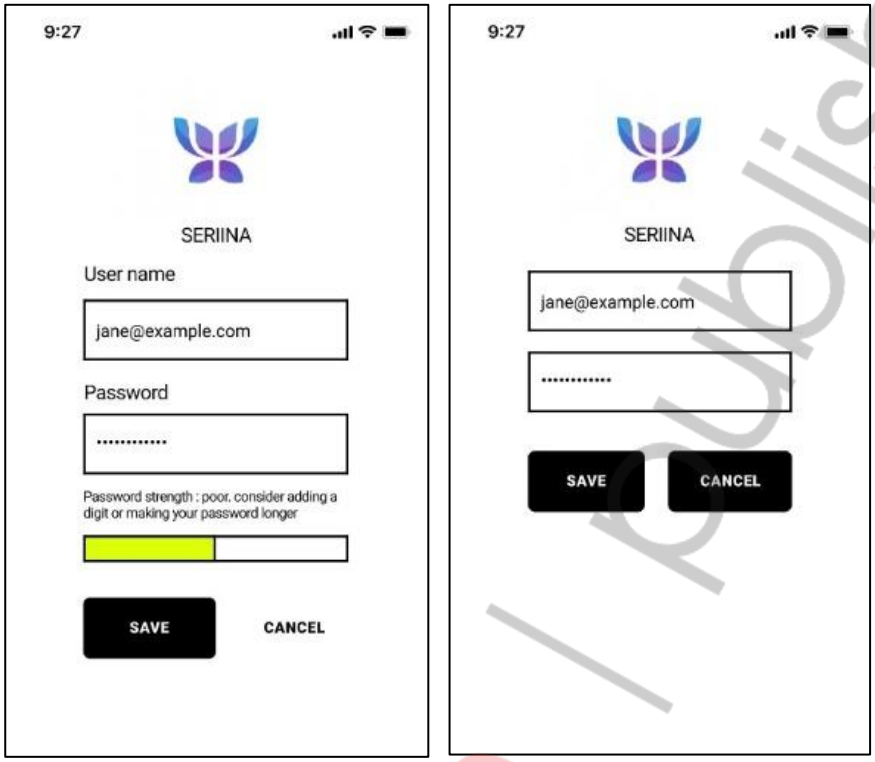
background untuk setiap *button*, sehingga pengguna akan fokus pada *button* yang lebih penting.

Gambar 8. Hirarki button

Gambar 8 menjelaskan hierarki visual *button* yang dapat dilihat pada tingkat prioritas, yaitu *primary*, *secondary* dan *tertiary*. Desainer akan memperhatikan pengalaman pengguna pada saat dihadapkan pada sebuah tampilan. Pada halaman awal aplikasi, *button log in* adalah tombol yang memiliki tingkat kepentingan tertinggi, sedangkan *forgot password* menjadi tombol dengan tingkat kepentingan paling rendah, sehingga tidak memerlukan *background* dan *line*.

Usability merupakan bagian dari UX. *Usability* adalah persepsi pengguna akhir tentang bagaimana seseorang dapat secara efektif, efisien dan puas dalam menyelesaikan tugas saat menggunakan aplikasi. Gambar 9 menunjukkan *visibility of system status*. Desain 9 (a) dapat dikatakan lebih baik dari pada 9 (b), karena sistem yang baik akan memberi tahu pada pengguna, sejauh mana tingkat kerumitan *password*, sehingga pengguna mendapatkan gambaran keamanan akunnya.

deepublish publisher



Gambar 9. Visibility of system status

deepublish | publisher

BAB II
USABILITY

"Usability is about people and how they understand and use things, not about technology – Steve Krug*"*

Usability telah banyak di definisikan dalam beberapa penelitian dengan banyak cara. Definisi *usability* menurut Nielsen, yaitu konteks penerimaan sistem secara keseluruhan (Jacob Nielsen, 1993). Dengan demikian, *usability* memastikan produk mudah dipelajari, efektif digunakan, dan menyenangkan pengguna (ShaRp*et al.*, 2002). Berdasarkan ISO 25010, *usability* menilai sejauh mana produk atau sistem dapat digunakan oleh pengguna tertentu, untuk mencapai tujuan tertentu (International Organization for Standardization, 2011). Sedangkan penelitian lain mengungkapkan, *usability* merupakan tujuan akhir dari desain (Dix *et al.*, 2004). Tabel 1 menunjukkan definisi *usability* berdasarkan pendapat masing-masing peneliti.

Tabel 1. Definisi usability

Peneliti	Definisi
Sharp *et al.* (2002)	*Usability* memastikan bahwa produk mudah dipelajari, efektif digunakan, dan menyenangkan dari perspektif pengguna
Dix *et al.* (2004)	*Usability* merupakan tujuan akhir dari desain interaktif sistem
	Keberhasilan sebuah produk, mencakup tiga kata "guna", yaitu "berguna", "dapat digunakan" dan "digunakan"
	Berguna: memenuhi apa yang dibutuhkan penggunanya, misalkan menjalankan musik, memasak makan malam, melakukan format dokumen

deepublish publisher

Peneliti	Definisi
	Dapat digunakan: pengguna dapat menggunakannya dengan mudah dan alami, misalkan tanpa kesalahan berbahaya
	Digunakan: membuat orang ingin menggunakannya, menarik, menyenangkan
ISO 25010 (2011)	Sejauh mana produk atau sistem dapat digunakan oleh pengguna tertentu untuk mencapai tujuan tertentu dengan efektivitas, efisiensi dan kepuasan dalam konteks tertentu digunakan
	Usability dapat diukur sebagai karakteristik kualitas produk, maupun diukur secara langsung yang merupakan bagian dari kualitas penggunaan
Nielsen (1993)	*Usability* adalah atribut kualitas yang menilai betapa mudahnya antarmuka yang digunakan. Kata "kegunaan" juga mengacu pada metode untuk meningkatkan kemudahan penggunaan selama proses desain

Tidak ada aturan atau kesepakatan khusus mengenai karakteristik *usability*, sehingga, setiap peneliti dan *developer* perangkat lunak dapat lebih fleksibel dalam menentukan parameter yang akan diuji. Akan tetapi, pada umumnya, pengukuran *usability* memiliki karakteristik seperti yang dijelaskan pada Tabel 2.

Tabel 2. Karakteristik usability

Karakteristik	Nielsen (1993)	Sharp *et al.* (2002)	Dix *et al.* (2004)	ISO 25010 (2011)
Effectiveness		√		√
Efficiency	√	√		√
Safety		√		
Utility		√		
Learnability	√	√	√	√
Memorability	√	√		
Flexibility			√	
Robustness			√	
Appropriateness recognizability				√
User Error Protection				√
User Interface aesthetics				√
Accessibility				√

deepublish publisher

Karakteristik	Nielsen (1993)	Sharp *et al.* (2002)	Dix *et al.* (2004)	ISO 25010 (2011)
Satisfaction	√			√
Errors	√			
Operability				√
Jumlah karakteristik	**5**	**6**	**3**	**9**

ISO 25010 atau yang biasa disebut ISO SQuaRE (*Software Quality Requirement and Evaluation*) merupakan perbaikan dari model sebelumnya, yaitu ISO 9126. Dari Tabel 2 dapat dilihat, bahwa model ISO 25010 memiliki jumlah karakteristik yang paling banyak. Meskipun demikian, beberapa peneliti dapat mengurangi atau menambahkan karakteristik di luar standar yang ditetapkan ISO. Penambahan tersebut dilakukan berdasarkan kebutuhan pengguna dan sistem perangkat lunak. Perbedaan pemilihan terminologi karakteristik *usability* beserta deskripsinya dipetakan pada Tabel 3 berikut ini (Wahyuningrum & Azhari, 2017).

Tabel 3. Pemetaan terminologi dalam usability

No	Terminologi/Sinonim	Deskripsi
1	*Learnability, ease to learn*	Tingkat kemudahan pengguna dalam mempelajari sebuah sistem, menyelesaikan tugas dasar untuk pertama kali, besarnya usaha dan waktu yang diperlukan, sehingga pengguna baru dapat mencapai interaksi efektif dan mendapatkan performa maksimal
2	*Effectiveness, effective to use*	Tingkat keakuratan dan kelengkapan sistem dalam mencapai tujuan tertentu.
3	*Efficiency, efficient to use*	Tingkat kecepatan dan sumber daya yang dibutuhkan pengguna dalam menyelesaikan tugas.
4	*Memorability, ease to remember*	Tingkat kemudahan untuk diingat, dalam penggunaan sistem, meskipun baru dipelajari sekali, atau tidak digunakan dalam jangka waktu tertentu.

deepublish | publisher

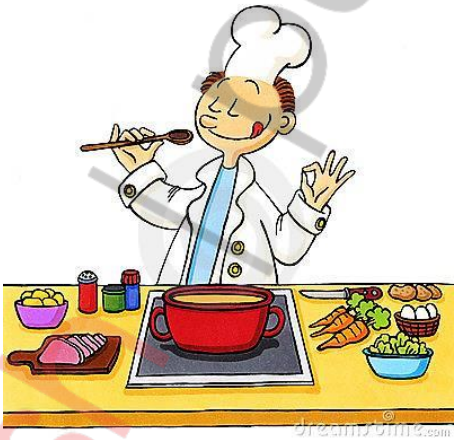
Tabel 4. Pemetaan terminologi dalam usability (lanjutan)

No	Terminologi/Sinonim	Deskripsi
5	*User error protection, safe to use, safety*	Tingkat perlindungan pengguna dalam melakukan kesalahan, keadaan bahaya, atau situasi yang tidak diinginkan.
6	*Errors*	Jumlah kesalahan yang dibuat pengguna, jenis kesalahan (kesalahan berat atau ringan), dan tingkat kemudahan sistem dalam menanggulangi kesalahan.
7	*Satisfaction*	Tingkat kepuasan pengguna dalam penggunaan produk atau sistem pada konteks tertentu, disesuaikan dengan kebutuhan pengguna.
8	*User Interface aesthetics, satisfaction*	Tingkat kepuasan dan kenyamanan pengguna dalam berinteraksi dengan sistem.
9	*Utility*	Tingkat ketersediaan sistem secara fungsional sehingga pengguna dapat melakukan apa yang mereka butuhkan dan inginkan
10	*Flexibility*	Banyaknya cara di mana pengguna dapat saling bertukar informasi
11	*Accessibility, technical accessibility*	Tingkat jangkauan pengguna dan kemampuan untuk mencapai tujuan tertentu, berisi semua atribut yang memungkinkan pengguna untuk menjalakan sistem
12	*Robustness*	Tingkat dukungan yang diberikan pengguna dalam menentukan keberhasilan pencapaian dan penilaian dari tujuan
13	*Appropriateness recognizability*	Tingkat pengenalan pengguna terhadap sistem yang sesuai dengan kebutuhan pengguna
14	*Operability, ease of operation, ease of use, ease of interaction*	Tingkat kemudahan sistem untuk dioperasikan dan di kendalikan.
15	*Understand ability, simplicity, comprehensibility*	Kemampuan untuk dipahami, fitur pendekatan yang jelas, tanpa ambiguitas.
16	*Attractiveness*	Tingkat ketertarikan dan penerimaan pengguna terhadap fitur dan perangkat sistem.

deepublish publisher

No	Terminologi/Sinonim	Deskripsi
17	*Interpretability*	Fitur pendekatan diatur ke dalam unit sesuai informasi untuk keterampilan pengguna.
18	*Helpfulness, help*	Ketersediaan bantuan ketika pengguna membutuhkannya.

Menurut Jacobsen (1999), metode evaluasi *usability* dibagi menjadi empat macam berdasarkan tujuan evaluasinya, yaitu *exploratory*, *predictive*, *formative* dan *summative*. Analogi dari keempat metode tersebut disebut analogi mencicipi sup (Stake, 1976).



"When the cook tastes other cooks' soups, that's exploratory.
When the cook assesses a certain recipe, that's predictive.
When the cook tastes the soup while making it, that's formative.
When the guests (*or food critics*) *taste the soup, that's summative."*

1. ***Exploratory***–bagaimana penggunaannya?
   - Dilakukan sebelum pembangunan antarmuka,
   - Mempelajari bagaimana *software* digunakan, seberapa sering dan untuk apa,
   - Mengumpulkan data yang digunakan–catatan statistik dan observasi dari penggunaan.

deepublisher publisher

2. ***Predictive***–memprediksi (akan) seberapa baik?
   - Memperkirakan seluruh kualitas dari antarmuka (seperti *summative evaluation*, tapi prediksi dilakukan di awal,
   - Dilakukan sekali saat desain selesai dibuat, tapi sebelum proses implementasi.

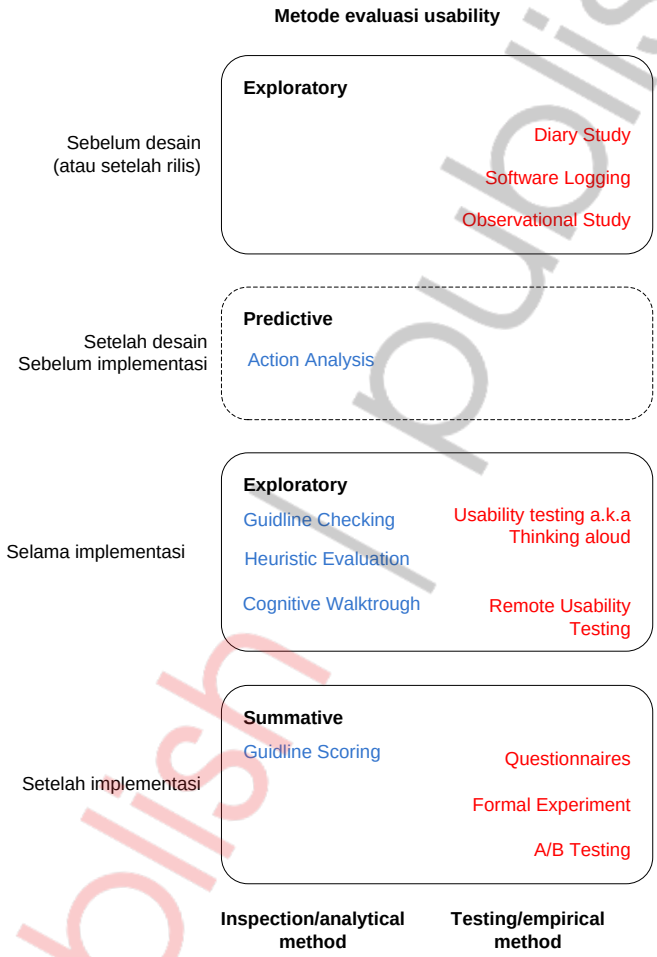
3. ***Formative***–bagaimana dapat menjadi lebih baik?
   - Memberitahukan proses desain dan membantu meningkatkan antarmuka selama desain;
   - Dilakukan selama pembangunan antarmuka;
   - Mempelajari mengapa sesuatu menjadi salah, tidak hanya pada saat terjadi kesalahan;
   - Mengumpulkan data proses-pengamatan kualitatif dari apa yang terjadi dan bagaimana.

4. ***Summative***–seberapa baik?
   - Penilaian dari semua kualitas dari antarmuka;
   - Dilakukan satu kali pada antarmuka (kurang lebih) telah selesai;
   - Membandingkan desain alternatif, atau menguji kebutuhan performansi yang telah didefinisikan.

Berdasarkan pengujinya, variasi metode evaluasi *usability* dikategorikan menjadi dua, yaitu analitik dan empiris. Evaluasi *usability* analitik atau dapat disebut inspection method merupakan teknik evaluasi yang melibatkan para pakar diantaranya adalah *Heuristic Evaluation, Cognitive Walktrough, Guidelines*, dan sebagainya (Gray & Salzman, 1998). Evaluasi *usability* empiris dilakukan berdasarkan pada pengalaman pengguna, seperti *Usability Testing* atau *Thinking Aloud, User Performance Test, Remote Usability Testing, Beta Test, Forum Test, Cooperative evaluation, Coaching Method*. Selain itu, terdapat pula evaluasi *usability* dengan melibatkan pernyataan pengguna seperti *User Satisfaction Questionnaire, Field Observation, Focus Group, Interviews* (Jacobsen, 1999). Beberapa metode sudah jarang

deepublish | publisher

digunakan karena terlalu kompleks dan tidak efisien. Secara umum, metode evaluasi usability dapat dijelaskan pada gambar 10.



Gambar 1. Metode evaluasi usability

Usability Testing

Usability Testing merupakan pengujian yang umum digunakan untuk melakukan evaluasi antarmuka pengguna. Metode ini biasanya dilakukan menggunakan *think aloud protocol* (Hom,

deepublish | publisher

1998). Pengguna diberikan beberapa tugas dalam lingkungan pengujian dan didorong untuk berpikir keras ketika menyelesaikan tugas-tugas tersebut. Dari pengujian tersebut, muncul informasi yang dibutuhkan tentang bagaimana desain antarmuka pengguna yang sesuai dengan cara alami manusia dalam berpikir (Kaikkonen & Kallio, 2005).

Remote Usability Testing

Remote Usability Testing merupakan pengujian yang memisahkan evaluator dan pengguna dalam ruang dan atau waktu yang berbeda. Metode *Remote Usability Testing* terbagi menjadi *synchronous* dan *asynchronous* (Riku Luostarinen, et.al, 2016). Metode *synchronous*, memisahkan pengguna dan evaluator secara spasial (ruang) dan temporal (waktu), pada prinsipnya sama dengan usability testing dalam laboratorium, dan menunjukkan kesamaan dengan metode konvensional. Sedangkan pada metode asynchronous, pengguna dan evaluator terpisah ruang saja, sehingga metode tersebut dinilai tidak efisien dalam waktu dan identifikasi masalah *usability* (Andreasen, et.al, 2007).

A/B Testing

Pengujian A/B (juga dikenal dengan pengujian bucket atau pengujian split run) merupakan sebuah metode untuk membandingkan dua jenis dari tampilan aplikasi, sistem informasi, website maupun produk digital lainnya. Pendekatan metode ini banyak digunakan karena data yang diperoleh dari pengujian A/B memberikan bukti langsung tentang perilaku pengguna dalam mendukung konsep “perpetual beta” (Young, 2014). Umpan balik pengguna yang melalui proses desain berulang terus menerus berupaya untuk meningkatkan pengalaman pengguna terhadap produk tersebut. Pengujian A/B dirancang untuk mengungkapkan solusi optimal di antara serangkaian solusi alternatif yang telah ditetapkan, cocok untuk penelitian yang telah ditentukan pertanyaan desain yang sudah jelas dengan hasil yang terukur secara kuantitatif.

deepublish publisher

Diary study

Diary study digunakan untuk mengumpulkan data kualitatif yang berkaitan dengan perilaku, aktifitas dan pengalaman pengguna dari waktu ke waktu, pada sebuah periode waktu tertentu (Salazar, 2016). Para partisipan yang terlibat dalam diary study diminta untuk menuliskan informasi yang spesifik tentang kegiatan yang sedang dipelajari dalam sebuah buku catatan. Para peneliti akan mengingatkan para partisipan untuk melakukan hal tersebut secara rutin. Fokus diary study salah satunya mencakup hal-hal seperti produk atau situs web, perilaku pengguna *smartphone,* aktivitas umum seperti berbagi informasi melalui sosial media, atau aktivitas yang spesifik seperti merencanakan liburan, membeli perabot elektronik, dll. Diary study memerlukan banyak waktu dan tenaga dibandingkan dengan metode lainnya, tetapi menghasilkan informasi yang sangat rinci terhadap perilaku dan pengalaman kehidupan nyata pengguna secara *real-time*.

Heuristic Evaluation

Heuristic Evaluation merupakan metode untuk menemukan masalah usability dalam desain user interface berdasarkan hasil pemeriksaan dan penilaian terhadap kepatuhan dengan prinsip usabililty yang diakui (disebut “heuristic”) oleh sekumpulan kecil pakar (Jakob Nielsen, 1992). Pada awalnya, Jakob Nielsen mengembangkan evaluasi heuristik bekerja sama dengan Rolf Molich pada tahun 1990 (Jakob Nielsen & Molich, 1990). Empat tahun kemudian, Nielsen menyempurnakan heuristik berdasarkan analisis faktor dari 249 masalah *usability* (Jakob Nielsen, 1994) untuk mendapatkan satu set heuristik dengan kekuatan penjelas maksimum, menghasilkan set heuristik yang direvisi.

Pada tahun 2020, Jakob Nielsen menambahkan lebih banyak penjelasan, contoh, dan tautan terkait (Jakob Nielsen, 2020). Sementara penyempurnaan bahasa definisi dilakukan, 10 heuristik tetap relevan dan tidak berubah sejak tahun 1994. Ketika sesuatu tetap benar selama 26 tahun, kemungkinan akan berlaku untuk

deepublish | publisher

antarmuka pengguna generasi mendatang juga. Kriteria evaluasi *usability heuristic* terdiri dari sembilan bagian, yaitu

1) Visibilitas status sistem,
2) Kecocokan antara sistem dan dunia nyata,
3) Kontrol dan kebebasan pengguna,
4) Konsistensi dan standar,
5) Pencegahan kesalahan,
6) Recognizing diutamakan daripada recalling,
7) Fleksibilitas dan efisiensi penggunaan,
8) Desain estetika dan minimalis,
9) Membantu pengguna mengenali, mendiagnosis dan memulihkan dari kesalahan,
10) Bantuan dan dokumentasi (Jakob Nielsen, 2020).

Cognitive Walktrough

Metode evaluasi *Cognitive Walktrough* merupakan salah satu teknik yang berfokus pada evaluasi awal desain sistem untuk eksplorasi kemudahan dipelajari *(learnability)*. Metode ini lebih terstruktur dibandingkan dengan metode *heuristic evaluation* dan secara tersurat dipandu oleh tugas pengguna, sehingga evaluator sebaiknya adalah seorang ahli *usability* yang menganalisis proses kognitif pengguna dalam menyelesaikan tugas-tugas yang diberikan (Jaspers, 2009).

Questionnaire

Salah satu metode evaluasi *usability* yang popular adalah mengumpulkan data dari responden melalui kuesioner. Metode ini mengandung penilaian subjektif para pengguna, sehingga perlu dikombinasikan dengan metode lain, seperti *usability testing*. Hasil umpan balik dari pengguna melalui kuesioner, menjadi dasar untuk menyeimbangkan metode lain (Roger, 2011).

Tullis and Albert (2008) menjelaskan, bahwa pengukuran *usability* dapat dilakukan dengan dua cara yaitu mengandalkan asumsi pembuat program, dan menggunakan *usability metric*. Pengukuran menggunakan *usability metric* mencakup 4 tahapan

deepublish publisher

yaitu (i) Pemilihan kuesioner, (ii) Memilih populasi atau partisipan, (iii) Menenetukan ukuran, (iv) Mengolah dan menginterpretasikan data sesuai dengan hasil penelitian. Tabel 5 menjelaskan jenis-jenis kuesioner *usability*.

Tabel 5. Jenis kuesioner usability (Perlman, 2016)

Singkatan	Kuesioner	Referensi	Institusi	Jumlah pertanyaan
QUIS	*Questionnaire for User Interface Satisfaction*	(Chin et.al, 1988)	Maryland	27
PUEU	*Perceived Usefulness and Ease of Use*	(Davis, 1989)	IBM	12
NAU	*Nielsen's Attributes of Usability*	(Nielsen, 1993)	Bellcore	5
NHE	*Nielsen's Heuristic Evaluation*	(Nielsen, 1993)	Bellcore	10
CSUQ	*Computer System Usability Questionnaire*	(Lewis, 1995)	IBM	19
ASQ	*After Scenario Questionnaire*	(Lewis, 1995)	IBM	3
PHUE	*Practical Heuristics for Usability Evaluation*	(Perlman, 1997)	OSU	13
PUTQ	*Purdue Usability Testing Questionnaire*	(Lin et al., 1997)	Purdue	100
USE	*USE Questionnaire*	(Lund, 2001)	Sapient	30
SUS	*System Usability Scale*	(Brooke, 1986)	Digital Equipment	10
UMUX	*Usability Metric for User Experience*	(Findad, 2010)	Intel	4
UMUX-Lite	*Usability Metric for User Experience-Lite*	(Lewis, 2013)	IBM	2

deepublish publisher

BAB III
UKURAN *USABILITY*

"If you can't measure it,
you can't manage it – Peter Drucker*"*

Kita telah mengenal berbagai macam ukuran (*metric*), seperti waktu, jarak, berat, tinggi, kecepatan, temperatur, volume, dan lain-lain. Setiap industri, aktivitas dan segala kebiasaan memiliki ukuran tersendiri. Istilah *horsepower* atau kekuatan kuda untuk mobil dikenal dalam industri otomotif; sedangkan dalam industri komputer, dikenal istilah Gigahertz sebagai ukuran kecepatan prosesor, atau Gigabyte untuk ukuran memori. Manusia menggunakan satuan kilogram untuk ukuran berat badan, meter untuk tinggi bangunan, bahkan ukuran ayam yang umum disajikan untuk acara tumpengan merupakan contoh ukuran dalam kehidupan sehari-hari.

Begitu pula *usability*, memiliki ukuran tersendiri. Ukuran-ukuran tersebut antara lain *task success*, *user satisfaction*, *error*, dan lain sebagainya. Data *usability* merupakan data kualitatif, namun demikian, tetap harus dapat diukur, dan dapat diubah menjadi data numerik atau sesuatu yang dapat dihitung. Seluruh ukuran *usability* (*usability metric*) harus dapat merepresentasikan beberapa aspek dari *user exprerience* yang disajikan dalam format angka. Sebagai contoh: 65% pengguna menyatakan puas terhadap suatu produk, 90% pengguna dapat menyelesaikan tugas kurang dari satu menit.

Perbedaan ukuran *usability* terhadap lainnya, yaitu karakteristiknya dalam mengukur pengalaman pengguna. Hal ini merupakan pengalaman personal dari setiap manusia terhadap

deepublish publisher

sesuatu. Ukuran *usability* menghasilkan interaksi antara pengguna dan sebuah benda, seperti aspek efektivitas (kemampuan menyelesaikan suatu tugas), efisiensi (jumlah usaha yang dikeluarkan untuk menyelesaikan tugas), dan kepuasan (derajat yang menyatakan pengguna senang dengan pengalaman mereka dalam mengerjakan suatu tugas).

Perbedaan lainnya, ukuran *usability* mengukur sesuatu tentang manusia dan perilaku mereka. Dikarenakan setiap orang memiliki perbedaan dalam kemampuan beradaptasi, maka para desainer dan developer memiliki tantangan tersendiri dalam mengukur *usability*. Ukuran *usability* membantu dalam pengambilan keputusan serta menyediakan jawaban dari setiap pertanyaan yang penting bagi organisasi yang tidak dapat dijawab dari ukuran lain. Ukuran *usability* dapat menjawab beberapa pertanyaan penting seperti:

- ✓ Apakah pengguna menyukai produk XYZ?
- ✓ Apakah produk baru ini lebih efisien daripada produk sebelumnya?
- ✓ Bagaimana *usability* sebuah produk dibandingkan dengan produk lainnya?
- ✓ Masalah *usability* apa yang paling signifikan dalam sebuah produk?
- ✓ Apakah terdapat perbaikan pada satu iterasi dengan iterasi selanjutnya?

A. Perencanaan pengukuran

Saat merencanakan studi *usability*, peneliti perlu memahami pengguna dan apa yang ingin mereka capai. Misalnya, apakah pengguna diharuskan memakai produk setiap hari sebagai bagian dari pekerjaan mereka? Apakah mereka menggunakan produk hanya sekali atau hanya beberapa kali? Apakah mereka sering menggunakan produk hanya sebagai hiburan? Sangat penting untuk memahami apa yang penting bagi pengguna. Apakah dorongan utama pengguna hanya ingin menyelesaikan tugasnya? Apakah pengguna sangat peduli pada estetika desain produk?

deepublish | publisher

Semua pertanyaan ini bermuara pada pengukuran dua aspek utama pengalaman pengguna: kinerja dan kepuasan.

Pengukuran kinerja

Pengukuran kinerja berkaitan dengan apa yang sebenarnya dilakukan pengguna dalam berinteraksi dengan produk. Hal ini termasuk mengukur sejauh mana pengguna dapat menyelesaikan tugas atau serangkaian tugas dengan sukses. Beberapa hal penting yang berkaitan dengan kinerja tugas-tugas, di antaranya waktu yang dibutuhkan untuk melakukan setiap tugas, jumlah upaya untuk melakukan masing-masing tugas (seperti jumlah klik mouse atau jumlah upaya kognitif), jumlah kesalahan yang dilakukan, dan jumlah waktu yang dibutuhkan untuk menjadi mahir dalam melakukan tugas (kemampuan belajar). Ukuran kinerja sangat penting untuk berbagai jenis produk dan aplikasi, terutama jika pengguna tidak benar-benar memiliki banyak pilihan dalam penggunaannya (seperti aplikasi internal perusahaan). Jika pengguna tidak dapat melakukan tugas-tugas utama dengan sukses ketika menggunakan suatu produk, maka kemungkinan akan produk tersebut juga gagal.

Pengukuran kepuasan

Kepuasan, yaitu semua yang dikatakan atau dipikirkan pengguna tentang interaksinya dengan produk. Pengguna dapat melaporkan bahwa sebuah produk mudah digunakan, membingungkan, atau melebihi harapannya. Pengguna mungkin memiliki pendapat tentang produk yang secara visual menarik atau tidak dapat dipercaya. Kepuasan pengguna memiliki banyak aspek berbeda. Kepuasan, dan banyak ukuran lain yang dilaporkan secara tersendiri, penting untuk produk yang memberikan pengguna untuk memiliki beberapa pilihan dalam penggunaannya. Hal ini tentu berlaku untuk sebagian besar situs web, aplikasi perangkat lunak, dan produk konsumen.

deepublish | publisher

Dashboard perencanaan

Pengukuran *usability* akan lebih *powerfull* dengan membuat sebuah *dashboard* perencanaan. Satu lembar ini cukup memudahkan para manajer dalam menjalankan pengukuran, pengembang dapat menempelkannya di sebelah papan kanban dan mempertimbangkan untuk menghadiri pengujian. Orang-orang di dalam tim desain dapat dengan mudah mengambil tindakan atas apa yang tertulis. Tabel 6 memperlihatkan sebuah contoh *dashboard* perencanaan pengukuran.

Tabel 6. Dashboard perencanaan
(diadopsi dari Usability Test Plan Dashboard oleh David Travis)
https://www.userfocus.co.uk/articles/usability_test_plan_dashboard.html)

Penulis		**Detail kontak**		**Tanggal akhir**
Tenia Wahyuningrum		tenia@ittelkom-pwt.ac.id		
Produk yang diukur Apa yang diuji? Apa tujuan bisnis dan pengalaman dari produk? *Website e-commerce* pada xyz.com [tuliskan test site]	**Tujuan pengukuran** Apa tujuan pengukuran *usability*? Adakah pertanyaan spesifik yang harus dijawab? Hipotesis apa yang akan diuji? Apakah ketentuan dan syarat pembelian barang telah dipahami oleh pengguna? Apakah pengguna memahami cara memesan barang? Apakah pengguna telah mengetahui cara melakukan pembayaran? Apakah fitur chat membantu pengguna dalam memahami detail barang yang dijual?	**Partisipan** Berapa banyak partisipan yang akan mengukur produk? Apa kriterianya? 5 partisipan yang direkrut via email. - Laki-laki atau perempuan - Usia 18-35 tahun - Memiliki smartphone	**Tugas** Tugas apa saja yang perlu dilakukan oleh pengguna? - Mencari ketentuan dan syarat pembelian dan pengembalian barang - Mencari sebuah barang dengan harga kurang dari satu juta, lebih dari lima ratus ribu. - Memasukkan barang ke keranjang belanja - Melakukan pembayaran - Melakukan chat dengan penjual	**Penanggungjawab** Siapa saja yang bertanggungjawab pada pengukuran dan apa tugasnya? Tenia Wahyuningrum (PM, Datalogger) Gita Fadila Fitriana (Moderator/fasilitator) Ariq Cahya Wardhana (Tech. support) Condro Kartiko (Rekrutmen partisipan)
Kasus bisnis Mengapa melakukan pengukuran? Apa manfaatnya? Apa yang akan terjadi jika tidak melakukan pengukuran? Pengukuran dilakukan untuk menjawab pertanyaan kunci		**Perlengkapan** Perlengkapan apa yang diperlukan? Bagaimana cara menyimpan data? Laptop/PC Kamera video Cermin Microphone Tripod Meja (2)		**Lokasi dan tanggal** Dimana dan kapan dilakukan pengujian? Kapan hasil pengujian dapat dibagikan? Laboratorium aplikasi ITTelkom Purwokerto, Jl. DI Panjaitan 128

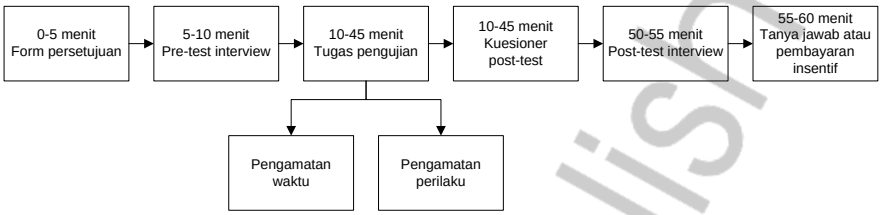
deepublish publisher

Perlu diingat bahwa *dashboard* pengukuran hanya digunakan untuk memudahkan dalam penggambaran keseluruhan pengukuran, akan tetapi, beberapa rincian rencana pengukuran mungkin terlewatkan pada *dashboard* tersebut. Beberapa hal yang dibutuhkan, yaitu, formulir rekrutmen, pernyataan persetujuan informasi/formulir persetujuan video, panduan diskusi (setidaknya dalam bentuk garis besar), formulir pilihan kata/kuesioner *post-test*, tangkapan layar untuk anotasi atau sistem pencatatan datalog.

Prosedur pengukuran

Prosedur pengukuran *usability* umumnya terdiri dari 6 tahap, seperti digambarkan pada Gambar 11. Lima menit pertama, partisipan akan mendapatkan penjelasan mengenai tujuan, tugas dan beberapa hal lain yang dibutuhkan selama pengukuran, serta menandatangani formulir persetujuan. Tahapan *pre-test interview* digunakan untuk menggali informasi personal partisipan, seperti nama, jenis kelamin, usia, durasi penggunaan gadget setiap hari, dan sebagainya. Selanjutnya, partisipan akan diminta untuk mengerjakan tugas. Para pengamat akan mencatat waktu penyelesaian tugas, dan mencatat perilaku. Apakah para partisipan merasa senang, biasa saja (netral), frustrasi, atau bahkan kesal saat menyelesaikan tugas. Setelah selesai melakukan tugas, partisipan akan diminta untuk mengisi kuesioner, beberapa jenis kuesioner *usability* seperti *System Usability Scale* (SUS), *Standardized User Experience Percentile Rank Questionnaire* (SUPR-Q), *Computer System Usability Questionnaire* (CSUQ), *Questionnaire For User Interaction Satisfaction* (QUIS), *Software Usability Measurement Inventory* (SUMI) atau yang lainnya. Lima menit berikutnya akan dilakukan *post-test interview,* biasanya para peneliti akan menggali informasi dari partisipan kesan selama menggunakan produk yang diuji. Apakah secara keseluruhan mereka merasakan kesulitan? Atau sangat mudah? Tahap akhir dari pengukuran adalah tanya jawab atau pembayaran insentif kepada partisipan.

deepublish publisher

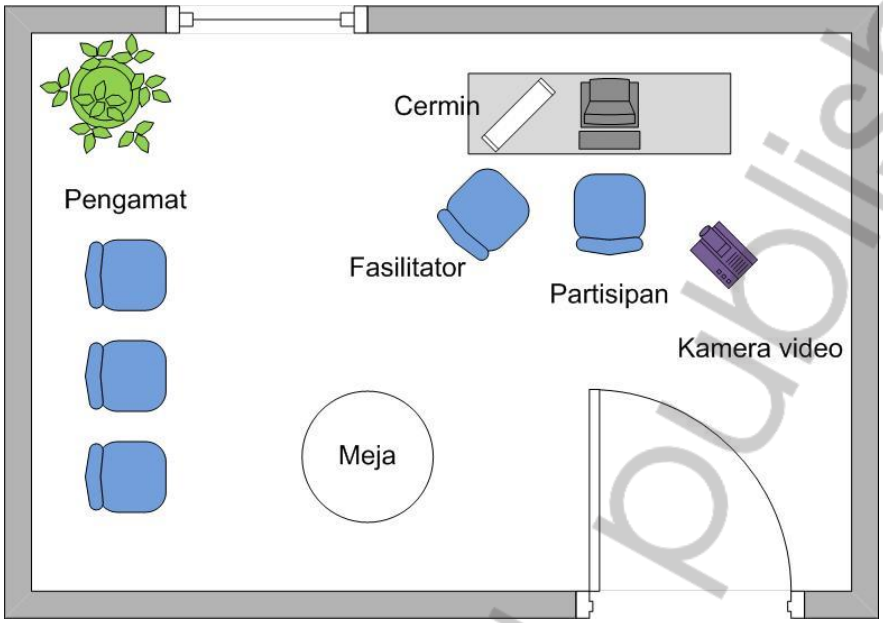


Gambar 11. Prosedur pengukuran usability

Persiapan ruangan

Sebenarnya, kita tidak memerlukan laboratorium *usability* secara khusus untuk melakukan pengukuran. Namun, jika organisasi atau perusahaan menciptakan banyak produk-produk digital yang membutuhkan pengukuran *usability* lebih dari sekali, perlu dipertimbangkan untuk menyiapkan ruangan khusus. Organisasi yang memiliki tingkat kematangan tinggi dan memiliki perhatian penuh pada pengalaman pengguna biasanya memiliki ruangan khusus untuk pengukuran *usability*. Gambar 12 merupakan contoh *layout* ruangan pengukuran *usability* tipe *single room*. Ruangan tersebut ditata sedemikian rupa, sehingga, pengamat, fasilitator dan partisipan berada dalam satu ruangan yang sama. Fasilitator duduk di sebelah partisipan untuk mendampingi partisipan jika mereka membutuhkan bantuan. Bantuan tersebut misalnya mengambilkan *mouse* yang terjatuh, membacakan instruksi (jika diperlukan), atau menyiapkan dokumen skenario tugas yang akan dikerjakan partisipan. Partisipan duduk menghadap komputer dan membelakangi kamera video. Cermin di sebelah partisipan untuk melihat ekspresi partisipan pada saat melakukan tugas. Para pengamat duduk di kejauhan sambil mencatat waktu dan mengamati ekspresi pengguna. Ekspresi tersebut dapat berupa senyuman (rasa senang), cemas, ataupun frustrasi.

deepublish publisher



Gambar 12. Contoh layout ruangan usability tipe single room

Kelebihan menggunakan tipe *single room* yaitu (Rubin & Chisnell, 2008),

1. Fasilitator memiliki pemahaman yang sangat baik tentang apa yang terjadi dengan peserta. Fasilitator tidak hanya dapat membaca isyarat verbal yang khas, tetapi juga memiliki pemahaman yang lebih baik tentang isyarat dan perilaku nonverbal peserta, seperti sedikit cemberut, mengernyitkan dahi, menggerakkan bibir atau mengangkat alis. Fasilitator dapat melihat dengan tepat apa yang dilakukan peserta hingga tingkah laku kecil dan perubahan halus dalam bahasa tubuh.
2. Selama tes eksplorasi, di mana banyak interaksi yang diinginkan untuk menginterogasi peserta selama tes, posisi ini menonjolkan rasa kerja tim. Pada tahap awal siklus hidup sebuah perangkat lunak, peserta hampir menjadi mitra yang membantu merancang produk atau dokumentasi. Faktanya,

deepublish publisher

selama tes eksplorasi, fasilitator dan peserta berada tepat di sebelah satu sama lain.

3. Untuk pengujian yang cukup rumit, di mana peserta harus berjuang dengan materi dan *task*, tipe ruangan *single room* memungkinkan fasilitator untuk mendorong dan mengatasi kesadaran diri peserta. Peserta tidak merasa hampir sendirian atau 'di kursi panas', seperti yang akan dia rasakan tanpa Anda hadir di ruang ujian. Juga lebih wajar bagi peserta untuk berpikir keras ketika moderator ujian berada di ruang ujian.
4. Pengamat dapat diundang oleh fasilitator untuk mengajukan pertanyaan lanjutan atau menuliskan pertanyaan mereka dan memberikannya kepada fasilitator untuk ditanyakan nanti.

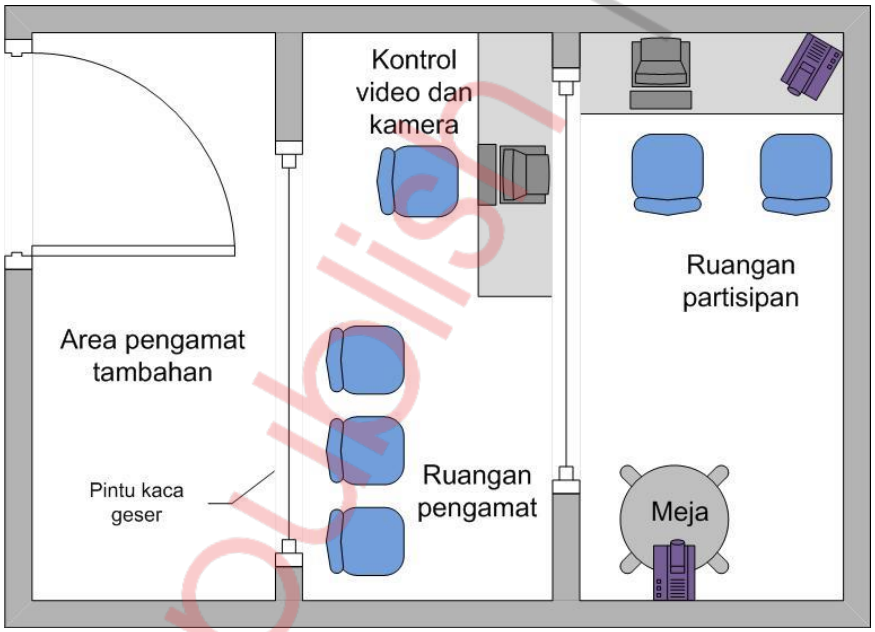
Kekurangan

1. Perilaku fasilitator dapat mempengaruhi perilaku partisipan. Fasilitator harus memperhatikan ucapan, tingkah laku, dan sikapnya, seingga tidak bereaksi terhadap apa yang dilakukan partisipan. Reaksi itu seperti menarik bahu, menghela nafas, menggelengkan kepala, dan lain sebagainya, sehingga dapat menyebabkan peserta menyimpang dari arah tujuan atau melihat ke halaman lain. Sangat penting untuk menghindari membuat catatan ketika tampaknya partisipan telah menyelesaikan tugas. Hal ini dapat memberi isyarat kepada partisipan bahwa dia telah menyelesaikan tugas dengan sukses meskipun peserta tidak yakin. Demikian pula, kegiatan mencatat di hadapan partisipan mungkin menunjukkan bahwa partisipan telah melakukan sesuatu yang salah, sehingga mempengaruhi aktivitas pengujian.
2. Ruangan yang sangat terbatas untuk pengamat, menyebabkan mereka sulit berekspresi. Perencanaan jumlah pengamat yang dapat masuk dalam satu ruangan perlu menjadi pertimbangan saat pengujian. Menghadirkan banyak pengamat dalam satu ruangan dan mengerumuni satu partisipan dapat sangat mengintimidasi. Keadaan ini menjadi

deepublish publisher

lebih lebih sulit bagi fasilitator untuk mengontrol perilaku pengamat ketika mereka begitu dekat dengan proses.

Tipe ruangan yang lain disebut *classical usability room*, tipe ini memisahkan antara partisipan dan pengamat dalam ruangan yang berbeda, dengan tujuan agar partisipan tidak merasa terintimidasi atau terganggu dengan keberadaan para pengamat. Pengamat dan partisipan dipisahkan oleh *one- way mirror*, yaitu sebuah teknologi kaca satu arah, di mana pengamat dapat melihat partisipan dari ruang berbeda, tapi partisipan tidak dapat melihat pengamat (Gambar 13). Kamera video dikontrol oleh operator yang berada satu ruangan dengan para pengamat. Secara umum, peralatan yang digunakan adalah kamera untuk merekam gerak gerik partisipan, cermin untuk mengamati ekspresi partisipan, perangkat lunak *screen recording* untuk merekam aksi layar, *tripod*, *microphone* atau *headset* untuk memperjelas suara atau komentar spontan yang dikeluarkan pengguna.



Gambar 13. Contoh layout classical usability room

deepublish publisher

Kelebihan ruangan tipe *classical* yaitu (Rubin & Chisnell, 2008),

1. Tipe ruangan *classical* memungkinkan pengumpulan data tidak terganggu oleh kegiatan para pengamat dan fasilitator. Hal ini dikarekanan fasilitator tidak ada di dalam satu ruangan, maka menghilangkan hampir semua kemungkinan efek bias karena komunikasi atau perilaku nonverbal yang tidak disengaja.
2. Berada di ruang yang kedap suara, tim pengujian dan pengamat dapat dengan mudah berkomunikasi di antara mereka sendiri tentang proses dan mendiskusikan kemungkinan peningkatan produk.
3. Lingkungan ini dapat menampung banyak pengamat (biasanya 4 hingga 12) yang melihat pengujian pada satu waktu.

Kekurangan

1. Memiliki ketergantungan pada keterampilan staf pengujian, karena pengaturan ruang kontrol dapat menciptakan lingkungan yang sangat impersonal. Hal ini kadang-kadang disebut sebagai sindrom "kelinci percobaan", dengan peserta merasa terlalu sadar diri selama tes. (Efek ini dapat diperburuk oleh jenis sistem interkom yang digunakan, beberapa di antaranya membuat instruksi fasilitator terdengar seperti "suara Tuhan.", mungkin mengagetkan para partisipan). Kesadaran diri ini lebih buruk jika memiliki peserta kurang mahir, sehingga merasa tidak percaya diri dengan kemampuannya.
2. Fasilitator mungkin perlu mengurangi efek ini dengan tetap berada di ruangan dengan peserta yang kurang mahir; atau jika peralatan memadai dan penempatan kamera yang ideal, sehingga dimungkinkan untuk dapat melihat dengan tepat apa yang sedang dilihat atau dialami peserta.

   Misalnya, terkadang fasilitator mungkin tidak dapat mengetahui di mana posisi dalam dokumen atau di layar yang dilihat peserta kecuali ada kamera khusus dan ditempatkan dengan baik (atau fasilitator dapat menggunakan peralatan

deepublish publisher

eye tracking). Untuk mengimbanginya, fasilitator mungkin perlu menggunakan protokol *thinking aloud*, yang mungkin tidak sesuai dengan desain pengujian.

3. Untuk tes *exploratory*, pengaturan ruang kontrol mungkin menawarkan sedikit keuntungan untuk memoderasi tes, karena fasilitator harus berada di sebelah peserta untuk mendampinginya. Namun, bagi pengamat, mereka dapat berdiskusi santai di ruang sebelah dengan nyaman dan ini cukup menguntungkan.
4. Pengamat dapat mulai mendesain ulang sambil berbincang dengna pengamat lainnya di ruang observasi tanpa mempertimbangkan semua data dan pengamatan.

B. Memilih Partisipan

Partisipan dalam studi *usability* memiliki dampak besar pada temuannya. Sangat penting bagi peneliti untuk merencanakan cara memilih peserta yang representatif dalam riset. Langkah-langkah dalam merekrut peserta pada dasarnya sama, yang berbeda adalah jumlah dan caranya tergantung dari keinginan peneliti ingin mengumpulkan metrik atau tidak. Tahapan pertama adalah mengidentifikasi kriteria rekrutmen yang akan digunakan untuk menentukan apakah orang tertentu memenuhi syarat untuk berpartisipasi dalam penelitian *usability*. Kriteria seharusnya sespesifik mungkin untuk mengurangi kemungkinan merekrut seseorang yang tidak sesuai dengan profil. Beberapa kriteria yang dapat digunakan untuk merekrut partisipan antara lain termasuk pengalaman mereka dengan *web*, pengguna yang telah mengalami pensiun dalam beberapa tahun, atau pengalaman dengan berbagai transaksi keuangan. Sebagai bagian dari mengidentifikasi kriteria, peneliti dapat mengelompokkan jenis peserta. Misalnya dengan cara merekrut sejumlah peserta baru serta yang memiliki pengalaman dengan produk yang ada. Setelah memutuskan jenis peserta yang diinginkan, peneliti perlu mencari tahu berapa banyak yang dibutuhkan. Tullis and Albert (2013) menyatakan bahwa

deepublish publisher

jumlah peserta yang diperlukan untuk tes *usability* adalah salah satu masalah yang paling diperdebatkan di lapangan.

Faktor yang dipertimbangkan dalam penentuan jumlah pengguna adalah keragaman populasi pengguna, kompleksitas produk, dan tujuan spesifik dari sebuah penelitian. Sebagai aturan umum, bagaimanapun, pengukuran dengan sekitar enam hingga delapan peserta untuk setiap iterasi dalam studi *formative* berhasil baik. Temuan *usability* yang paling signifikan akan diamati dengan enam atau lebih peserta pertama. Jika ada temuan kelompok pengguna yang berbeda, maka digunakan untuk membantu setidaknya empat partisipan dari masing-masing kelompok.

Jumlah partisipan

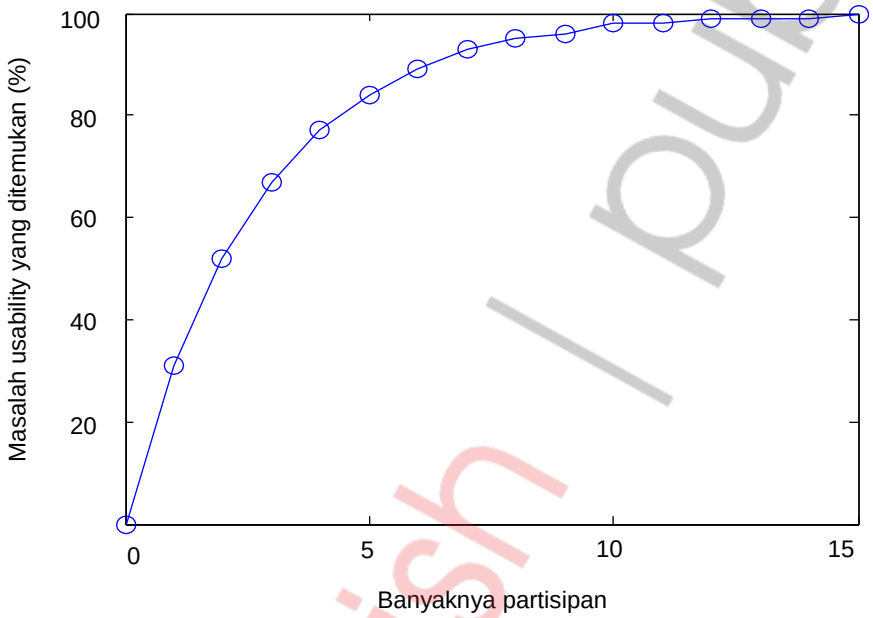
Studi *usability* jenis *summative*, pada umumnya menggunakan 50 hingga 100 pengguna representatif untuk setiap grup pengguna yang berbeda. Namun, pada keadaan krisis/genting, jumlah partisipan dapat dikurangi menjadi 30, tetapi hal ini menyebabkan varians data akan cukup tinggi, sehingga menyulitkan untuk menggeneralisasi temuan ke populasi yang lebih luas. Studi yang membutuhkan dampak perubahan desain yang berpotensi secara halus, disarankan memiliki setidaknya 100 partisipan untuk setiap kelompok pengguna yang berbeda.

Cukup lima partisipan

Beberapa pendapat yang berbeda ditemukan dalam menentukan seberapa banyak partisipan yang dibutuhkan dalam pengujian dan pengukuran *usability* untuk mengidentifikasi *usability* secara andal (Barnum *et al.*, 2003). Satu kubu percaya bahwa mayoritas, atau sekitar 80%, masalah kegunaan dapat diamati dengan lima peserta pertama (Lewis, 1994; Jakob Nielsen & Landauer, 1993; Virzi, 1992). Hal ini kemudian dikenal sebagai "angka ajaib 5" (Virzi, 1992). Salah satu cara paling penting untuk mengetahui berapa banyak peserta yang dibutuhkan dalam tes kegunaan adalah untuk mengukur *p*, atau probabilitas masalah kegunaan terdeteksi oleh satu peserta tes. Penting untuk dicatat

deepublish publisher

bahwa *p* ini berbeda dari nilai *p* yang digunakan dalam uji signifikansi. Peluangnya bervariasi dari penelitian ke penelitian, tetapi cenderung rata-rata sekitar 0,3, atau 30%. [Untuk tinjauan berbagai studi, lihat Turner, Nielsen, and Lewis (2002)]. Nielsen dan Landauer (1993) menemukan probabilitas rata-rata 31% berdasarkan 11 studi berbeda. Hal ini pada dasarnya berarti bahwa sekitar 31% dari masalah *usability* telah diamati.



Gambar 14.Partisipan dan masalah usability

Meskipun grafik pada Gambar 14 menunjukkan setidaknya ada 15 partisipan yang dilibatkan dalam evaluasi, akan tetapi, hal ini tidak dianjurkan. Penggunaan partisipan sebaiknya dalam jumlah yang sedikit, agar dapat menghemat waktu dan biaya, sehingga distribusi anggaran lebih baik digunakan untuk tes-tes kecil pada 3-5 orang saja. Berdasarkan alasan tersebut, penelitian ini menggunakan pakar sebanyak 5 orang, yaitu 3 pakar UI/UX dari lingkup akademisi maupun praktisi. Para pakar berperan dalam menentukan permasalahan apa yang sering muncul dalam

deepublish | publisher

pengukuran *usability* serta kriteria dalam pengukuran *usability* dan tingkat kepentingannya. Pakar yang lain berperan dalam menilai situs web sejumlah 2 orang sebagai penilaian tambahan selain penilaian pengguna.

Lima Partisipan Tidak Cukup

Peneliti lain telah menentang gagasan dari "angka ajaib 5" (Molich *et al.*, 2004; Spool & Schroeder, 2001; Woolrych & Cockton, 2001). Spool dan Schroeder (2001) meminta peserta untuk membeli berbagai jenis produk, seperti CD dan DVD, di tiga situs web elektronik yang berbeda. Mereka menemukan hanya 35% masalah kegunaan setelah lima peserta pertama—jauh lebih rendah dari 80% yang diprediksi oleh Nielsen (2000). Namun, dalam penelitian ini ruang lingkup situs web yang dievaluasi sangat besar, meskipun tugas untuk membeli sesuatu didefinisikan dengan sangat baik. Woolrych dan Cockton (2001) mengabaikan pernyataan bahwa lima peserta sudah cukup, terutama karena tidak memperhitungkan perbedaan individu. Analisis oleh Lindgaard *and* Chattratichart (2007) dari sembilan tes kegunaan *Comparative Usability Evaluation* (CUE-4) juga menimbulkan keraguan tentang angka ajaib 5. Mereka membandingkan hasil dari dua tim, A dan H, terungkap masing-masing 42% dan 43% dari set lengkap masalah kegunaan. Tim A hanya menggunakan 6 peserta, sedangkan Tim H menggunakan 12. Pada pandangan pertama, dapat dilihat sebagai bukti untuk angka ajaib 5, tim yang menguji hanya 6 peserta yang menemukan banyak masalah seperti tim yang menguji 12. Tapi analisis yang lebih rinci mengungkapkan kesimpulan yang berbeda. Jika dilihat secara khusus pada tumpang tindih masalah kegunaan antara hanya dua laporan ini, maka mereka hanya menemukan 28% kesamaan. Lebih dari 70% masalah terungkap hanya satu dari dua tim, mengesampingkan kemungkinan aturan lima peserta yang berlaku dalam kasus ini.

deepublish publisher

Cara merekrut partisipan

Setelah menentukan ukuran sampel, peneliti perlu merencanakan strategi perekrutan. Hal ini pada dasarnya adalah bagaimana peneliti benar-benar membuat orang berpartisipasi dalam penelitian ini. Peneliti dapat membuat daftar kemungkinan peserta dari data pelanggan dan kemudian menulis *screener* yang digunakan perekrut ketika menghubungi calon peserta.

Permintaan partisipasi dapat dilakukan melalui daftar distribusi *e-mail*. Peneliti dapat pula menyaring atau membuat segmentasi peserta melalui serangkaian pertanyaan latar belakang atau mungkin memutuskan untuk menggunakan pihak ketiga dalam menangani semua perekrutan. Beberapa dari perusahaan ini memiliki panel pengguna yang cukup luas untuk digunakan (contoh: *usability*science.com). Pilihan lainnya dapat dilakukan dengan cara mengunggah pengumuman di web atau mengirimkan *e-mail* ke grup tertentu dari calon peserta. Strategi yang berbeda berlaku untuk organisasi yang berbeda.

Apakah letak geografis berpengaruh?

Apakah peneliti perlu merekrut partisipan dari berbagai kota, wilayah, dan negara? Jawabannya biasanya tidak—geografi tidak menjadi masalah saat mengumpulkan data *usability*. Tom Tullis & Albert (2013) mengungkapkan sangat tidak mungkin bahwa para peserta di New York akan memiliki serangkaian masalah yang berbeda dari para peserta di Chicago, London, atau bahkan Washington. Meskipun demikian, ada beberapa pengecualian. Jika produk yang di evaluasi memiliki keberadaan perusahaan yang besar di satu lokasi, mungkin akan ada bias tanggapan. Misalnya, jika ingin menguji toserba YOGYA, pada *website* yogyaonline.co.id di kota asalnya di Bandung, Jawa Barat, para peneliti mungkin kesulitan mendapatkan serangkaian hasil yang netral dan tidak memihak. Selain itu, lokasi dapat berdampak pada sasaran pengguna untuk beberapa produk. Misalnya, jika peneliti mengevaluasi situs web pakaian *e-commerce*, mungkin data yang dikumpulkan berbeda dari para partisipan perkotaan atau

deepublish | publisher

pedesaan, atau partisipan di berbagai negara, di mana kebutuhan dan preferensi dapat sedikit berbeda. Beberapa peneliti menyiasati pertanyaan mengenai validitas hasil dengan memilih untuk menguji produk di berbagai daerah.

deepublish | publisher

BAB IV
UKURAN KINERJA

"The next big thing is the one that makes the last big thing usable
– Blake Ross, Co-creator of Mozilla Firefox*"*

Siapa pun yang menggunakan teknologi harus berinteraksi dengan beberapa jenis antarmuka untuk mencapai tujuan mereka. Misalnya, pengguna situs web mengklik tautan yang berbeda, pengguna aplikasi pengolah kata memasukkan informasi melalui *keyboard*, atau pengguna sistem *video game* menekan tombol pada kendali jarak jauh atau melambaikan pengontrol di udara. Tidak masalah teknologi, pengguna berperilaku atau berinteraksi dengan suatu produk dengan cara tertentu. Perilaku ini membentuk landasan metrik kinerja. Setiap jenis perilaku pengguna dapat diukur dengan beberapa cara. Perilaku yang mencapai tujuan bagi pengguna sangat penting bagi pengalaman pengguna.

Para peneliti dapat mengukur apakah pengguna mengklik melalui situs web (perilaku) menemukan apa yang mereka cari (tujuan). Peneliti juga dapat mengukur berapa lama pengguna untuk masuk dan memformat halaman teks dengan benar dalam aplikasi pengolah kata atau berapa banyak tombol yang ditekan pengguna dalam mencoba memasak makan malam beku dalam *microwave*. Semua metrik kinerja dihitung berdasarkan perilaku pengguna tertentu.

Metrik kinerja tidak hanya mengandalkan perilaku pengguna, tetapi juga pada penggunaan skenario tugas. Misalnya, jika Anda ingin mengukur kesuksesan, pengguna perlu memiliki tugas atau tujuan tertentu dalam pikiran. Tugasnya mungkin untuk menemukan

deepublish publisher

harga sweter atau menyerahkan laporan pengeluaran. Tanpa tugas, metrik kinerja tidak mungkin. Anda tidak dapat mengukur kesuksesan jika pengguna hanya menjelajahi situs web tanpa tujuan atau bermain dengan perangkat lunak. Bagaimana Anda tahu kalau dia berhasil? Tetapi ini tidak berarti bahwa tugas tersebut haruslah sesuatu yang sewenang-wenang diberikan kepada pengguna. Mereka dapat menjadi apa pun yang dilakukan pengguna untuk mengunjungi situs web langsung atau sesuatu yang dihasilkan oleh peserta dalam studi kegunaan sendiri. Seringkali kita memfokuskan studi pada tugas-tugas utama atau dasar.

Metrik kinerja juga berguna dalam memperkirakan besarnya masalah kegunaan tertentu. Sering kali tidak cukup mengetahui bahwa ada masalah tertentu. Anda mungkin ingin tahu berapa banyak orang yang mungkin mengalami masalah yang sama setelah produk dirilis. Misalnya, dengan menghitung tingkat keberhasilan yang mencakup interval kepercayaan, Anda dapat memperoleh perkiraan yang masuk akal tentang seberapa besar masalah kegunaan sebenarnya. Dengan mengukur waktu penyelesaian tugas, Anda dapat menentukan persentase audiens target Anda yang dapat menyelesaikan tugas dalam jumlah waktu tertentu. Jika hanya 20% dari pengguna target yang berhasil pada tugas tertentu, seharusnya cukup jelas bahwa tugas tersebut memiliki masalah *usability*.

Manajer senior dan pemangku kepentingan utama lainnya pada suatu proyek biasanya duduk dan memperhatikan ukuran kinerja, terutama ketika mereka disajikan secara efektif. Manajer ingin mengetahui berapa banyak pengguna yang dapat menyelesaikan serangkaian tugas inti dengan sukses menggunakan suatu produk. Mereka melihat metrik kinerja ini sebagai indikator yang kuat keseluruhan kegunaan dan prediktor potensial penghematan biaya atau peningkatan pendapatan.

Pengalaman pengguna (*user experience* ~ UX) memiliki 3 karakteristik utama, yaitu 1) adanya keterlibatan seorang pengguna, 2) adanya interaksi dengan produk, sistem, atau apa pun yang

deepublish publisher

berhubungan dengan antarmuka, 3) pengalaman pengguna dapat diamati atau diukur. Ukuran UX mengungkapkan sesuatu tentang interaksi antara pengguna dan produk (dalam ISO/IEC 9126 disebut *usability metric*, ukuran ini tetap digunakan pada versi terbarunya, ISO/IEC 25022) terdiri dari beberapa aspek **efektivitas** (mampu menyelesaikan tugas), **efisiensi** (jumlah upaya yang diperlukan untuk menyelesaikan tugas), atau **kepuasan** (sejauh mana pengguna senang dengan pengalamannya saat melakukan tugas) (Mifsud, 2015).

A. Efektivitas

Completion Rate

Efektivitas dapat dihitung dengan mengukur tingkat penyelesaian tugas (*task success*). Disebut sebagai ukuran *usability* dasar, tingkat penyelesaian dihitung dengan menetapkan nilai biner '1' jika peserta tes berhasil menyelesaikan tugas dan '0' jika dia tidak. Efektivitas dengan demikian dapat direpresentasikan sebagai persentase dengan menggunakan persamaan (1).

$$\bar{E} = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N} n_{ij}}{RN} \times 100\% \qquad (1)$$

Keterangan:

$\bar{E}$ adalah *completion rate*,

N adalah total tugas,

R adalah jumlah partisipan,

n_{ij} adalah hasil dari tugas *i* oleh partisipan *j*, n_{ij} akan bernilai 1 atau 0.

Contoh 4.1:

Jika 5 orang partisipan bekerja menggunakan sebuah aplikasi Google Mail (Gmail) dengan sebuah tugas mengirim *e-mail*. Tiga orang partisipan dapat menyelesaikan tugas dengan sukses, sedangkan dua yang lain gagal dalam menyelesaikan tugas. Maka,

deepublish publisher

menggunakan persamaan (1), *completion rate* dapat dihitung sebagai berikut.

$$\bar{E} = \frac{((3 \times 1) + (2 \times 0))}{5 \times 1} \times 100\% = 60\%$$

Contoh 4.2:

Ani dan Rudi adalah dua orang pengguna yang berpartisipasi pada pengukuran *usability* pada sebuah aplikasi *e-commerce* dengan dua tugas, yaitu mencari barang seharga Rp500.000,- dan melakukan transaksi pembelian. Ani dapat menyelesaikan tugas pertama, tetapi gagal pada tugas kedua. Sedangkan Rudi gagal melakukan tugas pertama, tetapi berhasil melakukan tugas kedua, maka *completion rate* nya, yaitu,

$$\bar{E} = \frac{((1 \times 1) + (1 \times 1))}{2 \times 2} \times 100\% = 50\%$$

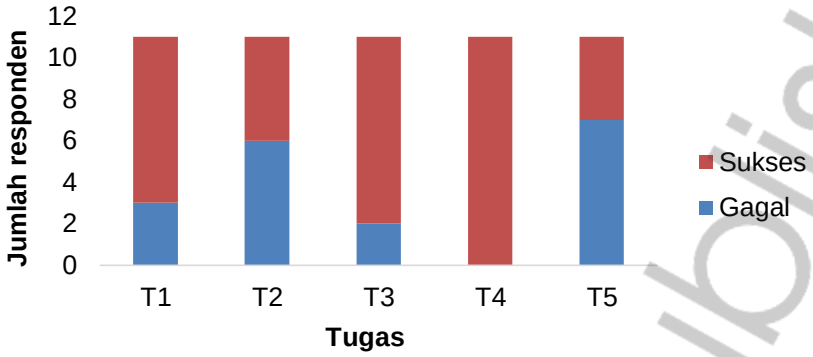
Contoh 4.3:

Partisipan mahasiswa sejumlah 11 orang bekerja pada 5 tugas pada aplikasi *e-learning*. Kelima tugas tersebut di antaranya *self enrollment*, membuka *course*, menyelesaikan *assignment,* mengerjakan *quiz,* dan melakukan tanya jawab di forum diskusi. Pada skenario tugas pertama dapat diselesaikan oleh 8 partisipan, tugas kedua dapat diselesaikan oleh 5 partisipan, tugas ketiga dapat diselesaikan oleh 9 partisipan, tugas keempat sukses dilakukan oleh 11 partisipan, dan tugas kelima dapat diselesaikan oleh 4 partisipan. Maka *completion rate* yang didapatkan, yaitu,

$$\bar{E} = \frac{(8 + 5 + 9 + 11 + 4)}{11 \times 5} \times 100\% = 67{,}2\%$$

Jika digambarkan dalam bentuk grafik, *completion rate* pada contoh 4.3 dapat dilihat pada Gambar 15.

deepublish publisher



Gambar 15. Completion Rate

Error rate

Pengukuran lain melibatkan penghitungan jumlah kesalahan yang dibuat partisipan saat mencoba menyelesaikan tugas. Kesalahan dapat berupa tindakan, slip, kesalahan atau kelalaian yang tidak diinginkan yang dilakukan pengguna saat mencoba tugas. Beberapa profesional pada pengalaman pengguna mempercayai bahwa *error* dan *usability* pada dasarnya adalah hal yang sama, tetapi sebenarnya keduanya merupakan sesuatu hal yang berbeda. Misal, jika pengguna mengalami masalah dalam menyelesaikan pembelian di situs web *e-commerce*, masalah (atau penyebabnya) mungkin adanya pelabelan produk yang membingungkan. *Error* yang dihasilkan mungkin tindakan memilih opsi yang salah untuk produk yang ingin mereka beli. Idealnya, para penguji harus memberikan deskripsi singkat, peringkat tingkat keparahan (*severity ratings*), dan mengklasifikasikan setiap kesalahan pada masing-masing kategori. Pengukuran *error* tidak selalu tepat untuk setiap situasi. Menurut Tom Tullis & Albert (2013), pengukuran *error* akan berguna pada tiga situasi umum seperti,

1. Ketika *error* akan mengakibatkan kerugian efisiensi yang signifikan. Misal, ketika *error* mengakibatkan hilangnya data, mengharuskan pengguna untuk memasukkan kembali

deepublish publisher

informasi, atau memperlambat pengguna secara signifikan dalam menyelesaikan tugas.

2. Ketika *error* akan mengakibatkan biaya yang signifikan untuk sebuah organisasi atau pengguna akhir. Misal, jika *error* akan mengakibatkan peningkatan volume panggilan untuk dukungan pelanggan atau peningkatan pengembalian produk.
3. Ketika *error* akan mengakibatkan kegagalan tugas. Misal, jika kesalahan akan terjadi menyebabkan pasien menerima obat yang salah, pengguna memilih calon karyawan yang salah secara tidak sengaja, atau pengguna web membeli produk yang salah.

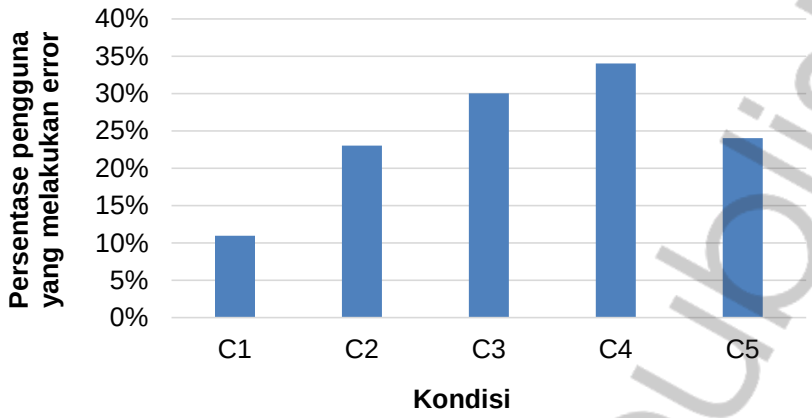
Pada umumnya *number of errors* diukur dengan memberikan tugas pada pengguna. Penguji cukup mencatat jumlah kesalahan untuk setiap tugas pada setiap pengguna. Jika diberikan satu kesempatan untuk melakukan kesalahan, angkanya menjadi 1 dan 0.

0 = Tidak ada kesalahan (*error*)
1 = Satu kesalahan (*error*)

Contoh 4.4:

Pada sebuah pengukuran *usability* sebuah aplikasi, pengguna diminta untuk melakukan tugas memasukkan *password*, dengan syarat pengguna hanya dapat melakukan satu kesalahan per tugas. Studi pada Gambar 16 menunjukkan bahwa pengguna hanya diberi satu kesempatan per tugas (memasukkan *password* dengan benar) yang dimungkinkan, dan grafik menunjukkan persentase partisipan yang melakukan kesalahan pada setiap kondisi.

deepublish | publisher



Gambar 16. Error rate

Perhitungan jumlah rata-rata kesalahan yang dibuat oleh setiap partisipan untuk setiap tugas akan memberi tahu tugas mana yang menghasilkan paling banyak kesalahan. Pengalaman pengguna yang berbeda akan menghasilkan jumlah kesalahan yang berbeda pula, misal, *novice user* akan mengalami kesalahan berkali-kali, sedangkan *expert user* barangkali tidak akan melakukan satu pun kesalahan, sehingga dengan mengambil jumlah rata-rata kesalahan untuk setiap pengguna, bias ini dapat dikurangi.

Jumlah partisipan yang optimum untuk uji keefektifan produk adalah 11-15 (Tullis & Albert, 2013). Jumlah partisipan ini cukup untuk mengungkapkan 90-95% dari semua kesalahan pengguna utama dalam produk, kesalahan statistik dari hasil penghitungan sekitar 10%, sehingga keefektifan produk secara keseluruhan dapat ditentukan dengan tingkat kepercayaan yang cukup.

0-50%	**50-75%**	**75-90%**	**90-100%**
awful (sangat buruk)	bad (buruk)	normal	good (baik)

deepublish publisher

B. Efisiensi

Time-based efficiency

Efisiensi dapat diukur dalam waktu tugas (*task time*). *Task time*, yaitu, waktu (dalam detik dan atau menit) yang dibutuhkan peserta untuk menyelesaikan tugas dengan sukses. Waktu yang dibutuhkan untuk menyelesaikan suatu tugas kemudian dapat dihitung dengan mengurangi waktu mulai dari waktu akhir seperti yang ditunjukkan pada persamaan di bawah ini:

Waktu Tugas = Waktu Berakhir - Waktu Mulai

$$\overline{P_t} = \frac{\sum_{j=1}^{R} \sum_{i=1}^{N} \frac{n_{ij}}{t_{ij}}}{NR} \tag{2}$$

Keterangan:

$\overline{P_t}$ adalah *time-based efficiency*,

R adalah jumlah partisipan

n_{ij} adalah hasil dari tugas *i* oleh partisipan *j*, n_{ij} akan bernilai 1 jika berhasil atau 0 jika gagal menyelesaikan tugas.

t_{ij} adalah waktu yang dibutuhkan partisipan *j* pada skenario tugas *i*.

Jika ada kasus penyelesaian skenario tugas yang tidak berhasil, diukur sampai saat partisipan menyerah pada tujuan atau keluar dari sistem.

Contoh 4.5:

Misal, ada 5 siswa SMP menggunakan *e-learning* "pintar" untuk melakukan satu tugas, yaitu mencari topik tentang "tumbuhan". Tiga pengguna berhasil menyelesaikannya-masing-masing membutuhkan waktu 2, 4, 5, dan 3 detik. Partisipan kelima membutuhkan waktu 10 detik dan kemudian menyerah tanpa menyelesaikan tugas.

Mengambil persamaan (2), persoalan tersebut dapat dijabarkan sebagai berikut,

N = Jumlah tugas = 1

R = Jumlah partisipan = 5

deepublish publisher

Partisipan 1: N_{ij} = 1 dan T_{ij} = 2
Partisipan 2: N_{ij} = 1 dan T_{ij} = 4
Partisipan 3: N_{ij} = 1 dan T_{ij} = 5
Partisipan 4: N_{ij} = 1 dan T_{ij} = 3
Partisipan 5: N_{ij} = 0 dan T_{ij} = 10

$$\overline{P_t} = \frac{\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{5}+\frac{1}{3}+\frac{0}{10}\right)}{1\times 5} = 0{,}25\ goal/sec$$

Efisiensi berdasarkan waktu yang didapatkan dari perhitungan adalah 0,25 *goal/sec* atau dalam satu detik, partisipan dapat menyelesaikan seperempat tugasnya. Efisiensi waktu dapat pula diartikan seberapa cepat para pengguna bekerja dengan produk.

Overall relative efficiency

Efisiensi relatif keseluruhan menggunakan rasio waktu yang dibutuhkan oleh partisipan yang berhasil menyelesaikan tugas dalam kaitannya dengan total waktu yang dibutuhkan oleh semua partisipan. Persamaan tersebut dengan demikian dapat direpresentasikan pada formula (3), di mana $\bar{P}$ adalah *overall relative efficiency*.

$$\bar{P} = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N} n_{ij}t_{ij}}{\sum_{j=1}^{R}\sum_{i=1}^{N} t_{ij}} \times 100\% \qquad (3)$$

Contoh 4.6:

Berdasarkan contoh 4.5, *overall relative efficiency* yang didapatkan adalah,

$$\bar{P} = \frac{\big((1\times 2)+(1\times 4)+(1\times 5)+(1\times 3)+(0\times 10)\big)}{(2+4+5+3+10)} \times 100\% = 58\%$$

deepublish publisher

Efisiensi relatif keseluruhan dari tugas mencari topik tumbuhan adalah 58%.

Contoh 4.7:

Diketahui partisipan bekerja dengan sebuah produk dengan diberikan 2 tugas. Pengguna pertama berhasil menyelesaikan tugas 1 namun gagal menyelesaikan tugas 2. Waktu penyelesaian tugas 1 adalah 1 detik, waktu untuk mengerjakan tugas 2 adalah 8 detik.

Sebaliknya, partisipan kedua gagal dengan tugas 1 namun berhasil menyelesaikan tugas 2. Waktu untuk menyelesaikan tugas 1 adalah 10 detik, sedangkan waktu penyelesaian tugas 2 adalah 4 detik. Di sisi lain, seorang ahli menyelesaikan skenario 1 dalam 1 detik, skenario 2 dalam 2 detik. Hitunglah *time-based efficiency*, *overall relative efficiency* dan *expert relative efficiency* nya!

$$\bar{P_t} = \frac{\left(\frac{1}{1} + \frac{0}{8} + \frac{0}{10} + \frac{1}{4}\right)}{2 \times 2} = 0{,}31\ goal/sec$$

Time-based efficiency yang didapatkan adalah 0,31 *goal/sec*, sedangkan *overall relative efficiency* nya adalah 21,7%.

$$\bar{P} = \frac{\left((1 \times 1) + (0 \times 8) + (0 \times 10) + (1 \times 4)\right)}{(1 + 8 + 10 + 4)} \times 100\% = 21{,}7\%$$

Expert relative efficiency adalah efisiensi produk relatif tambahan, yang biasa disebut efisiensi pakar. Dalam hal ini, selain pengguna yang termasuk dalam grup target, ada juga pakar yang terlibat dalam pengukuran. Pakar adalah arsitek sistem atau pengguna yang sepenuhnya menyadari arsitektur sistem dan oleh karena itu dapat melewati skenario tugas dengan kecepatan pengguna semaksimal mungkin dan penyelesaian yang sukses. Persamaan (4) digunakan untuk menghitung efisiensi pakar.

deepublish Publisher

$$\overline{P_E} = \bar{E} \times \frac{R \sum_{i=1}^{N} t_{0i}}{\sum_{j=1}^{R} \sum_{i=1}^{N} t_{ij}} \quad (4)$$

Keterangan:

$\overline{P_E}$ adalah efisiensi pakar

$\bar{E}$ adalah efektifitas (*completion rate*)

t_{0i} adalah waktu ideal yang digunakan pakar untuk menyelesaikan tugas *i*

Dengan demikian, pada contoh 4.7 dapat dihitung efisiensi pakar nya sebagai berikut,

$$\overline{P_E} = 50\% \times \frac{2 \times (1 + 2)}{(1 + 8 + 10 + 4)} = 13\%$$

Secara fisik, efisiensi pakar relatif adalah ukuran efisiensi potensial relatif terhadap efisiensi sistem aktual sehubungan dengan keefektifan penggunanya. Seperti halnya efektivitas, kesalahan statistik dalam pengukuran dengan jumlah partisipan yang sedikit cukup tinggi. Namun, bahkan sejumlah kecil responden sudah cukup untuk menunjukkan proporsi antara upaya produktif yang dihabiskan oleh pengguna saat bekerja dengan produk dan usaha yang terbuang percuma.

deepublish publisher

BAB V
UKURAN KEPUASAN

Pada umumnya, kepuasan pengguna diukur menggunakan kuesioner. Berbeda dengan efisiensi dan efektifitas, pada bagian ini, para peneliti mengamati dan mengevaluasi tindakan pengguna, sehingga ukuran kepuasan mengarah pada pemikiran subjektif pengguna. Kepuasan dapat diukur melalui kuesioner formal yang menampilkan skala kepuasan. Setiap kuesioner berisi sejumlah pernyataan yang mencerminkan pendapat subjektif pengguna tentang interaksi dengan produk. Pendapat subjektif tersebut misal, "Saya suka tampilan antarmuka aplikasi A" atau "Saya yakin produk B terlalu rumit digunakan".

A. Skala Likert

Skala Likert dipopulerkan oleh seorang psikolog sosial Amerika bernama Rensis Likert (Likert, 1932) dalam laporannya berjudul "*Technique for the Measurement of Attitudes*". Skala Likert merupakan skala psikometrik untuk mengukur pikiran dan perasaan orang terhadap sebuah survei opini hingga tes kepribadian. Pada pengukuran *usability*, responden memberikan tanggapan pada kuesioner dengan menentukan tingkat persetujuan mereka terhadap suatu pernyataan. Responden diminta untuk memilih salah satu dari pilihan yang tersedia, misalnya,

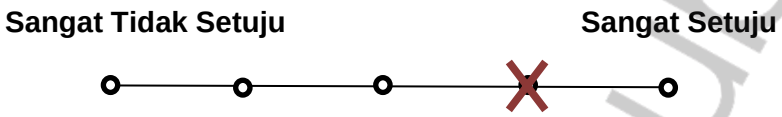
1. Sangat tidak setuju
2. Tidak setuju
3. Kurang setuju
4. Setuju
5. Sangat setuju

Skala Likert dapat bervariasi bentuknya, tetapi kebanyakan menggambarkan tingkat persetujuan dengan pernyataan-

deepublish publisher

pernyataan, dan menggunakan angka ganjil pada pilihan respons, hingga memungkinkan respons bersifat netral. Menurut kesepakatan, skala Likert yang bernilai “sangat setuju” umumnya ditampilkan di sebelah kanan saat disajikan.

Contoh:
Website memiliki tampilan yang ramah pengguna



Bentuk skala Likert tergantung dari tujuan yang diinginkan oleh peneliti. Beberapa peneliti menggunakan pilihan dengan lima skala seperti contoh, kadang digunakan juga skala dengan tujuh atau sepuluh tingkat. Penelitian yang dilakukan (Dawes, 2008) menunjukkan bahwa skala 5 atau 7 poin cenderung menghasilkan skor rata-rata yang sedikit lebih tinggi dibandingkan dengan yang dihasilkan dari skala 10 poin.

B. Analisis kepuasan pengguna

Task Level Satisfaction

Task level satisfaction adalah ukuran kepuasan pengguna, dilihat dari seberapa sulit sebuah tugas yang diberikan. Setelah pengguna melakukan tugas yang diminta, entah berhasil mengerjakannya atau tidak, mereka akan diberikan kuesioner *post-test*. Biasanya terdiri dari satu hingga lima pertanyaan, dan menggunakan skala Likert untuk memberikan wawasan pada peneliti terhadap kesulitan tugas sesuai perspektif pengguna.

Kuesioner yang biasa digunakan adalah *Single Ease Question* (SEQ) yang hanya terdiri dari 1 pertanyaan: “secara umum, bagaimanakah penilaian pengguna terhadap sistem?”. Jawaban dari pertanyaan tersebut berupa skala Likert 1-5 (semakin besar angkanya, semakin puas) (Sauro & Dumas, 2009).

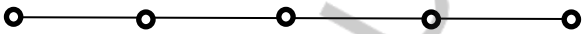
deepublish | publisher

Contoh:

Seorang UX *researcher* ingin mengetahui kepuasan pengguna pada saat melakukan pemesanan kendaraan dari aplikasi ojek *online*. Sejumlah 100 partisipan dikumpulkan, dan diminta melakukan *task* (tugas) pemesanan mobil dengan kapasitas 6 orang. Setelah melakukan *task*, para partisipan diberikan satu pertanyaan sebagai berikut,

Secara umum, tugas yang diberikan …

Sangat Sulit **Sangat Mudah**



Data yang telah terkumpul kemudian dikelompokkan berdasarkan jenis jawaban yang diperoleh. Misal, ditemukan jumlah data sebagai berikut,

- Partisipan yang menjawab sangat mudah (5) = 40 orang
- Partisipan yang menjawab mudah (4) = 25 orang
- Partisipan yang menjawab netral (3) = 5 orang
- Partisipan yang menjawab sulit (2) = 15 orang
- Partisipan yang menjawab sangat sulit (1) = 15 orang

Total skor dihitung menggunakan persamaan (5)

$$total\ skor = \sum_{i=1}^{x} w_i \times n_i \tag{5}$$

di mana w_i adalah bobot masing-masing skala, n_i adalah banyaknya partisipan yang menjawab pada skala tersebut, dan x adalah banyaknya skala.

$$total\ skor = (5 \times 40) + (4 \times 25) + (3 \times 5) + (2 \times 15) + (1 \times 15)$$
$$total\ skor = 360$$

deepublish publisher

Interpretasi skor perhitungan

Hasil interpretasi didapatkan dengan cara menghitung skor tertinggi ($skor\ max$) dan skor terendah ($skor\ min$) untuk *item* penilaian dengan rumus sebagai berikut,

$$skor\ max = \text{jumlah responden} \times \text{bobot tertinggi} \quad (6)$$
$$skor\ max = 100 \times 5$$
$$skor\ max = 500$$
$$skor\ max = \text{jumlah responden} \times \text{bobot tertinggi} \quad (7)$$
$$skor\ min = 100 \times 1$$
$$skor\ min = 100$$

Selanjutnya, perlu dihitung interval (rentang jarak) dan interpretasi persen agar mengetahui penilaian dengan mencari nilai interval (i).

$$i = {}^{100}/_{jumlah\ skor\ (Likert)} \quad (8)$$

Jadi, nilai interval yang didapatkan adalah,

$$i = {}^{100}/_{5} = 20$$

Berikut kriteria interpretasi skor berdasarkan interval yang didapatkan,

- Angka 0%–19,99% = Sangat sulit
- Angka 20%–39,99% = Sulit
- Angka 40%–59,99% = Netral
- Angka 60%–79,99% = Mudah
- Angka 80%–100% = Sangat mudah

Penyelesaian akhir

$$Persentase\ indeks = \frac{total\ skor}{skor\ maksimal} \times 100\% \quad (9)$$

deepublish | publisher

$$Persentase\ indeks = \frac{360}{500} \times 100\% = 72\%\ (Mudah)$$

Dengan demikian, dapat diambil kesimpulan bahwa pengguna merasa mudah (atau merasa puas) saat mengerjakan *task* memesan mobil menggunakan aplikasi ojek *online*.

Test Level Satisfaction

System Usability Scale Questionnaire (SUS)

Kuesioner SUS banyak digunakan dalam pengukuran kepuasan pengguna karena keunggulannya dalam pengelolaan nya yang cepat dan murah (Brooke, 2013). SUS adalah salah satu cara paling efisien untuk mengumpulkan data yang valid secara statistik dan memberi skor yang jelas dan cukup tepat. Oleh karena itu pengukuran SUS sering disebut "*Quick and Dirty test*". Kuesioner SUS telah digunakan lebih dari 25 tahun yang lalu, sebagai bagian dari teknik pengukuran *usability* yang andal (Brooke, 2013). Metode SUS telah tersedia secara bebas dan digunakan dalam evaluasi sistem oleh para kolega peneliti dan insinyur *usability* sejak 1986, dan tahun 1996 berkontribusi pada rekayasa *usability* dalam industri. SUS telah dimasukkan ke dalam perangkat evaluasi *usability* komersial seperti Morae, dan disebut sebagai "standar industri", meskipun belum pernah melalui proses standardisasi formal.

Kuesioner SUS memiliki keunggulan sebagai alat evaluasi yang mapan untuk pengukuran kualitas perangkat lunak (Alves *et al.*, 2016). Kriteria *ease of use* dikumpulkan menggunakan kuesioner SUS. Meskipun demikian, setiap pertanyaan disusun dengan hati-hati dengan cara mengumpulkan 50 pertanyaan kuesioner yang potensial, kemudian memilih dua contoh sistem perangkat lunak untuk diuji. Dua puluh orang dari kelompok karyawan, dengan pekerjaan mulai dari sekretaris hingga pemrogram sistem kemudian menilai kekuatan pertanyaan dengan 50 pernyataan kuesioner potensial untuk kedua sistem pada skala 5 poin mulai dari "sangat setuju" hingga "sangat tidak setuju". Brooke

deepublish publisher

(2013) memilih 10 pernyataan dengan nilai korelasi r = ± 0,7 hingga ± 0,9.

Sauro (2011) kemudian melakukan uji reliabilitas dan validitas terhadap *item* pertanyaan SUS menggunakan data yang lebih besar, yaitu 2324 kuesioner. Analisis faktor dari dua set data SUS independen mengungkapkan bahwa SUS sebenarnya memiliki dua faktor-*usability* (8 *item*) dan *learnability* (2 *item*). Pertimbangan baru ini memiliki nilai reliabilitas yang masuk akal (koefisien Alpha Cronbach masing-masing 0,91 dan 0,70) yang berkorelasi tinggi dengan SUS keseluruhan (masing-masing 0,985 dan 0,784); hasil lainnya, yaitu berkorelasi secara signifikan satu sama lain (r = 0,664), tetapi pada tingkat yang cukup rendah untuk digunakan sebagai skala terpisah.

Validitas SUS telah diuji dengan korelasi signifikan 0,806. Dengan demikian, keandalan dan validitas SUS dinilai sudah mapan. Meskipun terjemahan tidak resmi kuesioner SUS ke dalam bahasa Spanyol, Perancis dan Belanda telah memiliki keandalan yang serupa dengan versi bahasa Inggris asli, tetapi tingkat keandalan (reliabilitas) dan uji validitas tetap perlu diuji (Brooke, 2013). Kuesioner SUS dalam Bahasa Inggris terdiri dari 10 *item* pertanyaan sebagai berikut,

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.

deepublish publisher

Kuesioner SUS dalam Bahasa Indonesia kemudian di konversi oleh Sharfina & Santoso (2017) dan telah diukur reliabilitasnya. Hasil pengukuran menunjukkan nilai koefisien Alpha Cronbach sebesar 0,841. Nilai tersebut menunjukkan bahwa kuesioner SUS dalam terjemahan Bahasa Indonesia cukup andal.

1. Saya berpikir akan menggunakan sistem ini lagi.
2. Saya merasa sistem ini rumit untuk digunakan.
3. Saya merasa sistem ini mudah untuk digunakan.
4. Saya membutuhkan bantuan dari orang lain atau teknisi dalam menggunakan sistem ini.
5. Saya merasa fitur-fitur sistem ini berjalan dengan semestinya.
6. Saya merasa ada banyak hal yang tidak konsisten (tidak serasi) pada sistem ini.
7. Saya merasa orang lain akan memahami cara menggunakan sistem ini dengan cepat.
8. Saya merasa sistem ini membingungkan.
9. Saya merasa tidak ada hambatan dalam menggunakan sistem ini.
10. Saya perlu membiasakan diri terlebih dahulu sebelum menggunakan sistem ini.

Kuesioner SUS pada umumnya menggunakan skala Likert 1-5, nilai 1 mengindikasikan “sangat tidak setuju”, kemudian meningkat sesuai dengan bertambahnya nilai angka, nilai 2 berarti “tidak setuju”, nilai 3 merepresentasikan “netral”, nilai 4 artinya “setuju” dan akhirnya nilai 5 berarti “sangat setuju”. Interval skor kuesioner SUS adalah 0–100 dengan skor rata-rata pada situs web adalah 68, skor di atas 68 berarti pengguna merasa puas. Berdasarkan studi Lewis (2018), Tabel 7 merupakan representasi dari *Sauro–Lewis curved grading scale* (CGS) untuk melihat apakah situs web, aplikasi atau sistem informasi berada pada tingkat rendah, sedang atau tinggi pada skor *usability*.

deepublish | publisher

Tabel 7. Sauro Lewis CGS

SUS Score range	Grade	Percentile range
84.1-100	A+	96–100
80.8–84.0	A	90–95
78.9–80.7	A−	85–89
77.2–78.8	B+	80–84
74.1–77.1	B	70–79
72.6–74.0	B−	65–69
71.1–72.5	C+	60–64
65.0–71.0	C	41–59
62.7–64.9	C−	35–40
51.7–62.6	D	15–34
0.0–51.6	F	0–14

Pada *item* pertanyaan genap, kontribusi skor adalah 5 dikurangi posisi skala (Lewis, 2018). Oleh karena itu, setiap kontribusi berkisar dari 0 sampai 4. SUS adalah jumlah dari semua kontribusi skor untuk 10 *item* dikalikan dengan 2.5 seperti yang ditunjukkan pada persamaan (6) di mana U_i mengacu pada *item* ke-*i*. Skor SUS berkisar antara 0 dan 100 nilai yang lebih tinggi mencerminkan kepuasan yang lebih tinggi pula dari pengguna. Persamaan (6) merupakan perhitungan SUS (Harrati *et al.*, 2016; Wahyuningrum *et al.*, 2020).

$$SUS = 2{,}5 \times \left[\sum_{n=1}^{5} (U_{2n-1} - 1) + (5 - U_{2n}) \right] \quad (6)$$

Sebagai contoh, Tabel 8 adalah jawaban kuesioner dari seorang responden pada saat menilai sebuah situs web menggunakan SUS. Pada *item* pertanyaan pertama (Q_1) responden menjawab netral, atau setara dengan nilai 3. Demikian seterusnya sampai pada pertanyaan ke sepuluh (Q_{10}). Maka skor SUS adalah ((3-1) + (5-3) + (4-1) + (5-2) + (3-1) + (5-3) + (4-1) + (5-3) + (3-1) + (5-2)) x 2,5 = 60.

deepublish publisher

Tabel 8. Contoh jawaban responden SUS

***Item* Pertanyaan**	**Jawaban**	**Nilai Q_i**	***Item* Pertanyaan**	**Jawaban**	**Nilai Q_i**
Q_1	Netral	3	Q_6	Netral	3
Q_2	Netral	3	Q_7	Setuju	4
Q_3	Setuju	4	Q_8	Netral	3
Q_4	Tidak setuju	2	Q_9	Netral	3
Q_5	Netral	3	Q_{10}	Tidak Setuju	2
Skor SUS = 60					

Usability Metric for User Experience (UMUX)

UMUX lebih luas daripada kuesioner *single ease question*, tetapi lebih pendek dari standar industri, SUS. UMUX menggunakan skala Likert 7 poin dan 4 *item* pertanyaan yang digunakan untuk penilaian subjektif dari kegunaan yang dirasakan situs web, aplikasi atau perangkat lunak lainnya. UMUX dirancang untuk memberikan hasil yang serupa dengan yang diperoleh SUS pada 10 *item usability*, dan diatur berdasarkan definisi *usability* ISO 9241-11. Penelitian (Finstad, 2010) menunjukkan bahwa kedua skala, baik UMUX maupun SUS berkorelasi dengan baik, dapat diandalkan, dan keduanya selaras pada satu faktor *usability* yang mendasarinya. Selain itu, UMUX cukup ringkas untuk berfungsi sebagai modul *usability* dalam metrik pengalaman pengguna yang lebih luas. *Item* pertanyaan pada kuesioner UMUX dapat dilihat sebagai berikut,

1. [This system's] capabilities meet my requirements.
2. Using [this system] is a frustrating experience.
3. [This system] is easy to use.
4. I have to spend too much time correcting things with [this system].

$$UMUX = \frac{1}{24} \times \left[\sum_{n=1}^{7} (U_{2n-1} - 1) + (7 - U_{2n}) \right] \times 100 \qquad (5.2)$$

Untuk penilaian, *item* ganjil diberi skor (skor–1) dan *item* genap sebagai (7–skor). Jumlah dari semua skor ini kemudian

deepublish publisher

dibagi dengan 24 dan dikalikan dengan 100. Tabel 9 jawaban kuesioner dari seorang responden pada saat menilai sebuah situs web menggunakan UMUX. Maka skor UMUX adalah

= 1/24 × (6-1) + (7-2) + (5-1) + (7-2) × 100 = 79,1.

Tabel 9. Contoh jawaban responden UMUX

***Item* Pertanyaan**	**Jawaban**	**Nilai Q_i**	***Item* Pertanyaan**	**Jawaban**	**Nilai Q_i**
Q_1	Netral	6	Q_3	Netral	5
Q_2	Netral	2	Q_4	Setuju	2
Skor SUS = 79,1					

UMUX-Lite

UMUX-Lite dikembangkan oleh (Lewis et al., 2013), merupakan versi pendek yang hanya berisi pernyataan positif. UMUX dan UMUX-Lite dapat ditempatkan pada Skala Likert 5 atau 7 poin. Namun, rekomendasi keseluruhannya adalah jika menjalankan kuesioner ini untuk pertama kalinya, lebih baik menggunakan skala 7 poin untuk mendapatkan data yang lebih terperinci. Dua buah item pertanyaan tersebut yaitu,

1. This website/ product/ tool/ software/ prototype capabilities meet my requirements.
2. This website/ product/ tool/ software/ prototype is easy to use.

Perhitungan UMUX-Lite dibandingkan dengan SUS yaitu,

$$UMUX-Lite = 0{,}65 \times \begin{bmatrix} (UMUX-Liteitem1 + \\ UMUX-LiteItem2-2) \\ \times (100/12) \end{bmatrix} + 22{,}9 \qquad (5.2)$$

Kuesioner UMUX-LITE masih dalam penelitian lebih lanjut, sehingga tidak direkomendasikan penggunaannya secara independen dari SUS. Meskipun validitas dan reliabilitasnya cukup menjanjikan sejauh ini, bagaimanapun, para peneliti

deepublish publisher

merekomendasikan bahwa penggunaan SUS yang dikombinasikan dengan item UMUX-LITE dalam pekerjaan mereka harus dibangun *database* independen untuk evaluasi masa depan dari keandalan, validitas, dan kepekaan.

deepublish | publisher

BAB VI
STUDI KASUS

Diketahui manajemen sebuah toko ritel *online* bernama "Syopi" ingin mengetahui seberapa besar tingkat *usability*/kegunaan *website*-nya. Hasil analisis *usability website* akan digunakan oleh manajer toko untuk melakukan perbaikan desain. UX *researcher* toko "Syopi" kemudian melakukan pengukuran *usability* dengan menggunakan metode *Heuristic Evaluation* dan *Questionnaire*. Kedua metode tersebut dipilih karena dianggap cukup baik dalam menilai website, sehingga praktisi dan peneliti sering merekomendasikan penggunaan metode tersebut untuk saling melengkapi (Devine *et al.*, 2014; Paz *et al.*, 2015). UX *researcher* kemudian menentukan *usability plan* untuk mengerjakan proyek ini.

A. Pemilihan partisipan

Partisipan yang terlibat dalam pengukuran *usability* ini dibedakan menjadi dua level, yaitu ***usability expert*** dan ***intermittent user***.

Kriteria usability expert

Usability expert atau seorang spesialis kegunaan, atau kadang disebut analis kegunaan, memastikan bahwa produk memenuhi persyaratan kegunaan dan keandalan perusahaan yang membuatnya. Dalam banyak kasus, orang di bidang ini juga bekerja untuk memastikan bahwa produk menyenangkan bagi konsumen. Seringkali, spesialis kegunaan mengerjakan produk teknis, seperti perangkat lunak dan perangkat keras komputer, tetapi perusahaan juga dapat memanggilnya untuk jenis proyek lain.

Tugas pekerjaan

Usability expert menganalisis hasil pengujian dan menyiapkan laporan rekomendasinya, memberikan laporan tersebut kepada

deepublish publisher

teknisi yang bertanggung jawab untuk produksi. Dia juga meneliti tren produk dan industri saat ini untuk membuat rekomendasi.

Keterampilan

Keterampilan yang dibutuhkan seorang *usability expert, yaitu* memiliki keterampilan dalam layanan pelanggan, keterampilan penelitian dan analisis. Selain itu, *usability expert* juga harus memiliki keterampilan komunikasi dan menulis yang baik. Pekerjaan *usability expert* sangat berhubungan dengan teknologi baru dan tren terkini untuk menjawab kebutuhan konsumen, sehingga membutuhkan imajinasi untuk memikirkan cara meningkatkan pengalaman pengguna.

Pendidikan

Usability expert minimal memiliki gelar sarjana dari berbagai latar belakang jurusan. Beberapa yang paling umum adalah gelar sarjana dalam ilmu komputer, teknik, psikologi, atau pemasaran. Beberapa perusahaan memberikan preferensi kepada individu yang telah memperoleh gelar master di bidang ini serta di jurusan seperti psikologi faktor manusia dan interaksi komputer manusia. Namun, beberapa profesional memulai karier ini dengan gelar di bidang-bidang seperti seni rupa dan pendidikan. Menurut Asosiasi Profesional Kegunaan, lebih dari setengah spesialis kegunaan memiliki gelar master.

Metode *Heuristic Evaluation* hanya memerlukan sekitar 3-5 orang dalam pengukuran *usability*.

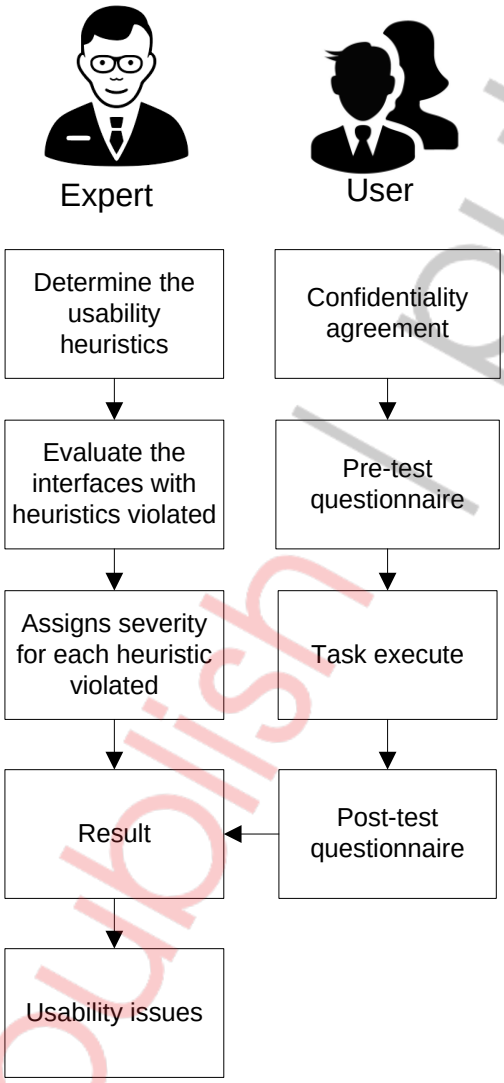
Kriteria intermittent user

Partisipan pada pengukuran *usability*, yaitu pengguna yang telah memiliki pengetahuan tentang *website* “Syopi”, berusia sekitar 17-35 tahun, laki-laki atau perempuan tidak dibedakan. Jumlah pengguna yang berpartisipasi adalah 30 orang, dan dipilih menggunakan metode *simple random sampling*.

deepublish publisher

B. Metode pengukuran *usability*

Usability diukur dengan dua karakteristik pengguna, yaitu *expert* dan *intermittent user* (Wahyuningrum *et al.*, 2020). Gambar 17 menunjukkan metode pengukuran *usability*.



Gambar 17.Metode pengukuran usability

deepublish publisher

C. Pengukuran oleh *expert*

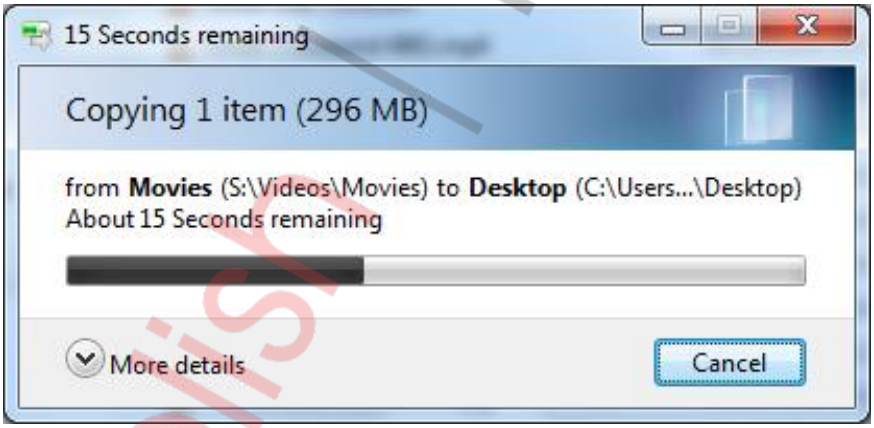
Pengukuran oleh expert melalui tahapan-tahapan sebagai berikut,

Langkah 1:

Peneliti dan ahli menentukan *usability heuristic*. *Heuristic evaluation* terdiri dari sepuluh kriteria yang telah dijelaskan oleh (Jakob Nielsen & Molich, 1990), yaitu

1. Visibilitas status sistem

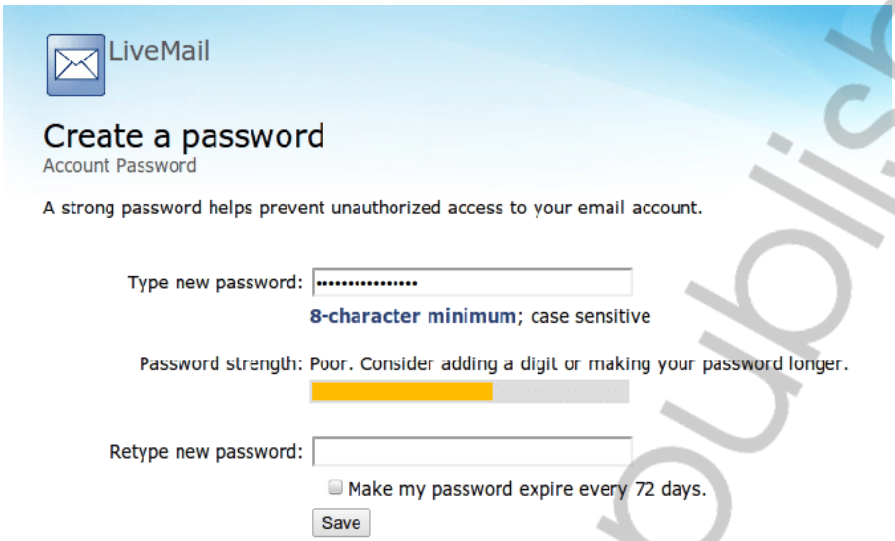
Sistem harus selalu memberikan informasi kepada pengguna tentang apa yang sedang terjadi, melalui umpan balik yang sesuai dalam waktu yang wajar. Contoh, pada saat menyalin sebuah *file* pada sistem biasanya akan ditampilkan *progress bar,* dengan animasi sebuah batang yang akan bergerak dari pangkal sampai ujung, dengan menampilkan pula estimasi waktu yang diperlukan untuk menyelesaikan tugas (Gambar 18).



Gambar 18. Progress bar saat menyalin file

Contoh yang lain ditunjukkan pada Gambar 19. Saat pengguna membuat *password*, sistem akan memberikan *feedback*, sejauh mana *password* tersebut, apakah cukup kuat, lemah atau sedang. Ilustrasi kuat lemahnya *password* dibuat dengan metode *dynamic validation* menggunakan bar berwarna merah, kuning, dan hijau untuk merepresentasikannya.

deepublish publisher



Gambar 19. Password pada live-mail

2. Kecocokan antara sistem dan dunia nyata

Sistem harus disajikan dalam bahasa yang dapat dipahami pengguna, dengan kata-kata, frasa dan konsep yang akrab bagi pengguna. Mengikuti konvensi dunia nyata, informasi yang muncul seharusnya dalam urutan yang wajar dan logis.

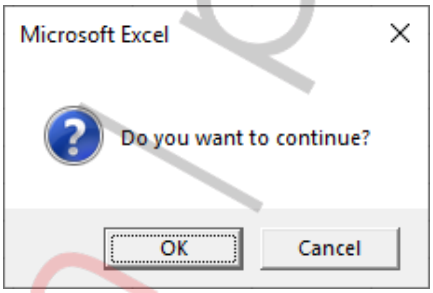
Gambar 20. Folder/map di dunia nyata dan di dalam sistem

Gambar 20 memperlihatkan bahwa bentuk *folder*/map di dunia nyata mirip dengan bentuk *icon folder* di sistem. Hal ini untuk memudahkan pengguna dalam membuat analogi berpikir. Contoh lain adalah pada penggunaan website dua bahasa, sehingga lebih memudahkan pengguna saat menjelajah.

deepublish publisher

3. Kontrol dan kebebasan pengguna

Pengguna sering memilih fungsi sistem secara tidak sengaja dan akan membutuhkan "pintu keluar darurat" yang ditandai dengan jelas untuk meninggalkan keadaan yang tidak diinginkan tanpa harus melalui dialog yang diperpanjang. Kontrol dan kebebasan pengguna mendukung adanya tombol *undo* dan *redo*, adanya tampilan *create, read, update* dan *delete* (CRUD), adanya pilihan untuk mengganti *theme*, dan kebebasan pengguna untuk melakukan *update sistem*, sekarang atau nanti. Gambar 21 memperlihatkan *dialog box* pertanyaan pada Microsoft Excel untuk memberikan kebebasan pengguna memilih apakah akan melanjutkan tugas atau tidak.

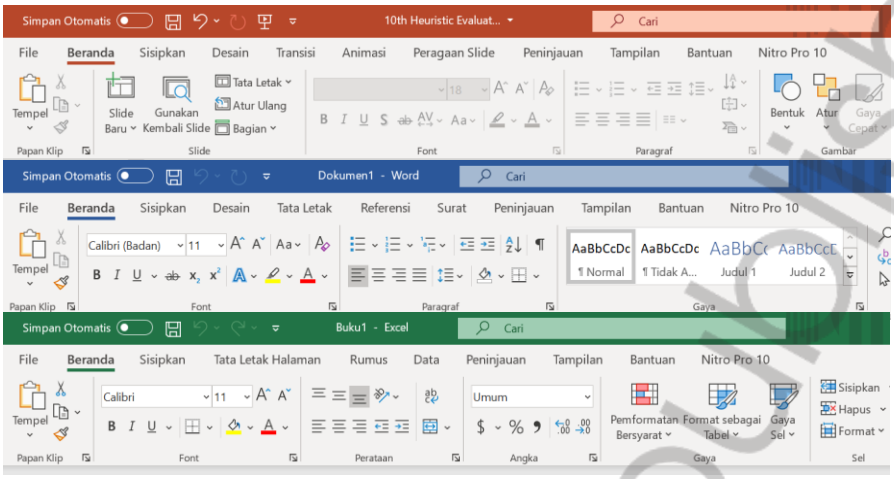


Gambar 21. Tombol validasi di Ms. Excel

4. Konsistensi dan standar

Pengguna tidak perlu bertanya-tanya apakah kata, situasi, atau tindakan yang berbeda memiliki arti yang sama. Sistem harus mengikuti konvensi *platform*. Gambar 22 menunjukkan kekonsistenan Microsoft dalam membuat perangkat lunak. Prinsip-prinsip desain dan budaya yang melekat pada pengguna diterapkan dalam merancang navigasi bar. Kekonsistenan dalam peletakan menu, mulai dari menu yang sering digunakan diletakkan pada sisi paling kiri, dan menu yang jarang digunakan diletakkan pada sisi paling kanan. Menu *file* pada ketiga *software* aplikasi tersebut memiliki kekonsistenan letak, yaitu paling kiri. Hal ini memudahkan pengguna saat akan berpindah dari satu aplikasi ke aplikasi lain.

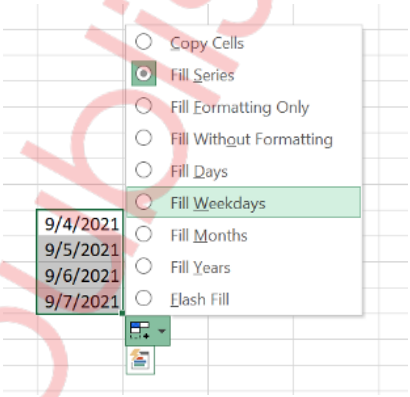
deepublish | publisher



Gambar 22. Navigasi bar pada Microsoft Excel, Word, dan PowerPoint

5. Pencegahan kesalahan

Pencegahan kesalahan lebih baik dari pesan kesalahan yang baik. Para desainer harus berhati-hati untuk mencegah masalah terjadi di tempat yang penting dan terjadi berulang pada tempat yang sama. Hilangkan kondisi rawan kesalahan dan berikan pengguna opsi konfirmasi sebelum mereka berkomitmen untuk melakukan tindakan.



Gambar 23. Autofill Microsoft Excel

deepublish | publisher

Gambar 23 menunjukkan fitur pencegahan kesalahan pada Microsoft Excel, dengan memberikan opsi *autofill*, pengguna akan dimudahkan saat mengisi sel kosong.

6. *Recognizing* diutamakan daripada *recalling*

Daripada memaksa pengguna untuk mengingat, lebih baik sistem dapat membuat pengguna dapat mengenali sistem tersebut. Pengguna dalam menjelajahi sistem atau aplikasi tentu akan melewati beberapa halaman. Agar pengguna tidak mengalami kebingungan, desainer dan developer harus membuat aplikasi yang membuat pengguna dapat mengenali pola desain yang dibuat sehingga mereka dapat tetap menggunakan aplikasi tanpa harus mengingat langkah yang harus dilakukan setelahnya. Contoh lain adalah pada fitur *autocorrect*, fitur *date time picker*, dan *dropdown menu* yang menyediakan pilihan-pilihan untuk mencegah pengguna salah input.

7. Fleksibilitas dan efisiensi penggunaan

Pada kriteria ini, sistem memberikan kemudahan pengguna ahli untuk mempercepat interaksi sehingga sistem dapat melayani pengguna yang tidak berpengalaman dan berpengalaman. Sistem mengizinkan pengguna untuk menyesuaikan tindakan yang sering dilakukan.

Contoh, pada Gambar 24, *website* memberikan kemudahan pengguna untuk melakukan navigasi menggunakan anak panah ← dan → pada *keyboard*. Contoh lain adalah tersedianya kombinasi tombol untuk *shortcut*.

deepublish | publisher

Gambar 24. Keyboard untuk menavigasi halaman

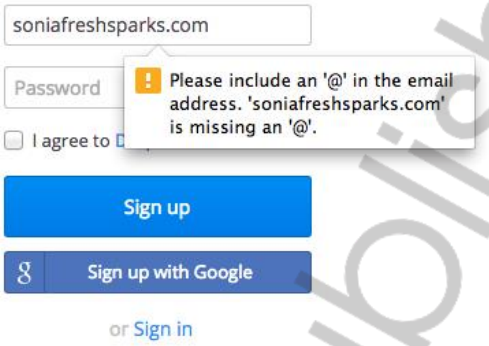
8. Desain estetika dan minimalis

Dialog tidak boleh berisi informasi yang tidak relevan atau jarang dibutuhkan. Setiap unit informasi tambahan dalam dialog bersaing dengan unit informasi yang relevan dan mengurangi visibilitas relatifnya. Contoh, Google menggunakan desain minimalis untuk menyajikan konten dan fitur dengan cara sederhana sehingga pengguna dapat langsung pada tujuannya, yaitu fokus pada pencarian.

9. Membantu pengguna mengenali, mendiagnosis dan memulihkan dari kesalahan

Desain yang baik dan nyaman tentu belum lengkap tanpa adanya penanganan *error* bila terjadi. Saat *error* terjadi, aplikasi seharusnya tidak hanya memberikan pesan *error* namun juga memberikan solusi. Gambar 25 menunjukkan pesan kesalahan pada sistem saat pengisian *form*. Pesan kesalahan tersebut juga menunjukkan bagaimana pengguna seharusnya mengisikan *text box e-mail*, yaitu harus menggunakan karakter “@”.

deepublish publisher



Gambar 25. Pesan kesalahan pada pengisian form

10. Bantuan dan dokumentasi

Meskipun sistem dapat digunakan tanpa dokumentasi, tetapi sistem perlu untuk menyediakan bantuan dan dokumentasi. Informasi semacam itu harus mudah dicari, fokus pada tugas pengguna, membuat daftar langkah-langkah konkret yang harus dilakukan, dan tidak terlalu besar ukurannya. Harapan pengguna dalam menggunakan sistem atau aplikasi tentu agar dapat menyelesaikan masalah dan pekerjaannya. Untuk membantu mereka dalam menyelesaikan masalah atau pekerjaannya perlu diberikan fitur bantuan dan dokumentasi dari kemungkinan kesalahan dalam penggunaan. Contohnya, setiap website atau aplikasi menyediakan fitur *help*, FAQ, ataupun *customer care*.

Langkah 2:

Para ahli mengevaluasi keselarasan antarmuka dengan heuristik Nielsen. Setiap pakar bekerja secara individu untuk mengevaluasi antarmuka situs *website* “Syopi” dengan memperhatikan kriteria heuristik yang dilanggar. Evaluasi didasarkan pada lima tugas yang mewakili transaksi *online* (Pokki, 2016), yaitu

1. Menemukan informasi tentang kebijakan perusahaan, termasuk praktik pengiriman dan pengembalian;
2. Menjelajahi dan menemukan produk dan detail dengan mudah;

deepublish / publisher

3. Menemukan, melihat, dan memodifikasi gerobak, menambahkan produk ke sana, dan terus membayar;
4. Mendaftar dan melihat informasi pribadi dan pesanan;
5. Menavigasi dari halaman ke halaman lain dengan sukses.

Langkah 3:

Masing-masing dari sepuluh kriteria evaluasi berisi tiga aturan yang harus diperiksa. Jika ada aturan yang dilanggar maka akan diberi nilai biner 0, jika tidak maka diberi nilai 1. Studi kasus ini menetapkan tujuh kriteria *severity*, seperti yang diilustrasikan pada Tabel 10.

Tabel 10. Tabel kriteria severity

Nilai	Aturan *usability* masing-masing kriteria
86-100%	Kepatuhan dengan 3 aturan *usability*
72-85%	Kepatuhan antara 2-3 aturan *usability*
58-71%	Kepatuhan dengan 2 aturan *usability*
44-57%	Kepatuhan antara 1 hingga 2 aturan *usability*
30-43%	Kepatuhan dengan 1 aturan *usability*
14-27%	Kepatuhan antara 0 hingga 1 aturan *usability*
0-13%	Kepatuhan dengan 0 aturan *usability*

Pada langkah terakhir, setiap evaluator ahli harus menghitung level kritis setiap masalah kegunaan dan membuat rata-rata dari semua evaluator lainnya. Kemudian hasil ini digunakan untuk memberikan informasi tambahan dari temuan SUS. Berikut ini kuesioner yang disusun dan digunakan para expert untuk menilai *website* “Syopi”. Setiap kriteria *heuristic*, terdapat tiga *item* yang harus dinilai oleh masing-masing *expert*, selanjutnya di rata-rata dijelaskan pada Tabel 11. (penjelasan lebih detail pada kriteria tiap *item* dapat dilihat pada Lampiran).

deepublish publisher

Tabel 11. Hasil rata-rata penilaian ketiga expert

No	Kriteria Evaluasi	Tugas 1	Tugas 2	Tugas 3	Tugas 4	Tugas 5	Re rata
1	Visibilitas status sistem	2,5	2,5	2,5	2,5	2	2,4
2	Kecocokan antara sistem dan dunia nyata	2,5	3	3	3	3	2,9
3	Kontrol dan kebebasan pengguna	3	3	3	3	3	3
4	Konsistensi dan standar	1,5	2,5	2,5	2,5	2,5	2,3
5	Pencegahan kesalahan	3	3	3	2	3	2,8
6	*Recognizing* diutamakan daripada *recalling*	1,5	2,5	3	3	3	2,6
7	Fleksibilitas dan efisiensi penggunaan	1,5	1	1,5	1,5	1,5	1,4
8	Desain estetika dan minimalis	3	3	2,5	3	3	2,9
9	Membantu pengguna mengenali, mendiagnosis dan memulihkan dari kesalahan	3	3	3	3	3	3
10	Bantuan dan dokumentasi	3	3	3	3	3	3
			Rata-rata				2,63

Berdasarkan Tabel 11, nilai terendah ada pada aspek ke-7, dan nilai tertinggi, yaitu pada aspek ke-3, 9, dan 10. Secara rata-rata para pakar memberikan nilai 2,63 atau setara dengan 72-85%. Artinya, website “Syopi” telah memenuhi kepatuhan antara 2-3 aturan *usability*, atau dapat dikatakan sudah cukup baik.

D. Pengukuran oleh *intermittent user*

Langkah 1:

Pada menit ke 0-5 pengguna sebagai partisipan dalam penelitian wajib mengisi *confidentiality agreement*, di mana peneliti menjamin kerahasiaan identitas dan persetujuan partisipan untuk menjadi responden penelitian.

Langkah 2:

Selanjutnya partisipan mengisi *pre-test questionnaire* kurang lebih 5 menit. Kuesioner tersebut di antaranya berisi identitas

deepublish publisher

personal partisipan biasanya memuat nama, umur, jenis kelamin, ataupun hal lain yang perlu diketahui peneliti, misal, berapa lama menggunakan aplikasi, seberapa sering mengakses aplikasi, dll.

Langkah 3:

Partisipan mengerjakan seluruh *task*, peneliti mengamati perilaku dan mencatat waktunya. Waktu pengerjaan *task* ditentukan oleh peneliti, misal, 10-40 menit. Pada kasus *e-commerce*, Pokki (2016) memberikan panduan 5 *task* yang perlu dilakukan pengguna, sama seperti *task* yang dilakukan pakar.

Langkah 4:

Setelah selesai mengerjakan *task*, partisipan mengisi *post-test questionnaire*. Kuesioner umumnya menggunakan instrumen yang telah teruji, salah satunya SUS. Sebagai contoh, Tabel 12 merupakan hasil pengukuran hasil kuesioner pada 30 responden.

Tabel 12. Hasil tabulasi kuesioner SUS

No	Usia	Pendidikan	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total skor
1	18	SMA	4	2	5	4	3	2	4	2	3	4	33
2	19	SMA	3	2	4	2	3	2	2	3	3	2	26
3	18	SMA	4	3	5	3	5	3	5	2	5	3	38
4	20	SMA	4	3	5	4	4	3	4	3	4	4	38
5	21	SMA	2	2	3	1	3	2	3	2	2	1	21
6	19	SMA	3	3	4	2	4	3	4	2	4	2	31
7	25	Sarjana	3	2	4	1	4	2	4	1	4	1	26
8	20	SMA	4	2	4	2	3	2	4	2	4	3	30
9	21	SMA	4	3	4	4	4	4	4	4	4	4	39
10	24	Sarjana	2	3	4	2	3	2	3	2	4	2	27
…													
30	28	Pasca Sarjana	5	1	5	1	5	1	5	1	5	1	30

Sebelum melakukan analisis lebih lanjut, peneliti perlu mengukur tingkat validitas dan reliabilitas instrumen.

deepublish publisher

Pengukuran validitas

Penelitian yang dilakukan menggunakan metode kuesioner perlu diuji validitasnya. Pengujian validitas berguna untuk mengukur kesahihan kuesioner yang digunakan dalam memperoleh data dari responden atau sampel penelitian. Salah satu uji validitas, yaitu dengan menggunakan korelasi *product moment Pearson*. Prinsipnya dengan menghubungkan masing-masing skor *item* kuesioner dengan total skor yang diperoleh dari jawaban responden.

Dasar pengambilan keputusan dalam uji validitas *product moment Pearson* dilihat dari nilai *r* hitung dan nilai *r* tabel atau dari nilai signifikansi dengan probabilitas tertentu.

Pada studi kasus di atas, maka perhitungan dapat dijelaskan menggunakan persamaan (7).

$$r_{xy} = \frac{N\sum XY(\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}} \quad (7)$$

Keterangan:

r_{xy} = Koefisien validitas

N = Banyaknya subjek

X = Nilai pembanding

Y = Nilai dari instrumen yang akan dicari validitasnya

Dari hasil perhitungan, didapatkan nilai r_{xy} untuk masing-masing *item* pertanyaan disajikan pada Tabel 13.

Tabel 13. Hasil tabulasi kuesioner SUS

***Item* pertanyaan**	r_{xy}	r_{tabel}	**Keterangan**
Q1	0,645	0,349	Valid
Q2	0,373	0,349	Valid
Q3	0,706	0,349	Valid
Q4	0,585	0,349	Valid
Q5	0,722	0,349	Valid
Q6	0,588	0,349	Valid
Q7	0,496	0,349	Valid

deepublish publisher

Item pertanyaan	r_{xy}	r_{tabel}	Keterangan
Q8	0,370	0,349	Valid
Q9	0,371	0,349	Valid
Q10	0.536	0,349	Valid

Pengambilan keputusan untuk menentukan apakah *item* pertanyaan valid atau tidak, dapat dilihat dari nilai *r* hitung dan nilai *r* tabel. Berdasarkan Tabel 12, nilai r hitung pada *item* soal nomor 1 adalah 0,645113. Selanjutnya adalah mencari nilai *r* tabel dengan N=30 pada signifikansi 5%, ditemukan nilai *r* tabel sebesar 0,349. Selanjutnya nilai *r* hitung dibandingkan dengan nilai r tabel.

1. Jika nilai *r* hitung > *r* tabel, maka *item* soal kuesioner tersebut valid
2. Jika nilai *r* hitung < *t* tabel, maka *item* soal kuesioner tersebut tidak valid.

Mengacu pada ketentuan dan Tabel 12, maka *item* soal nomor 1 sampai 10 dapat dikatakan valid. Jika setelah dilakukan pengujian validitas terdapat *item* soal yang tidak valid, maka beberapa solusi yang dapat dilakukan di antaranya, 1). Mengulang dan mengganti *item* kuesioner dengan kalimat yang lebih dimengerti, 2). Mengulang pengambilan data tanpa mengganti *item* kuesioner, 3). Menambah data responden, 4). Tidak mengubah *item* kuesioner dan tidak membagikan ulang pada responden, tetapi *item* angket yang tidak valid dihilangkan (dengan catatan *item* yang valid masih dapat menggambarkan dan mengukur variabel yang diteliti).

Pengukuran reliabilitas

Pengujian reliabilitas (keandalan) digunakan untuk melihat apakah kuesioner benar-benar dapat dipercaya sebagai alat ukur. Uji reliabilitas berfungsi untuk mengetahui tingkat konsistensi suatu kuesioner yang digunakan, sehingga kuesioner tersebut dapat diandalkan untuk mengukur penelitian meskipun telah dilakukan berulang-ulang. Salah satu uji reliabilitas, yaitu menggunakan Alpha Cronbach.

deepublish publisher

Dasar pengambilan keputusan untuk uji reliabilitas, yaitu

1. Jika nilai Alpha Cronbach > 0,60 maka kuesioner dinyatakan reliabel atau konsisten.
2. Jika nilai Alpha Cronbach < 0,60 maka kuesioner dinyatakan tidak reliabel atau tidak konsisten.

$$r_{ac} = \left(\frac{k}{k-1}\right)\left[1 - \frac{\sum\sigma_b^2}{\sigma_t^2}\right] \quad (8)$$

Keterangan:

r_{ac} : koefisien reliabilitas Alpha Cronbach,

k : banyak butir/*item* pertanyaan,

$\sum\sigma_b^2$: jumlah/total varians per-butir/*item* pertanyaan,

σ_t^2 : jumlah atau total varians.

Berdasarkan hasil perhitungan didapatkan nilai Alpha Cronbach, yaitu 0,734, sehingga dapat dikatakan butir-butir pertanyaan dalam kuesioner telah konsisten/reliabel.

Skor SUS

Skor SUS dihitung berdasarkan persamaan (6), kemudian dihitung pula rata-rata untuk keseluruhan responden. Hasil perhitungan menunjukkan rata-rata skor SUS pada studi kasus website “Syopi” adalah 78,4 atau setara dengan grade A. Hal ini dapat dikatakan bahwa responden puas dengan website “Syopi” berdasarkan aspek pengalaman pengguna.

Menentukan usability issues:

Hasil pengamatan dan perhitungan yang diperoleh dari perspektif pengguna dan para ahli kemudian dijadikan dasar untuk memberikan rekomendasi. Rekomendasi tersebut dilihat dari permasalahan yang timbul pada saat pengujian. Beberapa *usability issues* yang muncul, yaitu,

1. **Masalah:**

Pengguna tidak menemukan *shortcut* untuk mengakhiri kegiatan dalam menjelajahi dan menemukan produk dan detailnya.

deepublish publisher

Pengguna harus terus menerus menekan tombol "*back*" sehingga cukup menyulitkan.

Rekomendasi:

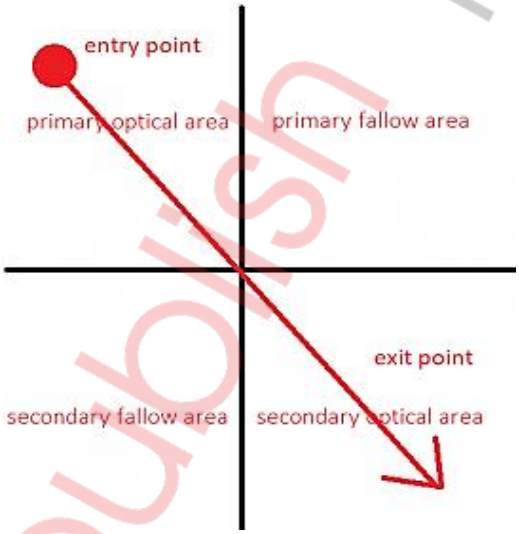
Menambahkan tombol "*exit*" sehingga memudahkan keluar sewaktu-waktu saat pengguna melakukan navigasi antar halaman.

2. **Masalah:**

Tombol dan *textbox* pencarian berada di sebelah kiri bawah, sehingga sulit dicari oleh pengguna.

Rekomendasi:

Berdasarkan prinsip Gutenberg, ruang desain dibagi menjadi 4 kuadran yang sama. Kuadran aktif adalah kiri atas dan kanan bawah. Kuadran pasif adalah kiri bawah dan kanan atas. Kuadran atas adalah primer di bandingkan kuadran bawah. Dengan demikian, area pencarian dapat diletakkan pada kanan atas atau kiri atas.



deepublish | publisher

3. **Masalah**:

Tombol “bayar” pada pengisian *form* sulit dikenali.

Rekomendasi:

Berdasarkan prinsip desain, tombol dirancang sedemikian rupa sehingga pengguna tertarik untuk menekan tombol tersebut (*call to action*). Tombol dengan *caption* “*submit*” sebaiknya diganti dengan “bayar”.

deepublish | publisher

LAMPIRAN

Lampiran 1. Pertanyaan dalam Evaluasi Heuristik pada studi kasus, dijabarkan sebagai berikut,

1. Visibilitas status sistem
1.1. Apakah pengguna mengetahui posisi mereka saat ini dalam perjalanan pengguna?
1.2. Apakah pengguna diberi tahu tentang perubahan dalam perjalanan pengguna mereka?
1.3. Apakah pengguna mengetahui faktor-faktor yang dapat berdampak signifikan pada pengalaman mereka?

2. Kecocokan antara sistem dan dunia nyata
2.1. Apakah elemen UI dan permintaan interaksi mudah dikenali?
2.2. Apakah pengalaman *online* meniru keakraban tindakan dan perilaku *offline*?
2.3. Apakah situs web menggunakan akronim, istilah teknis, atau jargon yang membutuhkan penjelasan? Jika akronim digunakan, apakah akronim tersebut dijelaskan dengan jelas?

3. Kontrol dan kebebasan pengguna
3.1. Apakah pengguna dapat keluar dari semua status seperti pop-up dan multimedia? Apakah status keluar konsisten dan jelas?
3.2. Apakah pengguna dapat menggunakan bagian inti situs web tanpa mendaftar?
3.3. Apakah pengguna memiliki kendali atas informasi pribadinya?

deepublish | publisher

4. Konsistensi dan standar

4.1. Apakah ada standar desain yang konsisten untuk semua ajakan bertindak (CTA) di situs web?

4.2. Apakah ada standar desain yang konsisten untuk kontrol formulir?

4.3. Apakah ada standar desain yang konsisten untuk *heading*?

5. Pencegahan kesalahan

5.1. Apakah ada kendala membantu yang mencegah pengguna melakukan kesalahan?

5.2. Apakah panduan pengguna dengan saran untuk mencegah tindakan yang salah?

5.3. Apakah pengguna disajikan dengan pemformatan yang memaafkan untuk informasi?

6. Pengakuan daripada mengingat

6.1. Apakah pengguna disajikan dengan daftar produk atau halaman yang baru dilihat

6.2. Apakah pengguna disajikan dengan konten yang disesuaikan berdasarkan tindakan sebelumnya?

6.3. Apakah pengguna disajikan dengan *item* navigasi yang mengurangi beban kognitif dan membantu mengingat?

7. Fleksibilitas dan efisiensi penggunaan

7.1. Apakah pengguna diberikan jalan pintas untuk mengakhiri tujuan?

7.2. Apakah pengguna dapat menyesuaikan tindakan yang sering dilakukan?

7.3. Apakah pengguna disajikan dengan informasi ambien untuk tindakan cepat?

8. Desain estetika dan minimalis

8.1. Apakah desain antarmuka pengguna sederhana dan mudah dipahami?

8.2. Apakah pengguna memahami dengan jelas arti semua ikon dan mengapa mereka disertakan dalam desain?

8.3. Apakah semua formulir mudah dipahami dan mudah diisi?

deepublish publisher

9. Membantu pengguna mengenali, mendiagnosis, dan memulihkan dari kesalahan

9.1. Apakah pengguna disajikan dengan pesan kesalahan (bukan tanpa pesan) saat menambahkan informasi yang salah dalam kontrol formulir?

9.2. Apakah pengguna disajikan dengan pesan kesalahan yang dapat dibaca manusia yang menawarkan informasi berguna tentang cara memperbaiki masalah?

9.3. Apakah pengguna diberikan pesan kesalahan sopan yang tidak menyalahkan pengguna atas kesalahan tersebut?

10. Bantuan dan dokumentasi

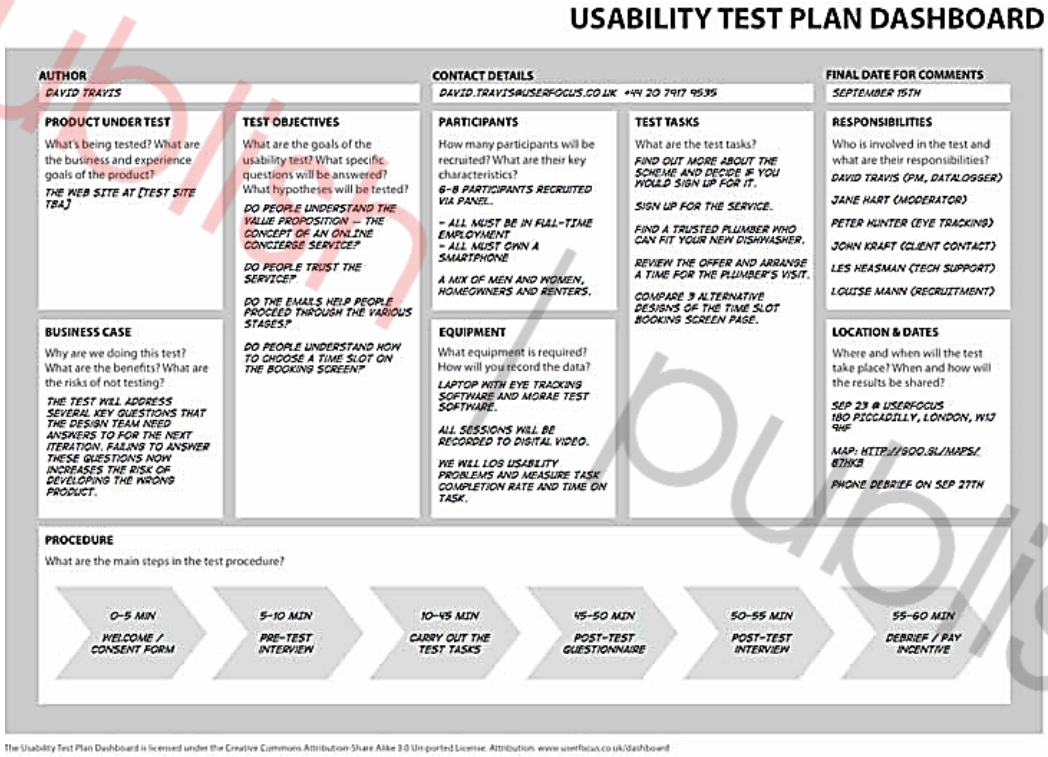
10.1. Apakah pengguna diberikan langkah/pedoman yang jelas untuk menggunakan layanan produk?

10.2. Apakah pengguna memiliki akses ke dokumentasi dengan topik yang relevan untuk membantu mencapai tujuan mereka?

10.3. Apakah pengguna disajikan dengan saluran komunikasi lain untuk meminta bantuan untuk mencapai tujuan mereka?

deepublish | publisher

Lampiran 2. Usability Test Plan Dashboard



Sumber: David Travis, https://www.userfocus.co.uk/articles/usability_ test_plan_dashboard.html

deepublish / publisher

Lampiran 3. Tabel r Product Moment pada sig. 0,05 (two tail)

N	r	N	r	N	r	N	r	N	r	N	r
1	0.997	41	0.301	81	0.216	121	0.177	161	0.154	201	0.138
2	0.95	42	0.297	82	0.215	122	0.176	162	0.153	202	0.137
3	0.878	43	0.294	83	0.213	123	0.176	163	0.153	203	0.137
4	0.811	44	0.291	84	0.212	124	0.175	164	0.152	204	0.137
5	0.754	45	0.288	85	0.211	125	0.174	165	0.152	205	0.136
6	0.707	46	0.285	86	0.21	126	0.174	166	0.151	206	0.136
7	0.666	47	0.282	87	0.208	127	0.173	167	0.151	207	0.136
8	0.632	48	0.279	88	0.207	128	0.172	168	0.151	208	0.135
9	0.602	49	0.276	89	0.206	129	0.172	169	0.15	209	0.135
10	0.576	50	0.273	90	0.205	130	0.171	170	0.15	210	0.135
11	0.553	51	0.271	91	0.204	131	0.17	171	0.149	211	0.134
12	0.532	52	0.268	92	0.203	132	0.17	172	0.149	212	0.134
13	0.514	53	0.266	93	0.202	133	0.169	173	0.148	213	0.134
14	0.497	54	0.263	94	0.201	134	0.168	174	0.148	214	0.134
15	0.482	55	0.261	95	0.2	135	0.168	175	0.148	215	0.133
16	0.468	56	0.259	96	0.199	136	0.167	176	0.147	216	0.133
17	0.456	57	0.256	97	0.198	137	0.167	177	0.147	217	0.133
18	0.444	58	0.254	98	0.197	138	0.166	178	0.146	218	0.132
19	0.433	59	0.252	99	0.196	139	0.165	179	0.146	219	0.132
20	0.423	60	0.25	100	0.195	140	0.165	180	0.146	220	0.132
21	0.413	61	0.248	101	0.194	141	0.164	181	0.145	221	0.131
22	0.404	62	0.246	102	0.193	142	0.164	182	0.145	222	0.131
23	0.396	63	0.244	103	0.192	143	0.163	183	0.144	223	0.131
24	0.388	64	0.242	104	0.191	144	0.163	184	0.144	224	0.131
25	0.381	65	0.24	105	0.19	145	0.162	185	0.144	225	0.13
26	0.374	66	0.239	106	0.189	146	0.161	186	0.143	226	0.13
27	0.367	67	0.237	107	0.188	147	0.161	187	0.143	227	0.13
28	0.361	68	0.235	108	0.187	148	0.16	188	0.142	228	0.129
29	0.355	69	0.234	109	0.187	149	0.16	189	0.142	229	0.129
30	0.349	70	0.232	110	0.186	150	0.159	190	0.142	230	0.129
31	0.344	71	0.23	111	0.185	151	0.159	191	0.141	231	0.129
32	0.339	72	0.229	112	0.184	152	0.158	192	0.141	232	0.128
33	0.334	73	0.227	113	0.183	153	0.158	193	0.141	233	0.128
34	0.329	74	0.226	114	0.182	154	0.157	194	0.14	234	0.128
35	0.325	75	0.224	115	0.182	155	0.157	195	0.14	235	0.127
36	0.32	76	0.223	116	0.181	156	0.156	196	0.139	236	0.127
37	0.316	77	0.221	117	0.18	157	0.156	197	0.139	237	0.127
38	0.312	78	0.22	118	0.179	158	0.155	198	0.139	238	0.127
39	0.308	79	0.219	119	0.179	159	0.155	199	0.138	239	0.126
40	0.304	80	0.217	120	0.178	160	0.154	200	0.138	240	0.126

deepublish publisher

DAFTAR PUSTAKA

Alves, J. M., Albino, D. B. L., Resener, M. C., Zannin, M., Savaris, A., Von Wangenheim, C. G., & Von Wangenheim, A. (2016). Quality evaluation of poison control information systems: A case study of the DATATOX system. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, *2016-Augus*, 30–35. https://doi.org/10.1109/CBMS.2016.53

Barnum, C., Bevan, Ni., Cockton, Gi., Nielsen, J., Spool, J., & Wixon, D. (2003). The " Magic Number 5 ": Is It Enough for Web Testing? *CHI 2003*, *5*, 5–7. https://doi.org/10.1145/765891.765936

Brooke, J. (2013). SUS: A Retrospective. *Journal of Usability Studies*, *8*(2), 29–40. https://doi.org/10.1074/jbc.R115.675280

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, *50*(1), 61–77. https://doi.org/10.1177/147078530805000106

Devine, E. B., Lee, C. J., Overby, C. L., Abernethy, N., McCune, J., Smith, J. W., & Tarczy-Hornoch, P. (2014). *Usability* evaluation of pharmacogenomics clinical decision support aids and clinical knowledge resources in a computerized provider order entry system: A mixed methods approach. *International Journal of Medical Informatics*, *83*(7), 473–483. https://doi.org/10.1016/j.ijmedinf.2014.04.008

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). Human-Computer Interaction. In *Human-Computer Interaction*: *Vol. Third* (Issue January). Prentice Hall. https://doi.org/10.1207/S15327051HCI16234

deepublish publisher

El-rayyes, E. K., & Abu-zaid, I. M. (2012). *New Model to Achieve Software Quality Assurance* (*SQA*) *in Web Application*. *2*(7), 423–426.

Finstad, K. (2010). The *usability* metric for user experience. *Interacting with Computers*, *22*(5), 323–327. https://doi.org/10.1016/j.intcom.2010.04.004

Harrati, N., Bouchrika, I., Tari, A., & Ladjailia, A. (2016). Exploring user satisfaction for e-learning systems via usage-based metrics and system *usability* scale analysis. *Computers in Human Behavior*, *61*, 463–471. https://doi.org/10.1016/j.chb.2016.03.051

International Organization for Standardization. (2011). Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models. In *ISO*/*IEC* (Vol. 2011, p. 34). https://doi.org/ISO/IEC FDIS 25010:2010(E)

Jacobsen, N. E. (1999). *Usability Evaluation Methods The Reliability and Usage of Cognitive Walkthrough and Usability Test*. University of Copenhagen, Denmark.

Lewis, J. R. (1994). *Sample Sizes for Usability Studies*: *Additional Considerations* (Issue 54.711). https://doi.org/10.1177/001872089403600215

Lewis, J. R. (2018). The System *Usability* Scale: Past, Present, and Future. *International Journal of Human-Computer Interaction*, *34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307

Likert, R. (1932). Technique for the Measurement of Attitudes. In *Archives of Psychology*. https://doi.org/10.4135/9781412961288.n454

Lindgaard, G., & Chattratichart, J. (2007). *Usability* testing: What have we overlooked? *Usability* Testing: What Have We Overlooked? *CHI 2007*, 1415–1424. https://doi.org/10.1145/1240624.1240839

Maxwell, D. B. S. (1993). Beer cans: A guide for the archaeologist. *Historical Archaeology*, *27*(1), 95–113. https://doi.org/10.1007/bf03373561

deepublish publisher

Mifsud, J. (2015). *Usability Metrics - A Guide To Quantify The Usability Of Any System*. *Usability*geek. http://*usability*geek.com/*usability*-metrics-a-guide-to-quantify-system-*usability*/

Molich, R., Ede, M. R., Kaasgaard, K., & Karyukin, B. (2004). Comparative *usability* evaluation. *Behaviour & Information Technology*, *23*(1), 65–74. https://doi.org/10.1080/0144929032000173951

Nielsen, Jacob. (1993). *Usability Engineering*. Academic Press.

Nielsen, Jakob, & Landauer, T. K. (1993). A Mathematical Model of the Finding of *Usability* Problems. *ACM INTERCHI Conference*, 206–213.

Nielsen, Jakob, & Molich, R. (1990). Heuristic Evaluation of User Interfaces. *Conference on Human Factors in Computing Systems*, *April*, 249–256.

Paz, F., Paz, F. A., Villanueva, D., & Pow-Sang, J. A. (2015). Heuristic Evaluation as a Complement to *Usability* Testing: A Case Study in Web Domain. *2015 12th International Conference on Information Technology - New Generations*, 546–551. https://doi.org/10.1109/ITNG.2015.92

Pokki, S. (2016). *Web usability in e-commerce Usability evaluation of four web shops*. Lahti University of Applied Sciences.

Sauro, J. (2011). *A practical guide to the System Usability Scale*: *Background, benchmarks, & best practices*.

Sauro, J., & Dumas, J. S. (2009). Comparison of three one-question, post-task *usability* questionnaires. *Conference on Human Factors in Computing Systems - Proceedings*, *August*, 1599–1608. https://doi.org/10.1145/1518701.1518946

Shah, K. A. (2012). *How To Improve Software Quality Assurance in Developing Countries*. *3*(2), 17–28.

Sharfina, Z., & Santoso, H. B. (2017). An Indonesian adaptation of the System *Usability* Scale (SUS). *2016 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2016*, 145–148. https://doi.org/10.1109/ICACSIS.2016.7872776

deepublish publisher

Sharp, H., Rogers, Y., & Preece, J. (2002). Interaction design: beyond human-computer interaction. In *Book* (Vol. 11). John Wiley & Sons, Inc. https://doi.org/10.1162/leon.2005.38.5.401

Sowunmi, Y. (2015). *An Empirical Evaluation of Software Quality Assurance Practices and Challenges in a Developing Country*. 867–871. https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.129

Spool, J., & Schroeder, W. (2001). Testing Web Sites : Five Users Is Nowhere Near Enough. *CHI 2011*. https://doi.org/10.1145/634067.634236

Stake, R. E. (1976). *Evaluating Educational Programmes*: *The Need and the Response*: *A Collection of Resource Materials*. OECD.

Tullis, T., & Albert, B. (2013). Measuring the User Experience: Collecting, Analyzing, and Presenting *Usability* Metrics: Second Edition. In *Measuring the User Experience*: *Collecting, Analyzing, and Presenting Usability Metrics*: *Second Edition*. https://doi.org/10.1016/C2011-0-00016-9

Turner, C. W., Nielsen, J., & Lewis, J. R. (2002). Current Issues in the Determination of *Usability* Test Sample Size : How Many Users is Enough ? *Usability Professionals Association*, *July 2002*.

Virzi, R. A. (1992). Refining the Test Phase of *Usability* Evaluation : How Many Subjects Is Enough ? *Human Factors*, *34*(4), 457–468.

Wahyuningrum, T., & Azhari. (2017). Tinjauan *usability* dalam evaluasi produk perangkat lunak. *Seminar Nasional Inovasi Teknologi*, 1–8.

Wahyuningrum, T., Kartiko, C., & Wardhana, A. C. (2020). *Exploring e-Commerce Usability by Heuristic Evaluation as a Compelement of System Usability Scale*. 1–5.

Woolrych, A., & Cockton, G. (2001). Why and When Five Test Users aren't Enough. *IHM-CHI*, 105–108.

deepublish publisher