

On the Value of CrowdRE in Research and Practice



Fabiano Dalpiaz
Requirements Engineering Lab
Utrecht University, Netherlands
September 21, 2021

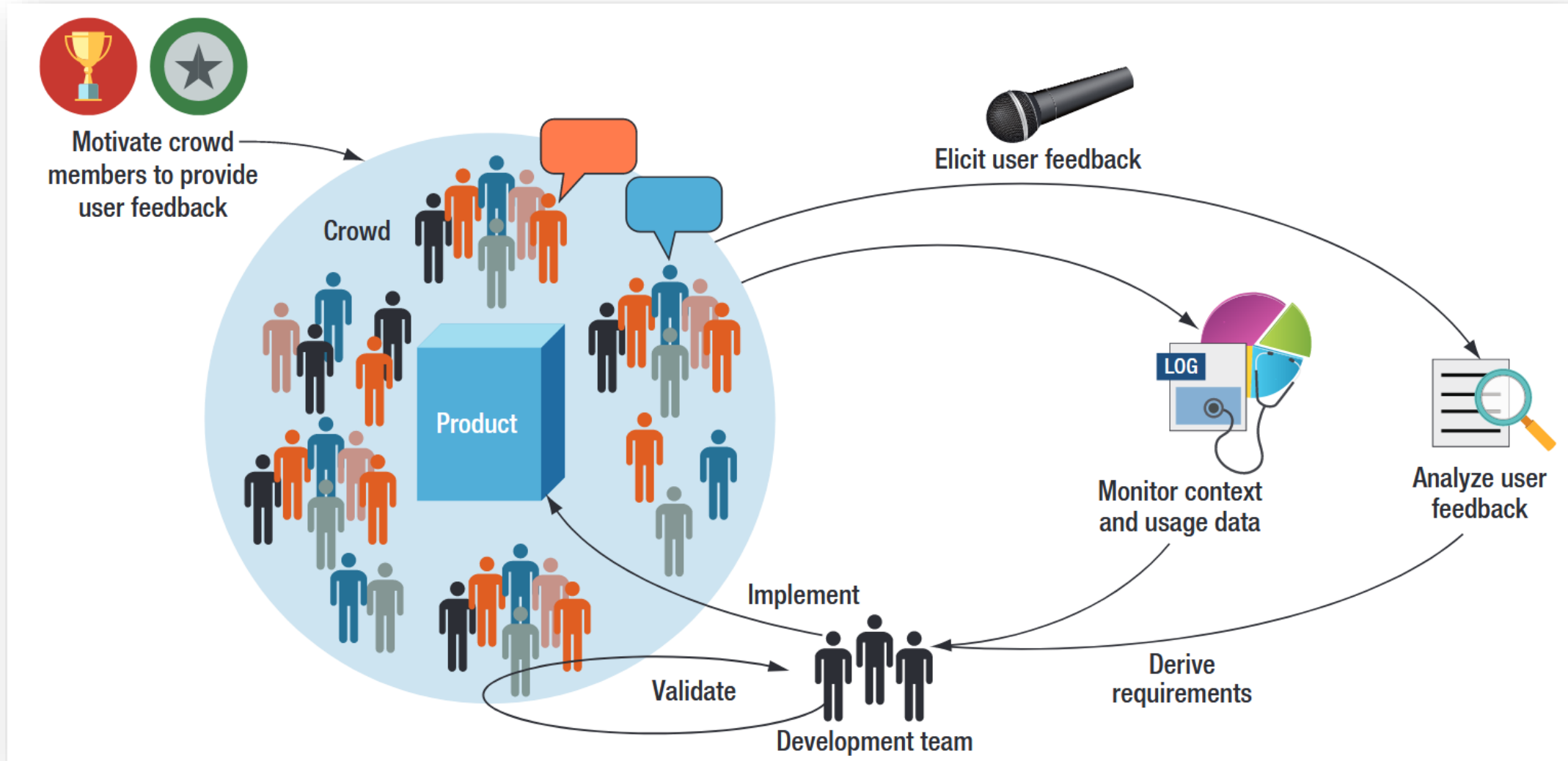


**Utrecht
University**

I. The status-quo in CrowdRE research

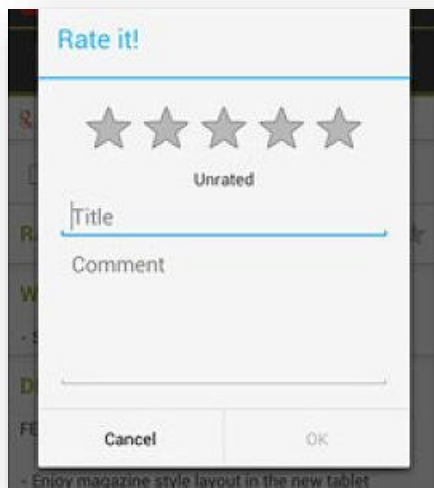


CrowdRE, illustrated

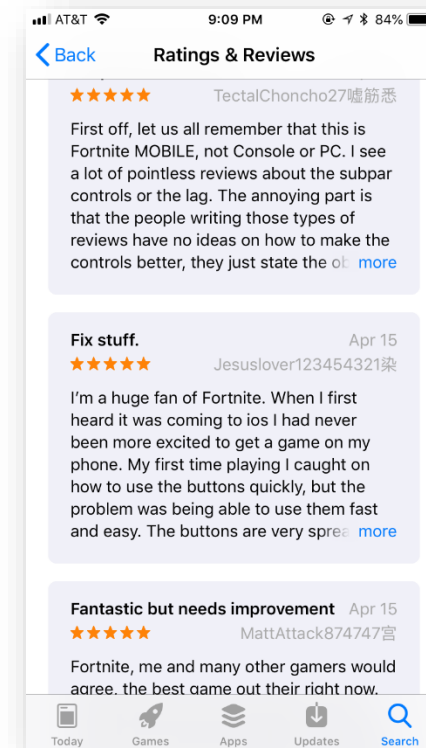


Pull and Push feedback

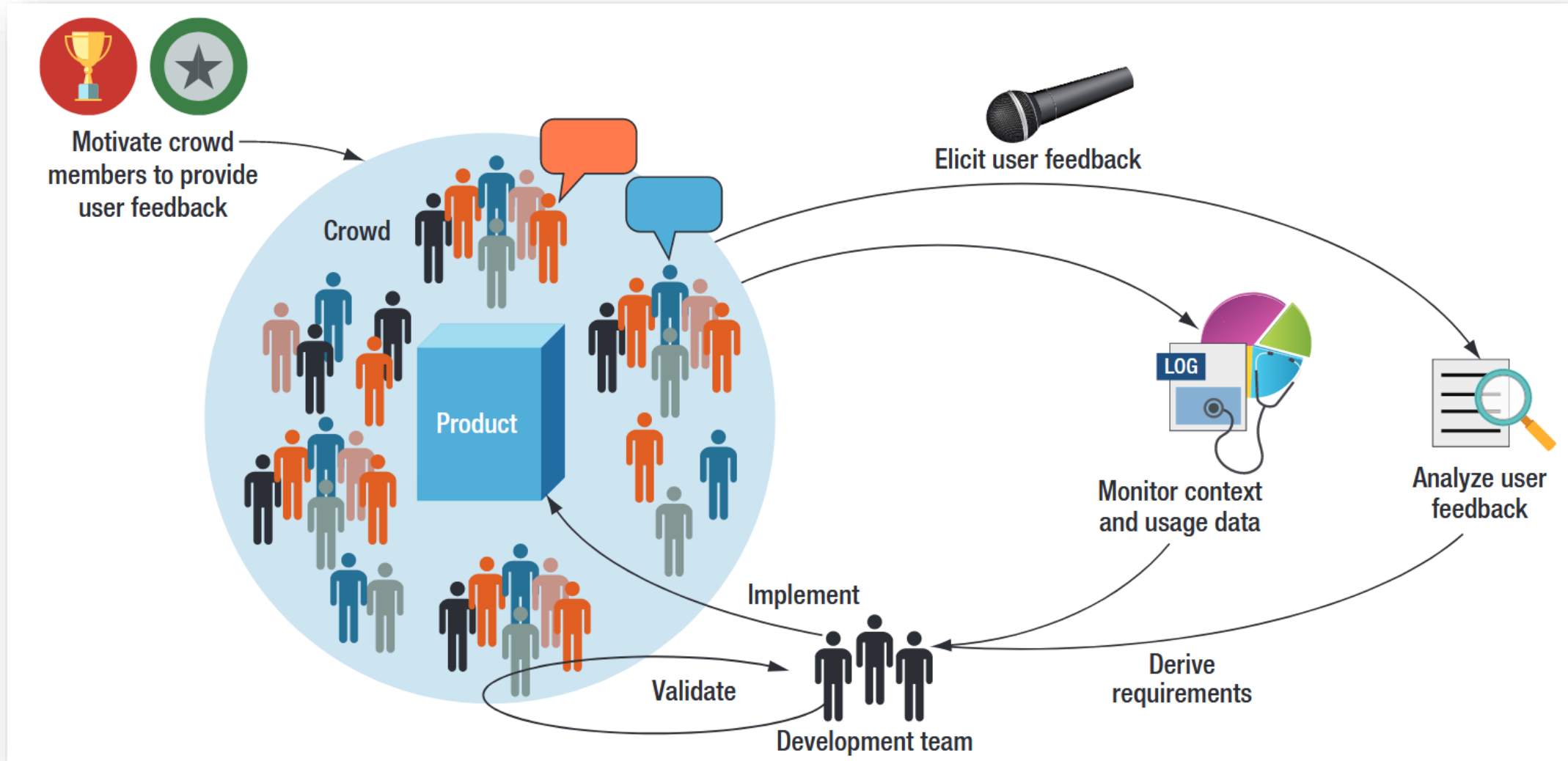
- ▶ **Pull feedback:** the company requests feedback to the users



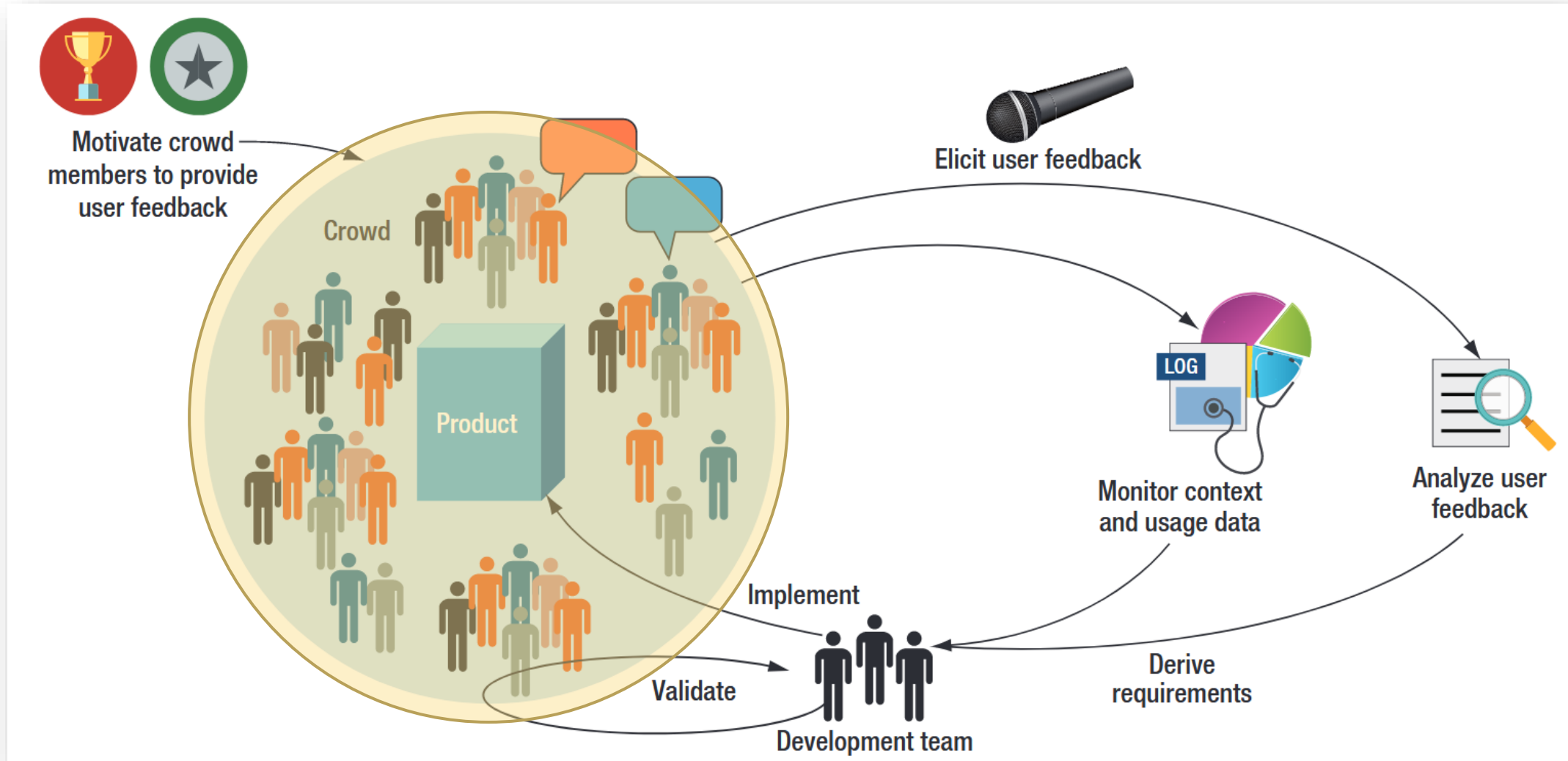
- ▶ **Push feedback:** the users spontaneously provide feedback



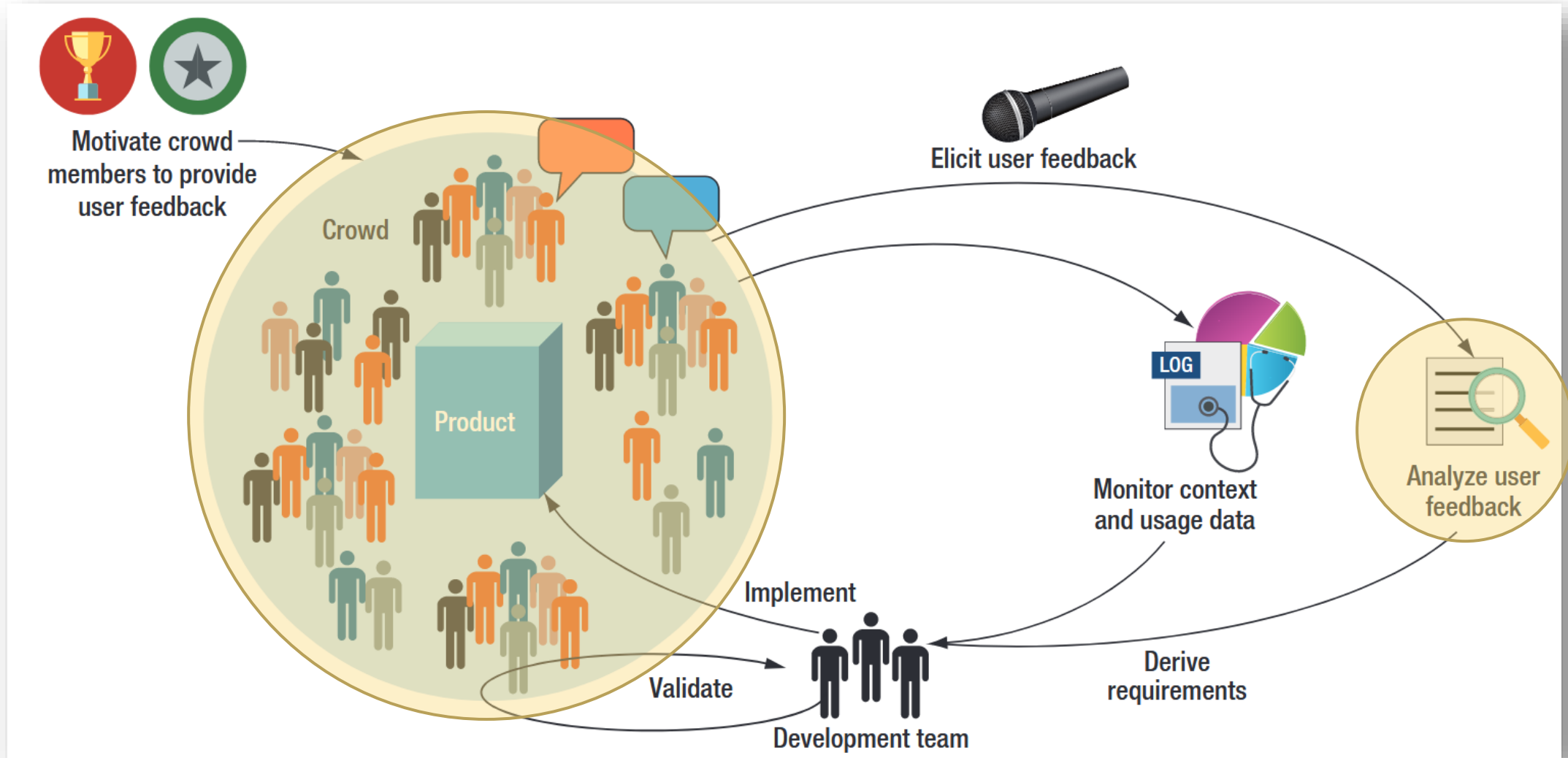
Human effort in CrowdRE



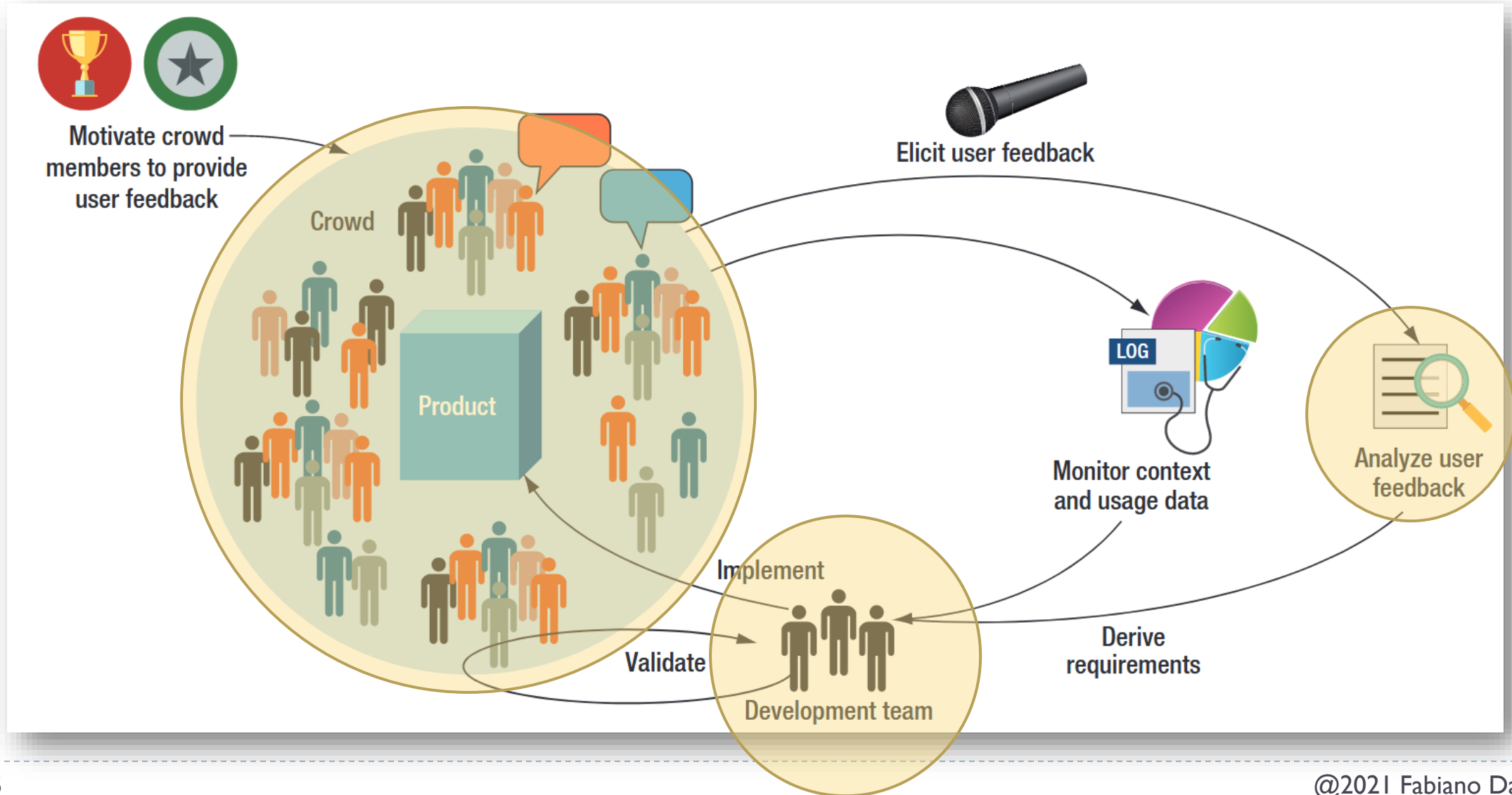
Human effort in CrowdRE



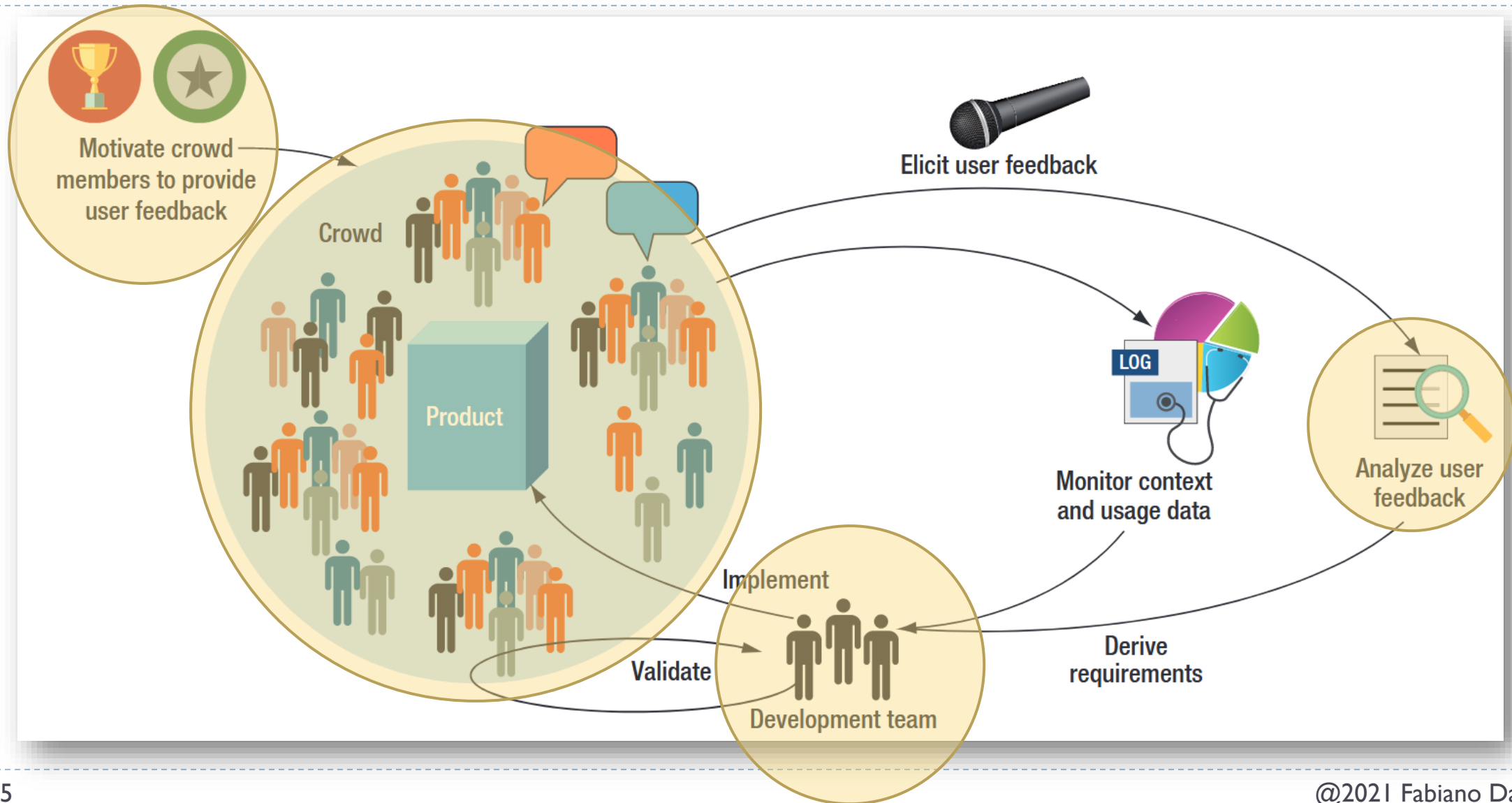
Human effort in CrowdRE



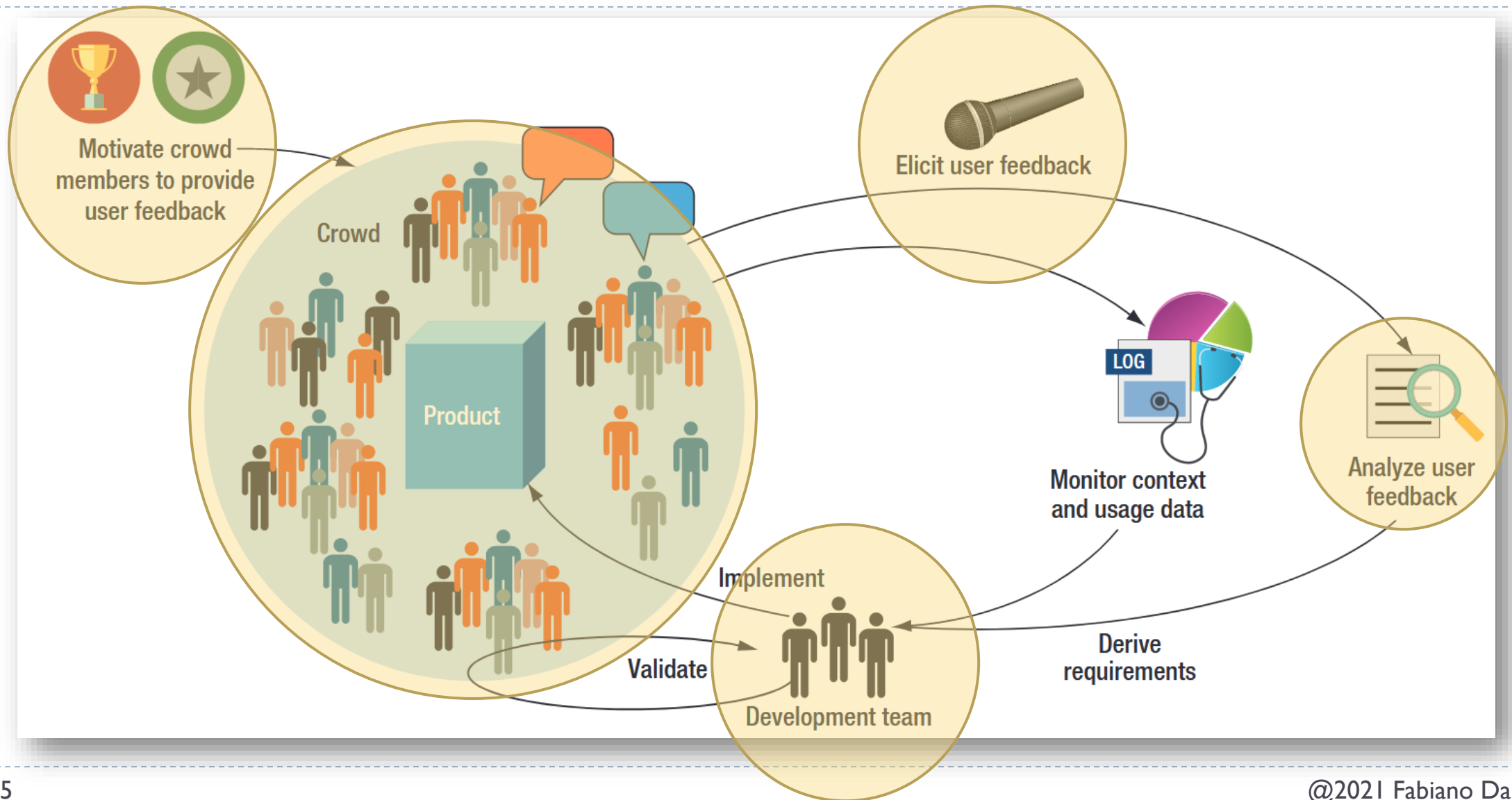
Human effort in CrowdRE



Human effort in CrowdRE



Human effort in CrowdRE



Research Question for the CrowdRE community

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



Automation to the rescue

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?

Automation to the rescue

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?

Requirements Eng (2016) 21:311–331
DOI 10.1007/s00766-016-0251-9

RE 2015

On the automatic classification of app reviews

Walid Maalej¹ · Zijad Kurtanović¹ · Hadeer Nabil² · Christoph Stanik¹

ML outperforms
keyword-based
approaches

Automation to the rescue

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?

Requirements Eng (2016) 21:311–331
DOI 10.1007/s00766-016-0251-9

RE 2015

On the automatic classification of app reviews

Walid Maalej¹ · Zijad Kurtanović¹ · Hadeer Nabil² · Christoph Stanik¹

ML outperforms
keyword-based
approaches

Superior accuracy
with less effort via
continuous training

App Review Analysis via Active Learning

Reducing Supervision Effort without Compromising Classification Accuracy

Venkatesh T. Dhinakaran¹, Raseshwari Pulle¹, Nirav Ajmeri², Pradeep K. Murukannaiah¹

Automation to the rescue

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?

Requirements Eng (2016) 21:311–331
DOI 10.1007/s00766-016-0251-9

RE 2015

On the automatic classification of app reviews

Walid Maalej¹ · Zijad Kurtanović¹ · Hadeer Nabil² · Christoph Stanik¹

ML outperforms
keyword-based
approaches

Superior accuracy
with less effort via
continuous training

App Review Analysis via Active Learning Reducing Supervision Effort without Compromising Classification Accuracy

Venkatesh T. Dhinakaran¹, Raseshwari Pulle¹, Nirav Ajmeri², Pradeep K. Murukannaiah¹

How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews

Emitza Guzman
Technische Universität München
Garching, Germany
emitza.guzman@mytum.de

Walid Maalej
University of Hamburg
Hamburg, Germany
maalej@informatik.uni-hamburg.de

What features are
mentioned? With
which sentiment?

Automation = Value for Research

High citation
figures

How do users like this feature? a fine grained sentiment analysis of app reviews

541

2014

E Guzman, W Maalej

2014 IEEE 22nd international requirements engineering conference (RE), 153-162

Bug report, feature request, or simply praise? on automatically classifying app reviews

387

2015

W Maalej, H Nabil

2015 IEEE 23rd international requirements engineering conference (RE), 116-125

How Can I Improve My App? Classifying User Reviews for Software Maintenance and Evolution

356

2015

S Panichella, AD Sorbo, E Guzman, A Visaggio, G Canfora, H Gall

Software Maintenance and Evolution (ICSME), 2015 IEEE International ...

Automation = Value for Research

High citation figures

How do users like this feature? a fine grained sentiment analysis of app reviews 541 2014
E Guzman, W Maalej
2014 IEEE 22nd international requirements engineering conference (RE), 153-162

Bug report, feature request, or simply praise? on automatically classifying app reviews 387 2015
W Maalej, H Nabil
2015 IEEE 23rd international requirements engineering conference (RE), 116-125

How Can I Improve My App? Classifying User Reviews for Software Maintenance and Evolution 356 2015
S Panichella, AD Sorbo, E Guzman, A Visaggio, G Canfora, H Gall
Software Maintenance and Evolution (ICSME), 2015 IEEE International ...

IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 43, NO. 9, SEPTEMBER 2017

A Survey of App Store Analysis for Software Engineering

William Martin, Federica Sarro, Yue Jia, Yuanyuan Zhang, and Mark Harman

An Overview of User Feedback Classification Approaches

Rubens Santos, Eduard C. Groen, Karina Villela
Fraunhofer IESE, Kaiserslautern, Germany

Extensive number of papers

Automation = Value for Research

High citation figures

How do users like this feature? a fine grained sentiment analysis of app reviews

541 2014

E Guzman, W Maalej

2014 IEEE 22nd international requirements engineering conference (RE), 153-162

Bug report, feature request, or simply praise? on automatically classifying app reviews

387 2015

W Maalej, H Nabil

2015 IEEE 23rd international requirements engineering conference (RE), 116-125

How Can I Improve My App? Classifying User Reviews for Software Maintenance and Evolution

356 2015

S Panichella, AD Sorbo, E Guzman, A Visaggio, G Canfora, H Gall

Software Maintenance and Evolution (ICSME), 2015 IEEE International ...

IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 43, NO. 9, SEPTEMBER 2017

A Survey of App Store Analysis for Software Engineering

William Martin, Federica Sarro, Yue Jia, Yuanyuan Zhang, and Mark Harman

An Overview of User Feedback Classification Approaches

Rubens Santos, Eduard C. Groen, Karina Villela
Fraunhofer IESE, Kaiserslautern, Germany

Extensive number of papers

Standard metrics to show success

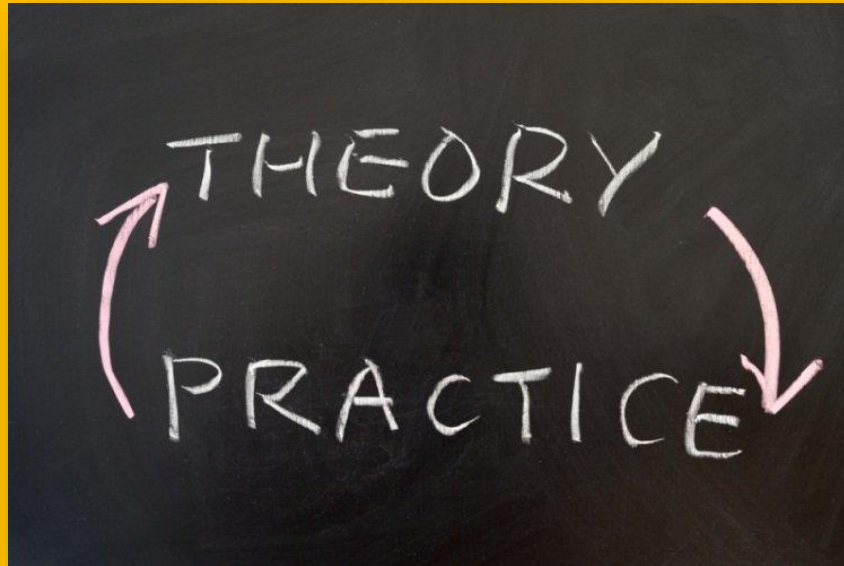
$$R = \frac{|\text{fnd} \cap \text{rel}|}{|\text{rel}|} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FN}|}$$

$$P = \frac{|\text{fnd} \cap \text{rel}|}{|\text{fnd}|} = \frac{|\text{TP}|}{|\text{FP}| + |\text{TP}|} \quad (2)$$

osite of recall and precision that is often called ' is captured by the *F*-measure:

$$F = 2 \times \frac{P \times R}{P + R} \quad (3)$$

2. How about value for practice?



The practitioner's standpoint: meet **Alex**




Value for practice in **push feedback**

Humans
Automation

Method	P1: Useless reviews			P1: Useful reviews		
	P	R	F_1	P	R	F_1
Crowdsourcing	0.93	0.84	0.88	0.83	0.93	0.88
SVM	0.90	0.79	0.84	0.83	0.92	0.87
Naïve Bayes	0.83	0.79	0.81	0.81	0.85	0.83
FastText	0.75	0.60	0.67	0.84	0.91	0.87
ELMo	0.83	0.80	0.82	0.81	0.84	0.82
BERT	0.95	0.88	0.92	0.88	0.96	0.92

Thu 23 Sep
Displayed time zone: **Eastern Time (US & Canada)** [change](#)
[Edit session information](#)

08:00 30m ☆ **Classifying User Requirements from Online Feedback in Small Dataset Environments using Deep Learning**
Talk
Research Papers [RESEARCH PAPER](#) 

Rohan Reddy Mekala Fraunhofer USA CESE,
Asif Irfan Fraunhofer USA Center Mid-Atlantic, Eduard C. Groen Fraunhofer IESE,
Adam Porter Fraunhofer USA CESE, Mikael Lindvall Fraunhofer USA CESE

[Media Attached](#)

Value for practice in **push feedback**

Humans



Method	P1: Useless reviews			P1: Useful reviews		
	P	R	F_1	P	R	F_1
Crowdsourcing	0.93	0.84	0.88	0.83	0.93	0.88
SVM	0.90	0.79	0.84	0.83	0.92	0.87
Naïve Bayes	0.83	0.79	0.81	0.81	0.85	0.83
FastText	0.75	0.60	0.67	0.84	0.91	0.87
ELMo	0.83	0.80	0.82	0.81	0.84	0.82
BERT	0.95	0.88	0.92	0.88	0.96	0.92

Automation



Value for practice in **push feedback**

	Method	P1: Useless reviews			P1: Useful reviews		
		<i>P</i>	<i>R</i>	<i>F</i> ₁	<i>P</i>	<i>R</i>	<i>F</i> ₁
Humans	Crowdsourcing	0.93	0.84	0.88	0.83	0.93	0.88
Automation	SVM	0.90	0.79	0.84	0.83	0.92	0.87
	Naïve Bayes	0.83	0.79	0.81	0.81	0.85	0.83
	FastText	0.75	0.60	0.67	0.84	0.91	0.87
	ELMo	0.83	0.80	0.82	0.81	0.84	0.82
	BERT	0.95	0.88	0.92	0.88	0.96	0.92

But I need experts to train ML/DL!



Value for practice in **push feedback**

Humans

Automation

Method	P1: Useless reviews			P1: Useful reviews		
	P	R	F_1	P	R	F_1
Crowdsourcing	0.93	0.84	0.88	0.83	0.93	0.88
SVM	0.90	0.79	0.84	0.83	0.92	0.87
Naïve Bayes	0.83	0.79	0.81	0.81	0.85	0.83
FastText	0.75	0.60	0.67	0.84	0.91	0.87
ELMo	0.83	0.80	0.82	0.81	0.84	0.82
BERT	0.95	0.88	0.92	0.88	0.96	0.92

How much better is this?



Value for practice in **push feedback**

Method	P1: Useless reviews			P1: Useful reviews		
	<i>P</i>	<i>R</i>	<i>F</i> ₁	<i>P</i>	<i>R</i>	<i>F</i> ₁
Crowdsourcing	0.93	0.84	0.88	0.83	0.93	0.88
SVM	0.90	0.79	0.84	0.83	0.92	0.87
Naïve Bayes	0.83	0.79	0.81	0.81	0.85	0.83
FastText	0.75	0.60	0.67	0.84	0.91	0.87
ELMo	0.83	0.80	0.82	0.81	0.84	0.82
BERT	0.95	0.88	0.92	0.88	0.96	0.92

Humans



Crowdsourcing

SVM

Naïve Bayes

FastText

ELMo

BERT

Automation



Value for practice in **pull feedback**

(i) *Amount of feedback data gathered and acceptance of the feedback tool.* SEnerCon received 81 feedback entries from 82 end-users in two weeks (seven feedback entries were excluded for further analysis due to invalid information, such as “test”). The feedback was mainly about (multiple-answers possible) ideas for improvements (37), computation error (17), others (17) and technical problems (13). “General error” was chosen in seven cases and “expert question” in four cases.

lack of knowledge on how software companies set up crowd involvement activities in practice. We observed a German software SME during their first steps in crowd-based feedback and monitoring data collection, and reported on their experiences during an initial implementation and evaluation phase.

We found that the end-users of the energy saving account communicate feedback very seldom and prefer a bilateral and non-public feedback communication with the software company. This could make it cumbersome for the software company

Stade et al., CrowdRE'19

Value for practice in **pull feedback**

(i) *Amount of feedback data gathered and acceptance of the feedback tool.* SENERCon received 81 feedback entries from 82 end-users in two weeks (seven feedback entries were excluded for further analysis due to invalid information, such as “test”). The feedback was mainly about (multiple-answers possible) ideas for improvements (37), computation error (17), others (17) and technical problems (13). “General error” was chosen in seven cases and “expert question” in four cases.

lack of knowledge on how software companies set up crowd involvement activities in practice. We observed a German software SME during their first steps in crowd-based feedback and monitoring data collection, and reported on their experiences during an initial implementation and evaluation phase.

We found that the end-users of the energy saving account communicate feedback very seldom and prefer a bilateral and non-public feedback communication with the software company. This could make it cumbersome for the software company

Stade et al., CrowdRE'19

Table 6 Overview of key figures

Platform visitors (see Table 7 and Fig. 10)	726	
Potential stakeholders (see Sect. 6.4.2)	244	
Stakeholders (see Sect. 6.4.3)	63	
Active stakeholders (see Sect. 6.4.3, Table 8, and Figs. 10 and 11)	32	
Users per expertise level (finally reached level; see Table 10)	I: 19; II: 3; III: 6; IV: 2; V: 2	
User activities	Total	Avg./User
Duration of participation (days from registration to last activity; see Table 9)	418	13.1
Number of logins (see Table 9)	177	5.5
Viewed posts (see Table 9)	468	14.6
Shared posts (see Table 10)	37	1.2
Shared sub-posts (see Table 10)	19	1.5
Ratings (see Table 10)	288	9
Votes on posts (see Table 10)	141	10.9
Votes on sub-posts (see Table 10)	19	1.9

Kolpondinos & Glinz, RE Journal 2020

Value for practice in **pull feedback**

(i) Amount of feedback data gathered and acceptance of the feedback tool. SENERCon received 81 feedback entries from 82 end-users in two weeks (seven feedback entries were excluded for further analysis due to ...)

The feedback was mainly ideas for improvements (3 (17) and technical problem in seven cases and “expert c

lack of knowledge on how involvement activities in software SME during their and monitoring data collec ences during an initial imple

We found that the end-communicate feedback very seldom and prefer a bilateral and non-public feedback communication with the software company. This could make it cumbersome for the software company

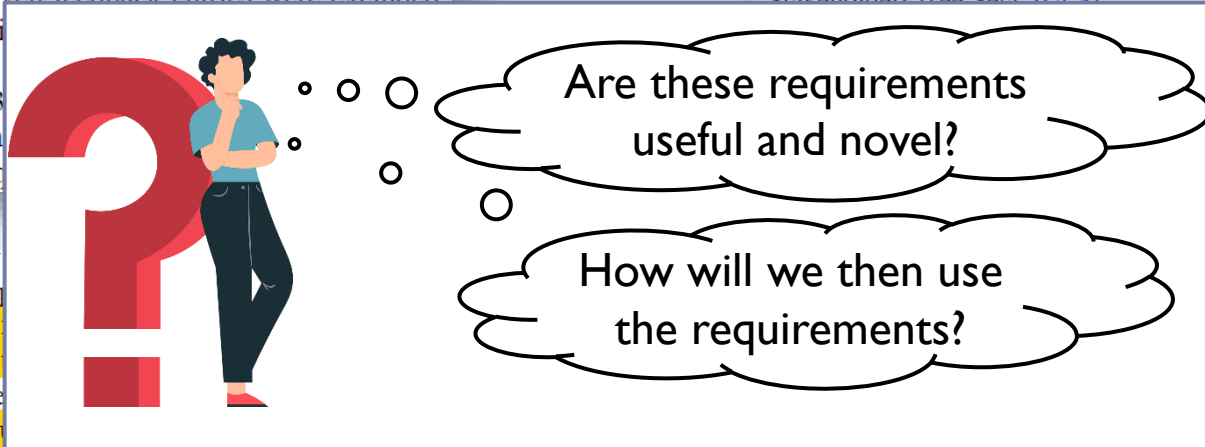


Table 6 Overview of key figures

Platform visitors (see Table 7 and Fig. 10)	726	
Potential stakeholders (see Sect. 6.4.2)	244	
Stakeholders (see Sect. 6.4.3)	63	
	32	
		I: 19; II: 3; III: 6; IV: 2; V: 2
	Total	Avg./User
	418	13.1
	177	5.5
	468	14.6
	37	1.2
Shared sub-posts (see Table 10)	19	1.5
Ratings (see Table 10)	288	9
Votes on posts (see Table 10)	141	10.9
Votes on sub-posts (see Table 10)	19	1.9

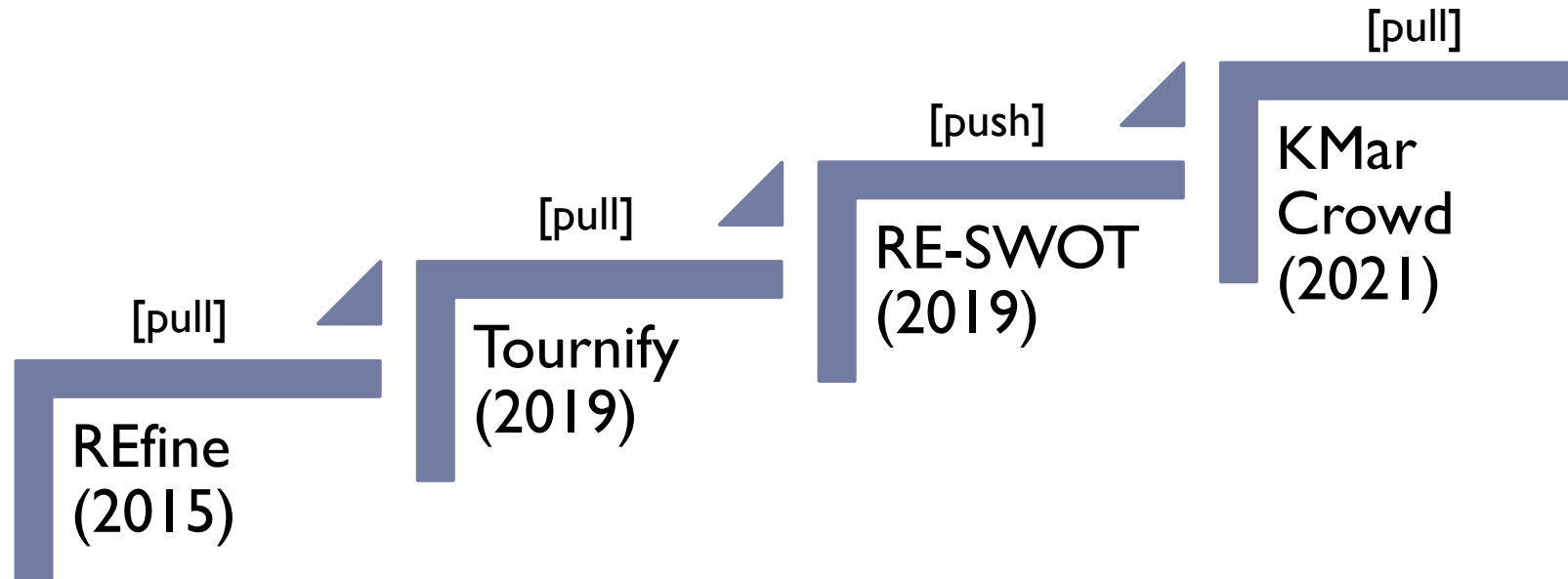
Stade et al., CrowdRE'19

Kolpondinos & Glinz, RE Journal 2020

3. A personal journey across research and practice



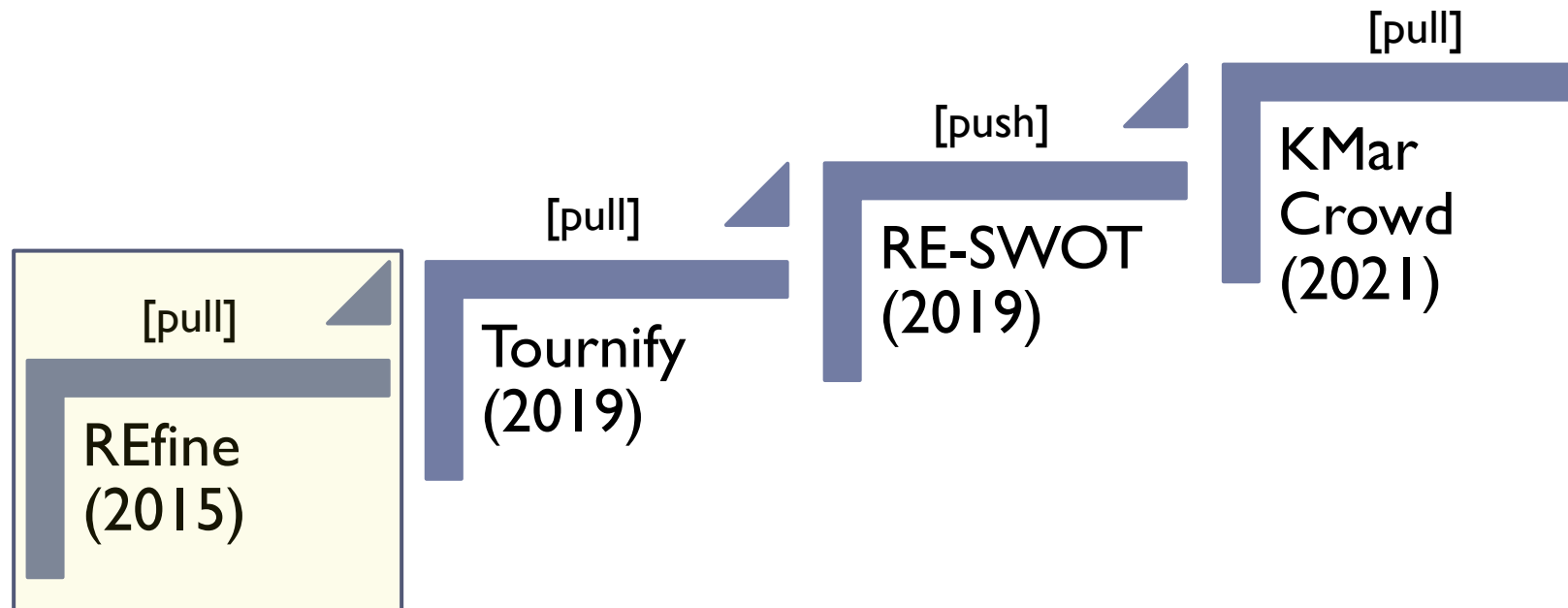
Personal Journey: Overview



Remco Snijders, Fabiano Dalpiaz, Sjaak Brinkkemper, Mahmood Hosseini, Raian Ali, Atilla Ozum:
REfine: A gamified platform for participatory requirements engineering. CrowdRE@RE 2015: 1-6



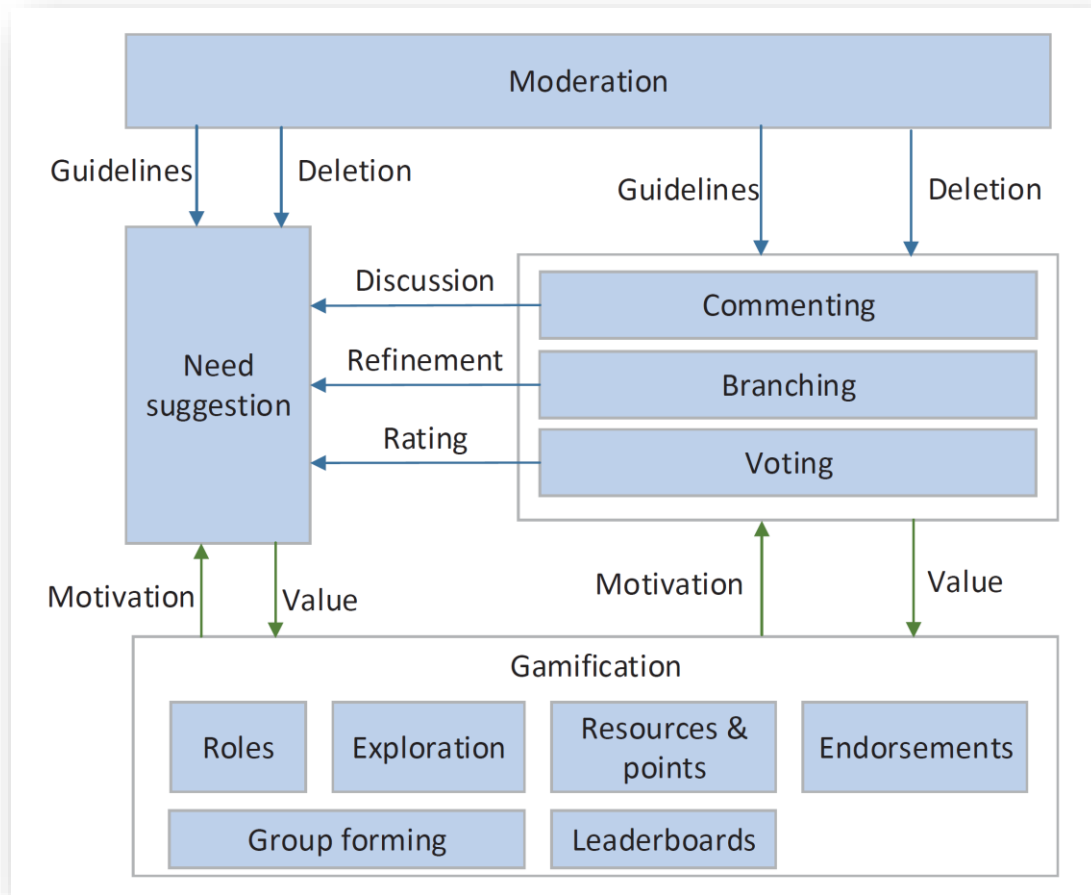
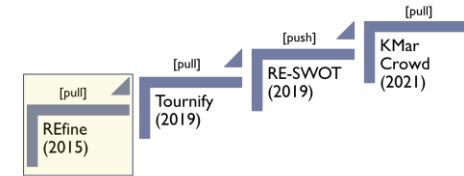
Personal Journey: Overview



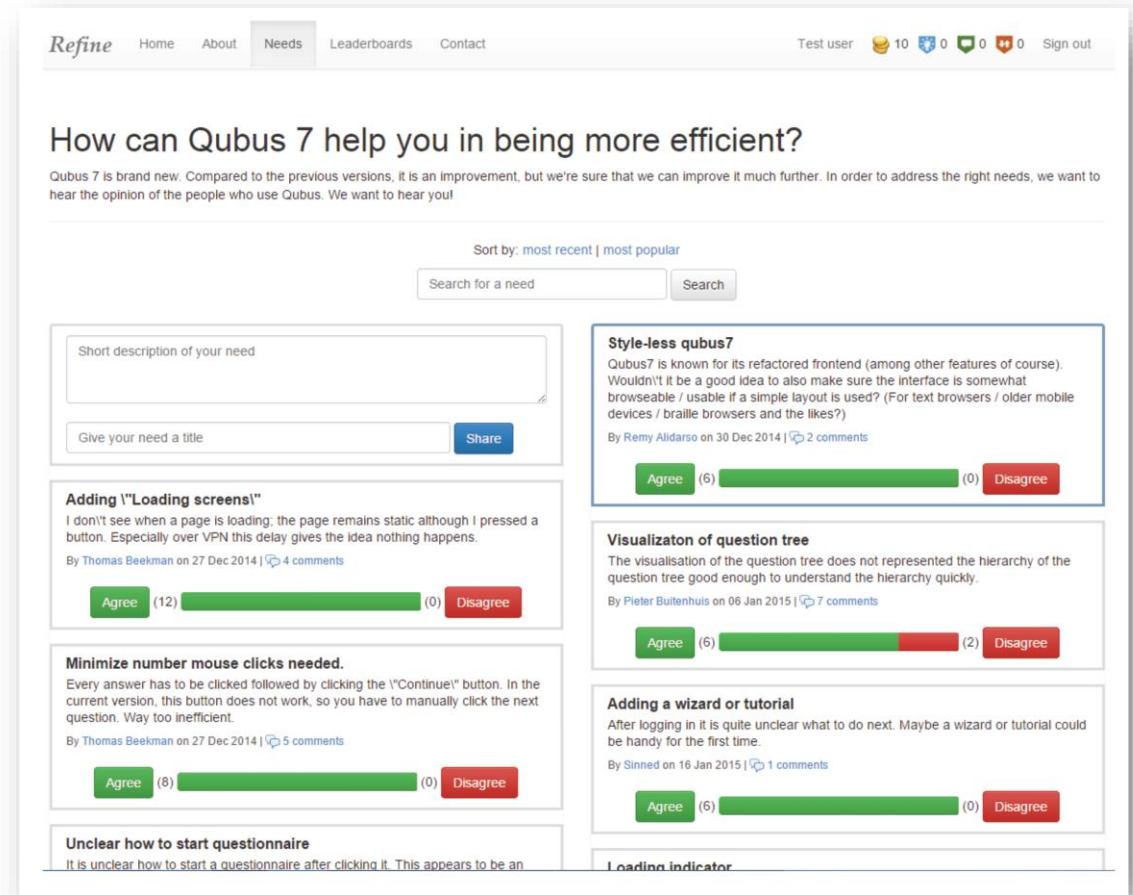
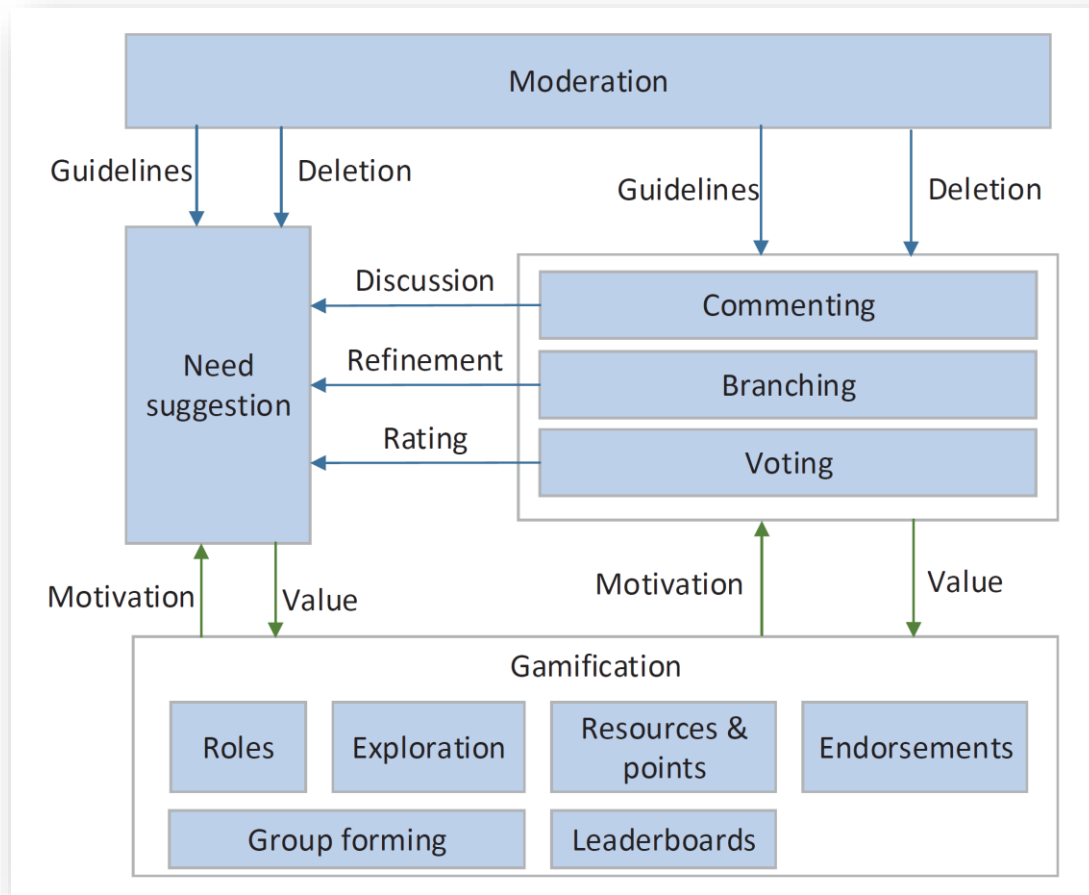
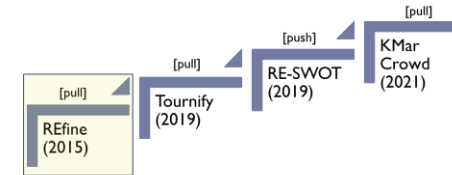
Remco Snijders, Fabiano Dalpiaz, Sjaak Brinkkemper, Mahmood Hosseini, Raian Ali, Atilla Ozum:
REfine: A gamified platform for participatory requirements engineering. CrowdRE@RE 2015: 1-6



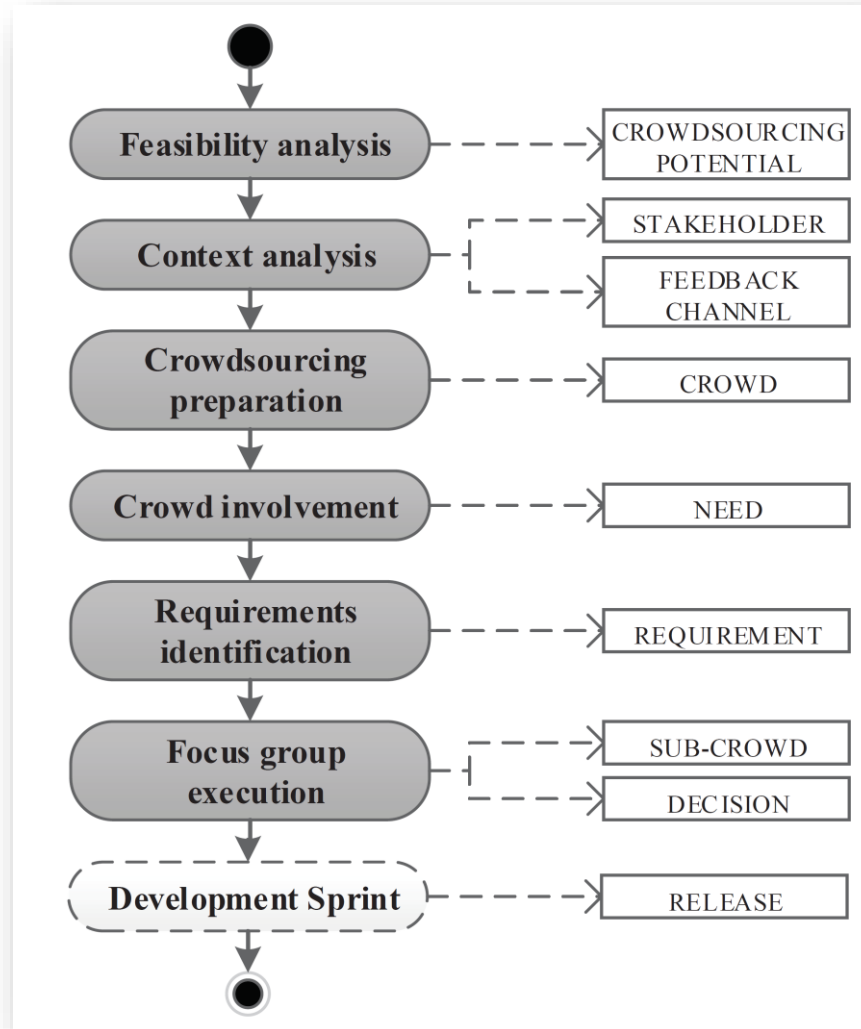
REfine: the start of a journey into CrowdRE



REfine: the start of a journey into CrowdRE

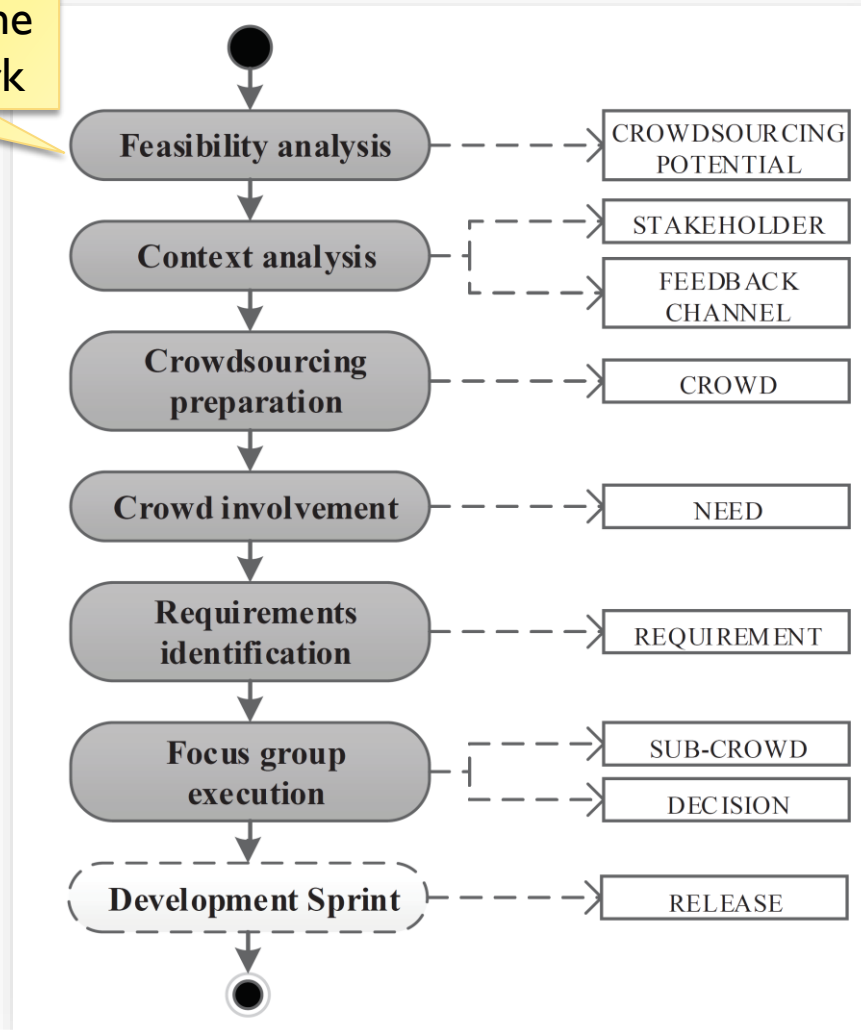


REfine and the **CCRE** method



REfine and the **CCRE** method

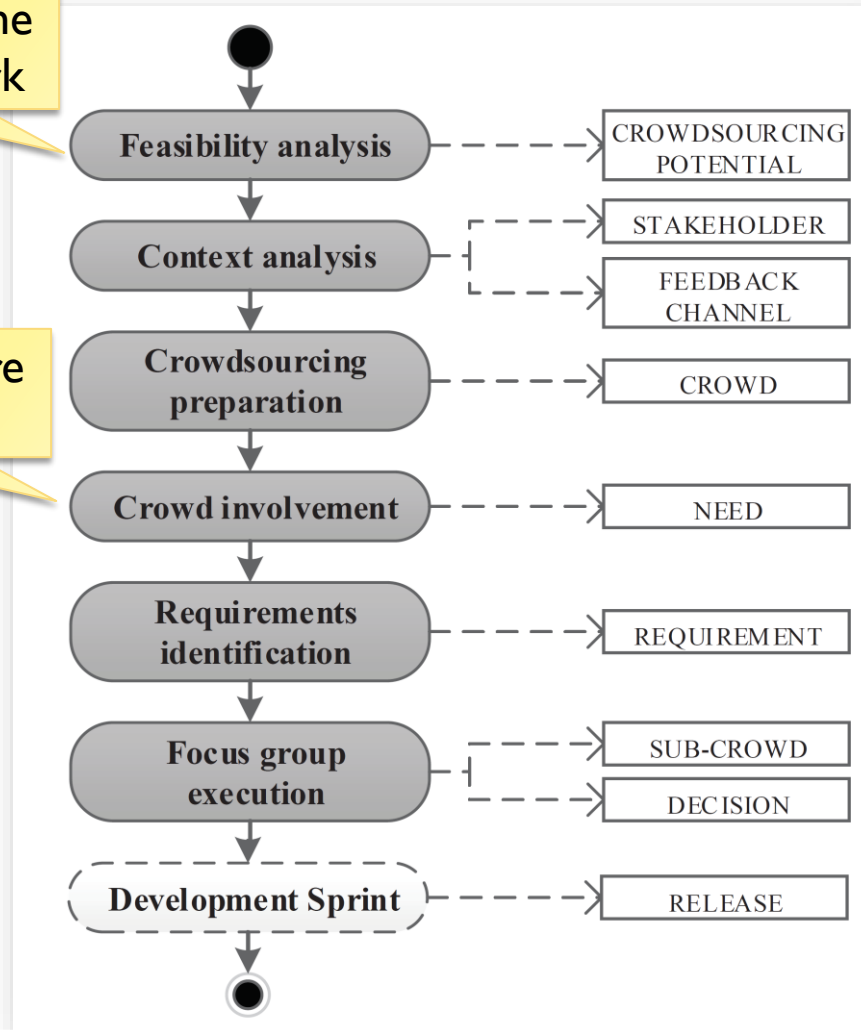
Don't always assume
CrowdRE will work



REfine and the **CCRE** method

Don't always assume CrowdRE will work

The only step where REfine is used

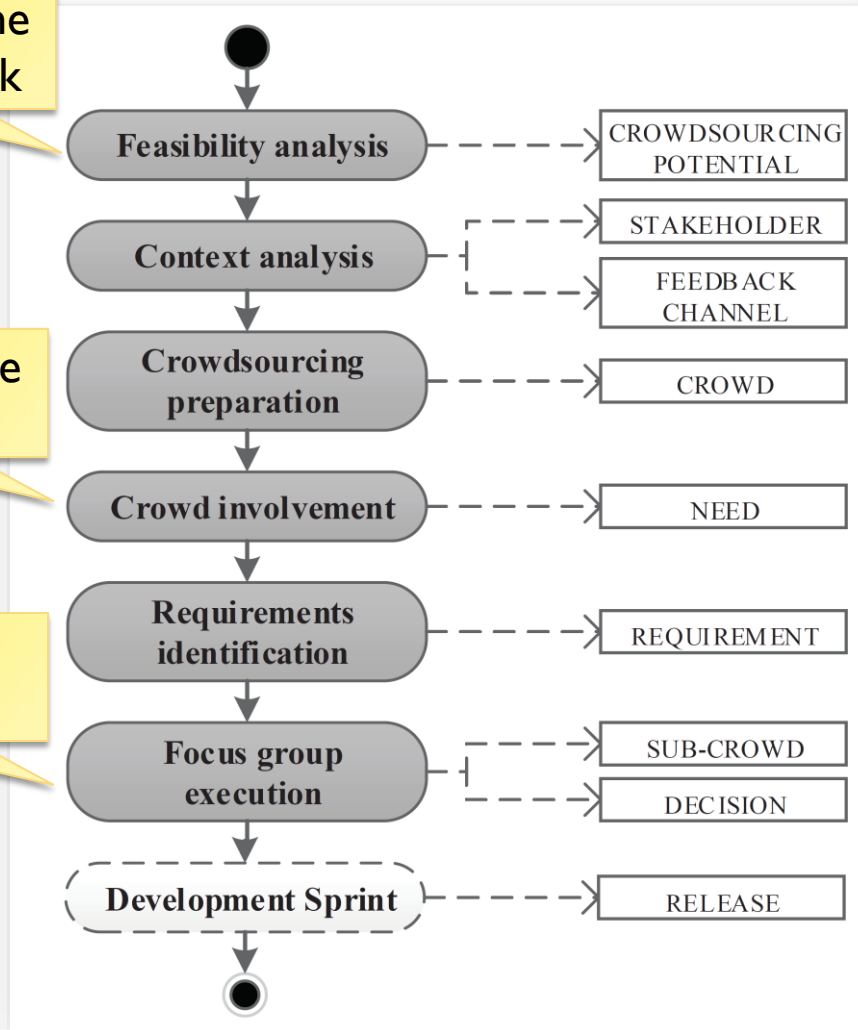


REfine and the **CCRE** method

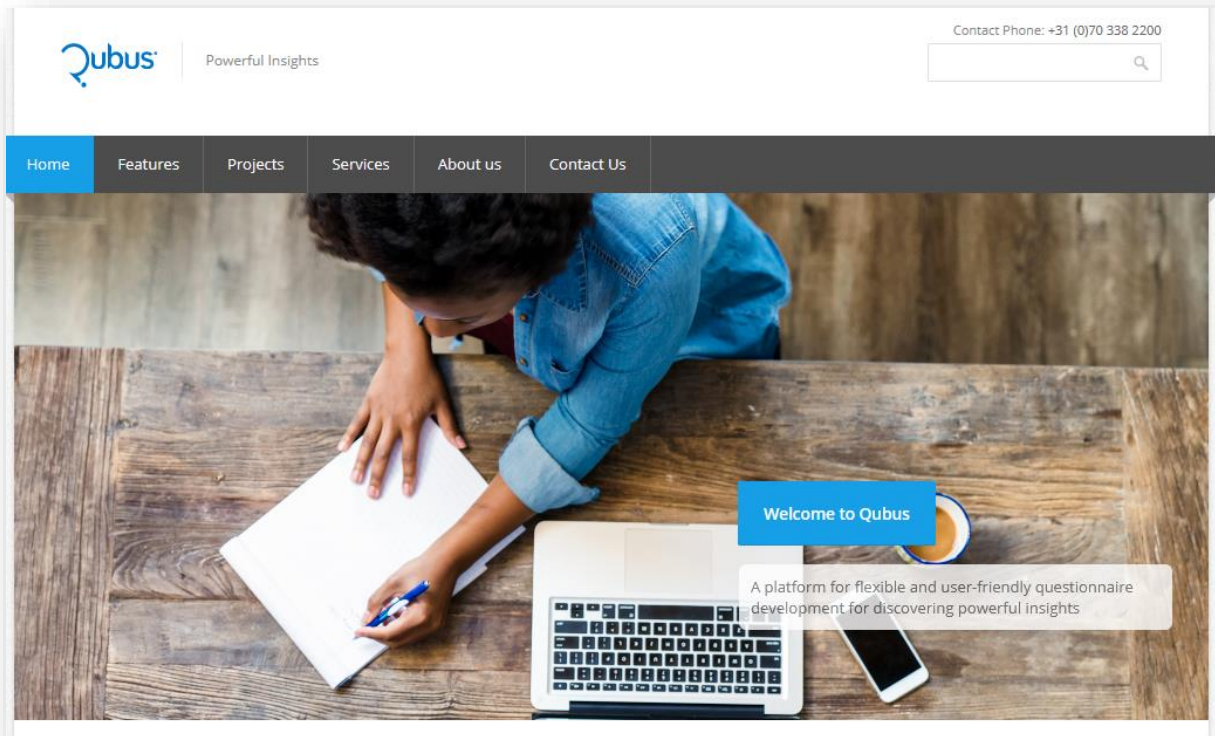
Don't always assume CrowdRE will work

The only step where REfine is used

Reward the most active participants



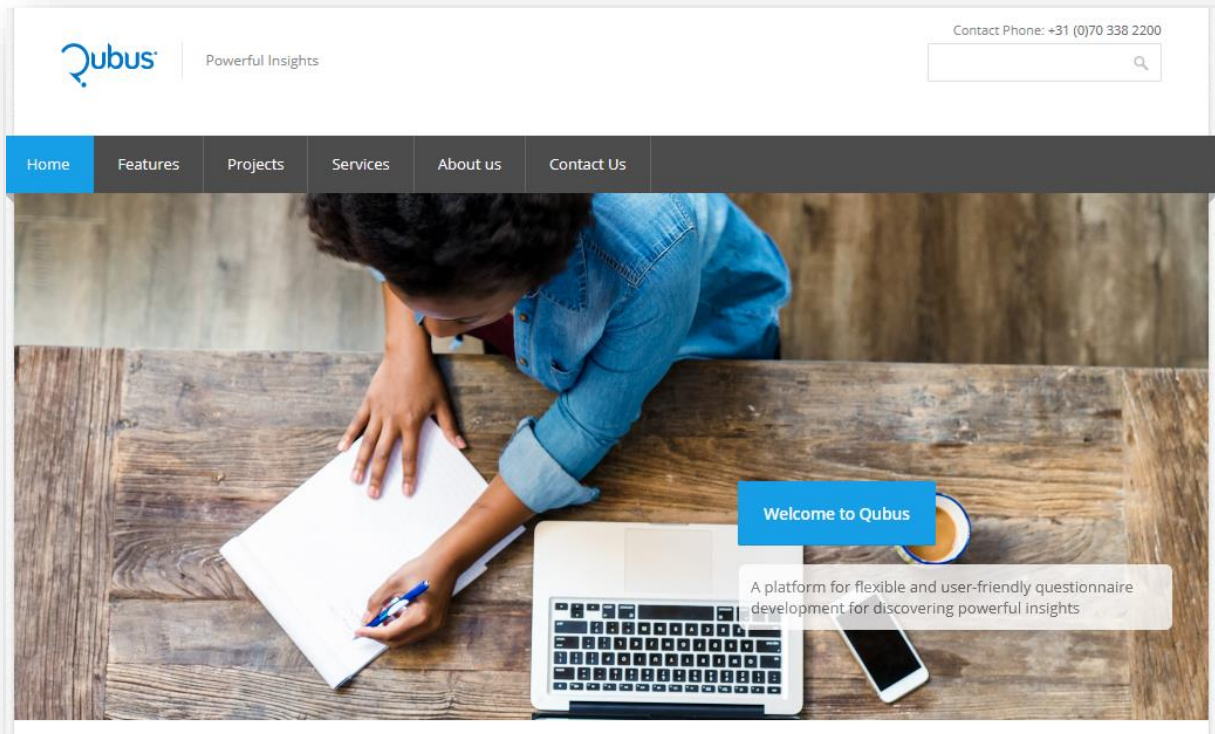
Case study: **KPMG's Qubus**



- ▶ Platform for compliance auditors
 - ▶ Questionnaire-based
 - ▶ Customizable workflow



Case study: **KPMG's Qubus**



- ▶ Platform for compliance auditors
 - ▶ Questionnaire-based
 - ▶ Customizable workflow
- ▶ **Focus of CrowdRE elicitation:**
Improved usability for mobile devices



Case study: KPMG's Qubus – the **users'** perspective

Stakeholder type	N	Needs		Comments		Votes		Points	
		avg.	tot.	avg.	tot.	avg.	tot.	avg.	tot.
Community manager	1	2	2	8	8	21	21	41	41
Product management	2	0	0	3	6	6	12	13.5	27
Development team	4	0.8	3	3.8	15	14.8	59	26.8	107
Expert	4	1	4	1.3	5	2.5	10	16	64
Client	1	0	0	0	0	3	3	3	3
End-User	1	1	1	0	0	3	3	11	11
Off-stage actor	6	1.8	11	0.5	3	4.2	25	20	120
<i>Total</i>	19	<i>1.1</i>	21	<i>1.9</i>	37	<i>6.8</i>	130	<i>19.6</i>	373



Case study: KPMG's Qubus – the **users' perspective**

Stakeholder type	N	Needs		Comments		Votes		Points	
		avg.	tot.	avg.	tot.	avg.	tot.	avg.	tot.
Community manager	1	2	2	8	8	21	21	41	41
Product management	2	0	0	3	6	6	12	13.5	27
Development team	4	0.8	3	3.8	15	14.8	59	26.8	107
Expert	4	1	4	1.3	5	2.5	10	16	64
Client	1	0	0	0	0	3	3	3	3
End-User	1	1	1	0	0	3	3	11	11
Off-stage actor	6	1.8	11	0.5	3	4.2	25	20	120
<i>Total</i>	19	<i>1.1</i>	21	<i>1.9</i>	37	<i>6.8</i>	130	<i>19.6</i>	373

Not that large, some members highly involved



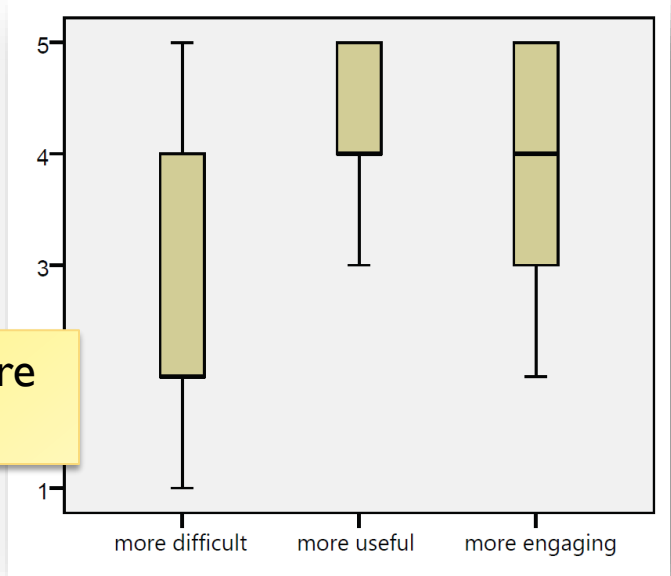
Case study: KPMG's Qubus – the **users' perspective**

Stakeholder type	N	Needs		Comments		Votes		Points	
		avg.	tot.	avg.	tot.	avg.	tot.	avg.	tot.
Community manager	1	2	2	8	8	21	21	41	41
Product management	2	0	0	3	6	6	12	13.5	27
Development team	4	0.8	3	3.8	15	14.8	59	26.8	107
Expert	4	1	4	1.3	5	2.5	10	16	64
Client	1	0	0	0	0	3	3	3	3
End-User	1	1	1	0	0	3	3	11	11
Off-stage actor	6	1.8	11	0.5	3	4.2	25	20	120
<i>Total</i>	19	<i>1.1</i>	21	<i>1.9</i>	37	<i>6.8</i>	130	<i>19.6</i>	373

Not that large, some members highly involved



Yet, the experience was more useful and more engaging



Case study: KPMG's Qubus – the **experts perspective**

- + Approachable method
- + Mostly useful for elicitation
- + Expected user adoption



Product managers
(2 Qubus, 3 external)



Case study: KPMG's Qubus – the **experts perspective**

- + Approachable method
- + Mostly useful for elicitation
- + Expected user adoption

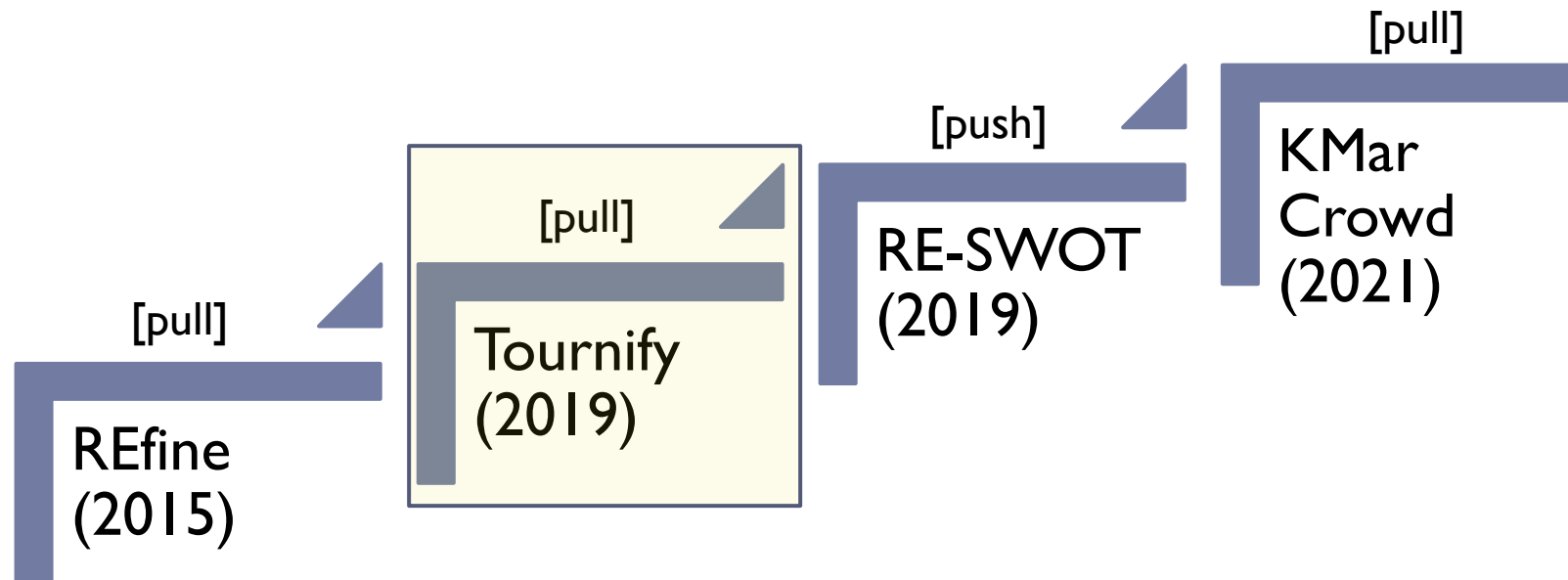


Product managers
(2 Qubus, 3 external)

- Little incentive to return
- Possibly trivial needs
- Needs inadequate for backlog



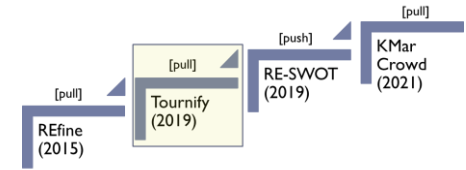
Personal Journey: Overview



Abel Menkveld, Sjaak Brinkkemper, Fabiano Dalpiaz: User Story Writing in Crowd Requirements Engineering: The Case of a Web Application for Sports Tournament Planning. RE Workshops 2019: 174-179



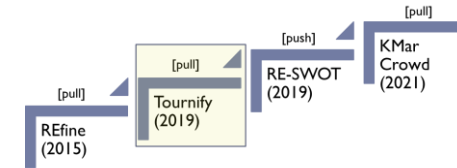
Tournify – User Story Writing




Product managers
(2 Qubus, 3 external)

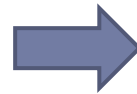
- Little incentive to return
- Possibly trivial needs
- Needs inadequate for backlog

Tournify – User Story Writing



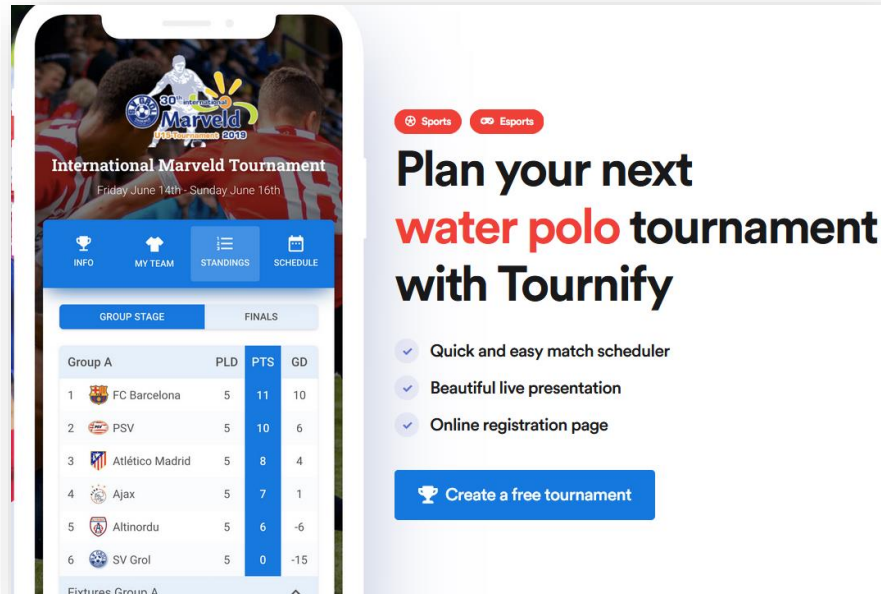

Product managers
(2 Qubus, 3 external)

- Little incentive to return
- Possibly trivial needs
- Needs inadequate for backlog



Can we obtain **more specific needs** by asking the crowd to write user stories?

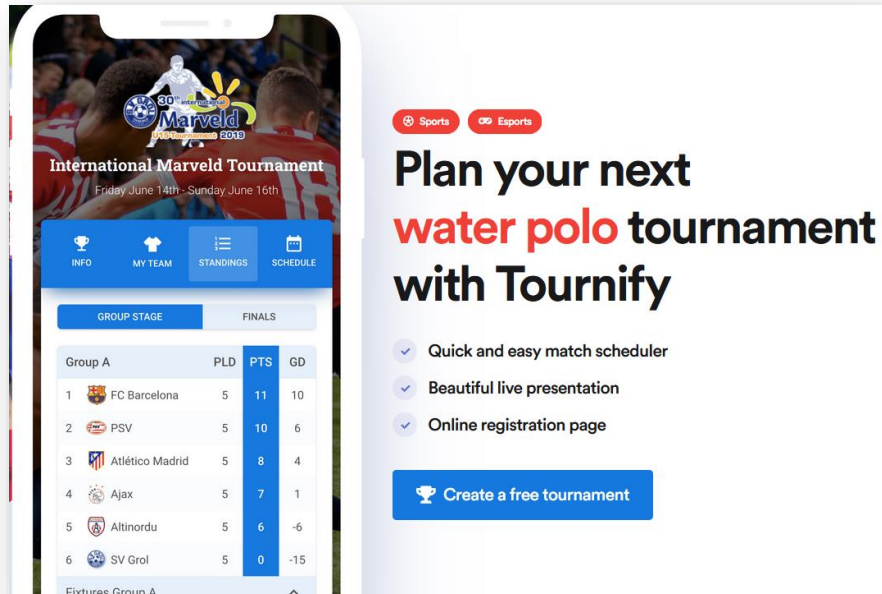
Tournify – Current Situation



- ▶ Previously, user feedback was collected via e-mail and a support chat

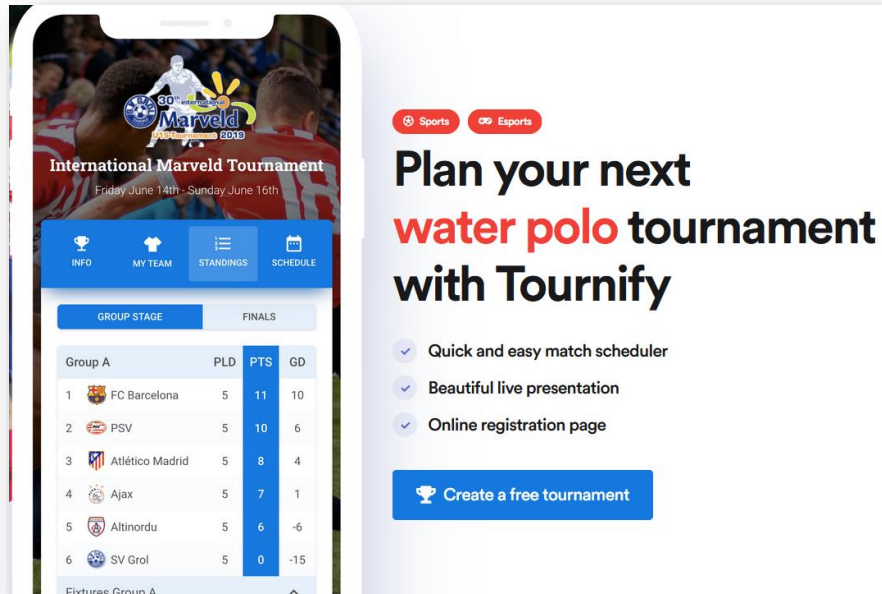


Tournify – Current Situation



- ▶ Previously, user feedback was collected via e-mail and a support chat
- ▶ The company adopts agile dev with user stories
 - ▶ For ~70% of the feedback, converting user feedback to user stories was easy/possible

Tournify – Current Situation



- ▶ Previously, user feedback was collected via e-mail and a support chat
- ▶ The company adopts agile dev with user stories
 - ▶ For ~70% of the feedback, converting user feedback to user stories was easy/possible
 - ▶ **However, the reason behind feature requests is often omitted**

Tournify – A CrowdRE Elicitation Platform

The screenshot displays the 'Feature Requests' section of the Tournify platform. The interface has a blue header with a back arrow, the title 'Feature Requests', and links for 'ENGLISH' and 'ACCOUNT'. The main content area is a grid of request cards. Each card contains a text description of a feature request, a date and category (e.g., '26 February 2019 - Presentation'), and interaction icons for comments and likes. One card is marked 'In development' with an orange pill. The requests are as follows:

- Request 1:** 'Als organisator wil ik de mogelijkheid om de ploegen zelf als scheidrechter te laten fungeren en dus mee te laten nemen in de berekening van de schema's, zodat er geen wedstrijden doorgaan waarvan de ploegen op dezelfde moment moeten arbitreran'. (26 February 2019 - General, 1 comment, 3 likes)
- Request 2:** 'wedstrijden per veld, zodat de wedstrijdsecretaris het schema alleen hoeft op te lezen zonder nadenken'. (26 February 2019 - Presentation, 1 comment, 3 likes)
- Request 3:** 'Als organisator wil ik dat bij de sheets voor de schermpresentaties start en eindtijden worden toegevoegd, zodat je kunt zorgen dat een sheet alleen op de gewenste tijdstippen wordt getoond zoals bijvoorbeeld voor het toernooi de welkomboodschap of bijv tijdens lunchtijd.'. (26 February 2019 - Presentation, 3 likes)
- Request 4:** 'Als organisator wil ik de dia's met een sleepfunctie of wijsbare volgordenummering kunnen laten wijzigen van volgorde, zodat je de presentatie niet volledig opnieuw hoeft te maken als er een extra dia tussenkomt.'. (26 February 2019 - Presentation, 5 likes)
- Request 5:** 'Als organisator wil ik dat fase 1 niet meer getoond wordt zodra de volgende fase begint, zodat deelnemers niet gedwongen worden om zelf te klikken op de volgende'. (26 February 2019 - Presentation, 3 likes)
- Request 6:** 'start aan altijd aan de rechterkant.'. (27 February 2019 - General)
- Request 7:** 'Als organisator wil ik graag de mogelijkheid om logo's toe te voegen van de deelnemende ploegen. Dit zoals het vlaggetje voor de deelnemersnaam maar dan hun eigen logo, zodat de logo's ook mooi kunnen getoond worden in de wedstrijdsschema's en standen'. (26 February 2019 - Presentation, 2 comments, 9 likes, marked 'In development')
- Request 8:** 'Als organisator wil ik graag de optie toevoegen om te kunnen kiezen voor een achtergrondkleur ipv het toevoegen van een achtergrondplaatje'. (26 February 2019 - Presentation, 2 comments, 9 likes)

Tournify – A CrowdRE Elicitation Platform

The screenshot displays the Tournify platform interface, which is used for crowd-referenced elicitation. It is divided into two main sections: a 'Feature Requests' sidebar and a central 'New idea' workflow.

Feature Requests Sidebar:

- Request 1:** "Als organisator wil ik de mogelijkheid om de ploegen zelf als scheidrechter te laten fungeren en dus mee te laten nemen in de berekening van de schema's, zodat er geen wedstrijden doorgaan waarvan de ploegen op dezelfde moment moeten arbitreran". (26 February 2019 - General, 1 comment, 3 likes)
- Request 2:** "Als organisator wil ik dat bij de sheets voor de schermpresentaties start en eindtijden worden toegevoegd, zodat je kunt zorgen dat een sheet alleen op de gewenste tijdstippen wordt getoond zoals bijvoorbeeld voor het toernooi de welkomboodschap of bijv tijdens lunchtijd." (26 February 2019 - Presentation, 3 likes)
- Request 3:** "wedstrijden per veld, zodat de wedstrijdsecretaris het schema alleen hoeft op te lezen zonder nadenken". (26 February 2019 - Presentation, 1 comment, 3 likes)
- Request 4:** "Als organisator wil ik de dia's met een sleepfunctie of wijsbare volgordenummering kunnen laten wijzigen van volgorde, zodat je de presentatie niet volledig opnieuw hoeft te maken als er een extra dia tussenkomt." (26 February 2019 - Presentation, 5 likes)
- Request 5:** "Als organisator wil ik dat fase 1 niet meer getoond wordt zodra de volgende fase begint, zodat deelnemers niet gedwongen worden om zelf te klikken op de volgende start aan altijd aan de rechterkant." (26 February 2019 - Presentation, 5 likes)

New idea Workflow (4 steps):

- Step 1 of 4:** "How do you use Tournify?"
 - I organize a tournament
 - I participate in a tournament
 - I support a team in a tournament(TO STEP 2)
- Step 2 of 4:** "What would you like to do with Tournify?"

I want to export the match schedule to PDF

(STEP BACK, TO STEP 3)
- Step 3 of 4:** "Why do you want to export the match schedule to PDF?"

So that I can print it out and have it as a backup in case the connection is lost.

(STEP BACK, TO STEP 4)
- Step 4 of 4:** "Verify your idea, choose a category and submit it."

As an organizer I want to export the match schedule to PDF, so that I can print it out and have it as a backup in case the connection is lost.

General Participants Format Schedule Presentation **Results** Other

(STEP BACK, SUBMIT IDEA)

Tournify – A CrowdRE Elicitation Platform

The screenshot displays the Tournify platform interface, which is used for eliciting requirements through a series of steps. The interface is divided into several sections:

- Feature Requests:** A list of user requests, such as "Als organisator wil ik de mogelijkheid om de ploegen zelf als scheidrechter te laten fungeren en dus mee te laten nemen in de berekening van de schema's, zodat er geen wedstrijden doorgaan waarvan de ploegen op dezelfde moment moeten arbitreran".
- Idea Elicitation Steps:** A series of steps for creating a new idea, including:
 - Step 1:** "How do you use Tournify?" with radio button options: "I organize a tournament" (selected), "I participate in a tournament", and "I support a team in a tournament".
 - Step 2:** "What would you like to do with Tournify?" with a text input field containing "I want to export the match schedule to PDF".
 - Step 3:** "Why do you want to export the match schedule to PDF?" with a text input field containing "So that I can print it out and have it as a backup in case the connection is lost".
 - Step 4:** "Verify your idea, choose a category and submit it." with a text input field containing "As an organizer I want to export the match schedule to PDF, so that I can print it out and have it as a backup in case the connection is lost." and a category selection menu with options: "General", "Participants", "Format", "Schedule", "Presentation", "Results" (selected), and "Other".

A yellow callout box labeled "Pre-defined roles" points to the radio button options in Step 1, indicating that these roles are predefined in the system.

Tournify – A CrowdRE Elicitation Platform

The image displays the Tournify platform interface, which is used for eliciting requirements through a crowd. The interface is divided into several sections:

- Feature Requests:** A list of user-submitted requests. For example, "Als organisator wil ik de mogelijkheid om de ploegen zelf als scheidrechter te laten fungeren en dus mee te laten nemen in de berekening van de schema's, zodat er geen wedstrijden doorgaan waarvan de ploegen op dezelfde moment moeten arbitreran".
- New idea (Step 1 of 4):** A form where users select a role. The question is "How do you use Tournify?". The options are:
 - I organize a tournament
 - I participate in a tournament
 - I support a team in a tournamentA yellow callout bubble labeled "Pre-defined roles" points to these options.
- New idea (Step 2 of 4):** A form where users describe what they want to do. The question is "What would you like to do with Tournify?". The user has entered "I want to export the match schedule to PDF". A yellow callout bubble labeled "What feature?" points to this text.
- New idea (Step 3 of 4):** A form where users explain why they want a feature. The question is "Why do you want to export the match schedule to PDF?". The user has entered "So that I can print it out and have it as a backup in case the connection is lost." A yellow callout bubble labeled "Why?" points to this text.
- New idea (Step 4 of 4):** A form where users verify their idea, choose a category, and submit it. The user has selected the "Results" category from a list including General, Participants, Format, Schedule, Presentation, and Results.

Tournify – A CrowdRE Elicitation Platform

The screenshot displays the Tournify platform interface, which is used for eliciting requirements through a crowd. The interface is divided into several sections:

- Feature Requests:** A list of user-submitted requests, such as "Als organisator wil ik de mogelijkheid om de ploegen zelf als scheidrechter te laten fungeren en dus mee te laten nemen in de berekening van de schema's, zodat er geen wedstrijden doorgaan waarvan de ploegen op dezelfde moment moeten arbitreran".
- New idea (Step 1 of 4):** A form where users select a role. The options are: "I organize a tournament" (selected), "I participate in a tournament", and "I support a team in a tournament". A yellow callout bubble labeled "Pre-defined roles" points to these options.
- New idea (Step 2 of 4):** A form where users describe what they want to do. The question is "What would you like to do with Tournify?". The user's response is "I want to export the match schedule to PDF". A yellow callout bubble labeled "What feature?" points to this text.
- New idea (Step 3 of 4):** A form where users explain why they want a feature. The question is "Why do you want to export the match schedule to PDF?". The user's response is "So that I can print it out and have it as a backup in case the connection is lost." A yellow callout bubble labeled "Why?" points to this text.
- New idea (Step 4 of 4):** A form where users verify their idea, choose a category, and submit it. The user's text is "As an organizer I want to export the match schedule to PDF, so that I can print it out and have it as a backup in case the connection is lost." Below the text are tags: "General", "Participants", "Format", "Schedule", "Presentation", "Results" (selected), and "Other". A yellow callout bubble labeled "Tags as metadata" points to these tags.

Tournify – Case Study Execution

- ▶ **337 invited participants**
 - ▶ Previous feature requesters
 - ▶ Newsletter subscribers
 - ▶ Recent paying clients
- ▶ **Data collection for 5 weeks**
 - ▶ Same duration as REfine, but users outside the organization
- ▶ **Users informed via a snack bar message on the Tournify page**

Tournify – Results

Value	Total	Unique users
Page views	247	157
Interactions	160	39
Requests	57	23
Votes	89	28
Comments	14	9



Tournify – Results

Larger crowd
than REfine's

Value	Total	Unique users
Page views	247	157
Interactions	160	39
Requests	57	23
Votes	89	28
Comments	14	9



Tournify – Results

Larger crowd
than REfine's

These are not
perfectly written
requirements...

Value	Total	Unique users
Page views	247	157
Interactions	160	39
Requests	57	23
Votes	89	28
Comments	14	9



criteria	# USs with defect	% USs with defect
Well-formed	3	5.4
Atomic	5	8.9
Minimal	24	42.9
Conceptual	5	8.9
Problem-oriented	8	14.3
Unambiguous	9	16.0
Full sentence	19	33.9
Estimatable	3	5.4

Tournify – Results

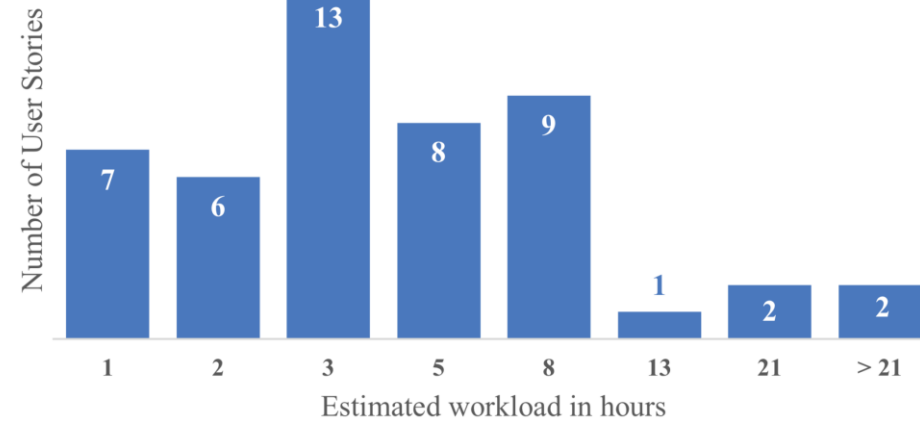
Larger crowd than REfine's

These are not perfectly written requirements...

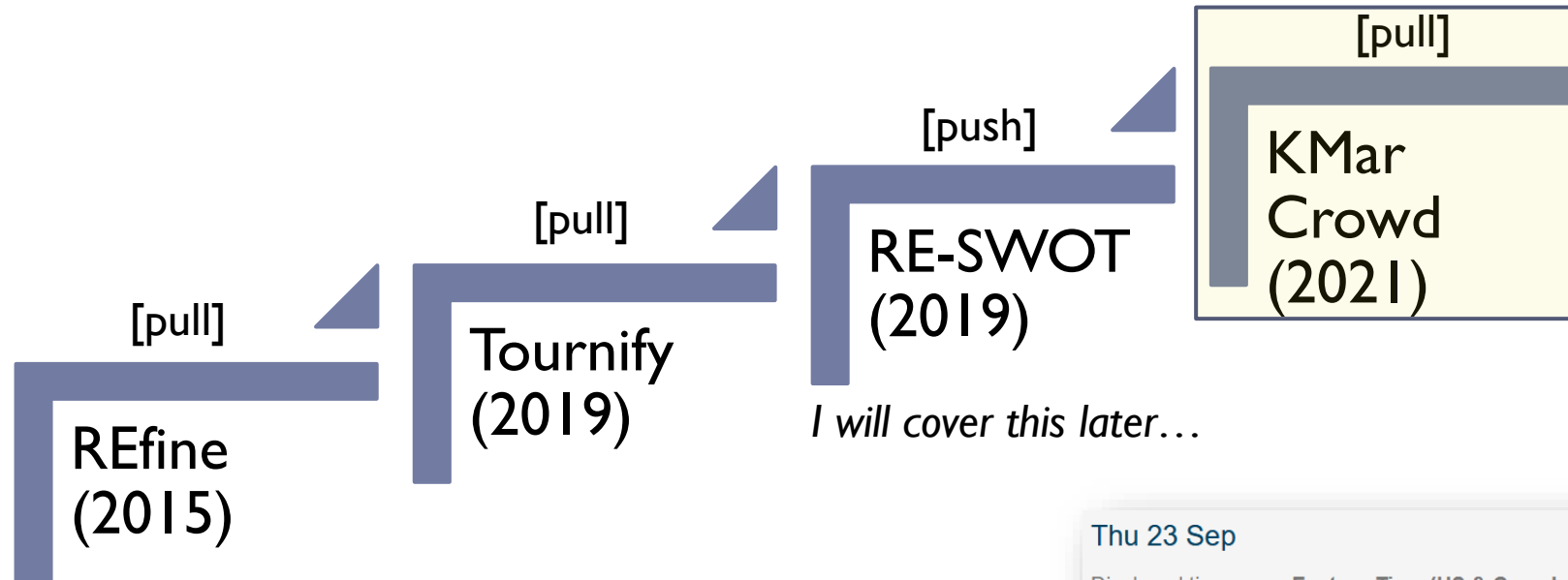
Yet, most are sufficiently specific

critierion	# USs with defect	% USs with defect
Well-formed	3	5.4
Atomic	5	8.9
Minimal	24	42.9
Conceptual	5	8.9
Problem-oriented	8	14.3
Unambiguous	9	16.0
Full sentence	19	33.9
Estimatable	3	5.4

Value	Total	Unique users
Page views	247	157
Interactions	160	39
Requests	57	23
Votes	89	28
Comments	14	9



Personal Journey: Overview



Thu 23 Sep

Displayed time zone: **Eastern Time (US & Canada)** [change](#)

10:00 30m ☆ **CrowdRE in a Governmental Setting: Lessons from Two Case Studies** INDUSTRIAL INNOVATION PAPER

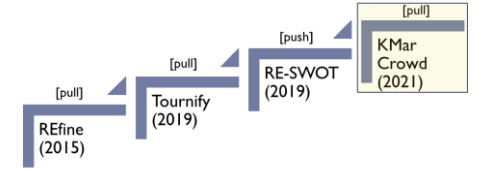
Talk

Industrial Innovation Papers

Jelle Wouters Royal Netherlands Marechaussee, Rob Janssen Royal Netherlands Marechaussee, Bas van Hulst Royal Netherlands Marechaussee, John van Veenhuizen Royal Netherlands Marechaussee, Fabiano Dalpiaz Utrecht University, Sjaak Brinkkemper Utrecht University

[Pre-print](#)

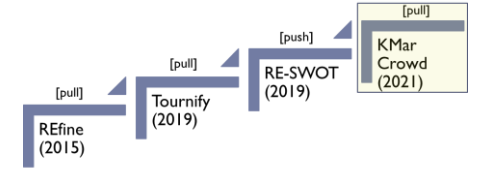
KMar Crowd – Premises



Measurement	Menkveld	REFine	GARUSO
Duration in days	35	35	92
Participants:			
- Invited	unk.	37	unk.
- Accessed	157	19	726
- Active	39	19	32
Ideas	57	21	56
Logins	247	unk.	unk.
Votes	89	130	160
Comments	14	37	unk.
Ideas / Accessed	0.36	1.11	0.08



KMar Crowd – Premises

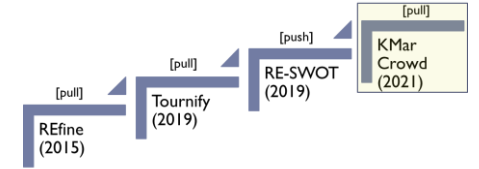


Can we gather a larger crowd?

Measurement	Menkveld	REfine	GARUSO
Duration in days	35	35	92
Participants:			
- Invited	unk.	37	unk.
- Accessed	157	19	726
- Active	39	19	32
Ideas	57	21	56
Logins	247	unk.	unk.
Votes	89	130	160
Comments	14	37	unk.
Ideas / Accessed	0.36	1.11	0.08



KMar Crowd – Premises



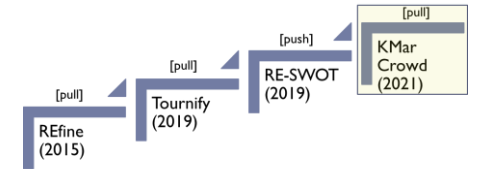
Measurement	Menkveld	REfine	GARUSO
Duration in days	35	35	92
Participants:			
- Invited	unk.	37	unk.
- Accessed	157	19	726
- Active	39	19	32
Ideas	57	21	56
Logins	247	unk.	unk.
Votes	89	130	160
Comments	14	37	unk.
Ideas / Accessed	0.36	1.11	0.08



Can we gather a larger crowd?

Are the ideas from CrowdRE useful?

KMar Crowd – Premises



Measurement	Menkveld	REfine	GARUSO
Duration in days	35	35	92
Participants:			
- Invited	unk.	37	unk.
- Accessed	157	19	726
- Active	39	19	32
Ideas	57	21	56
Logins	247	unk.	unk.
Votes	89	130	160
Comments	14	37	unk.
Ideas / Accessed	0.36	1.11	0.08



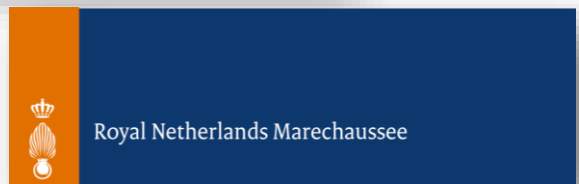
Can we gather a larger crowd?

Are the ideas from CrowdRE useful?

Can we engage the *right* crowd members?

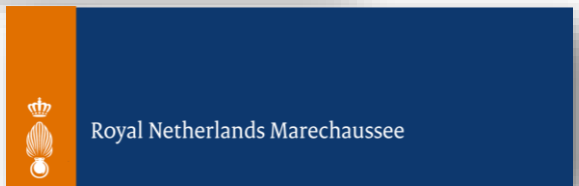
KMar Crowd – Case Studies

- ▶ Research conducted at the Royal Netherlands Marechaussee
 - ▶ 7,000 employees, split across 24 brigades
 - ▶ Canonical action research

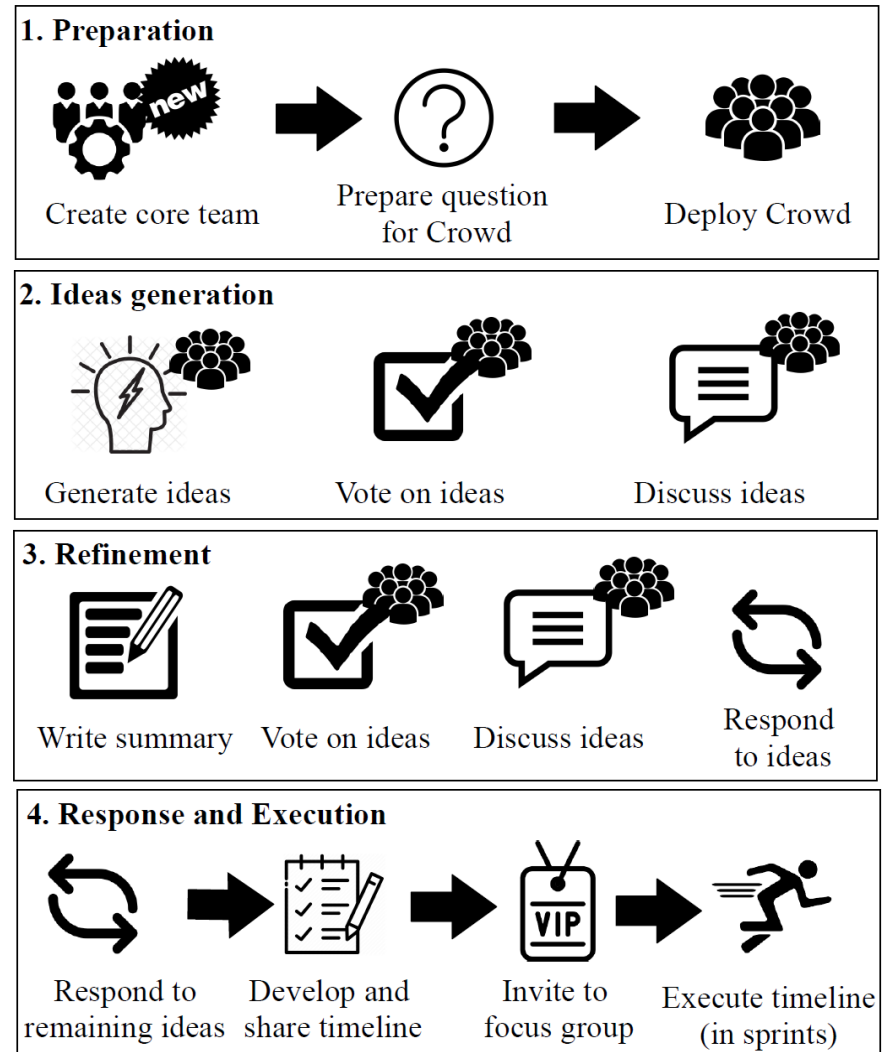


KMar Crowd – Case Studies

- ▶ Research conducted at the Royal Netherlands Marechaussee
 - ▶ 7,000 employees, split across 24 brigades
 - ▶ Canonical action research
- ▶ Three research cycles
 1. Pilot to test the Kmar-Crowd platform
 2. S-Sys: 1 brigade, **comparison to existing requirements**
 3. V-Sys: **13 brigades**, no prior requirements

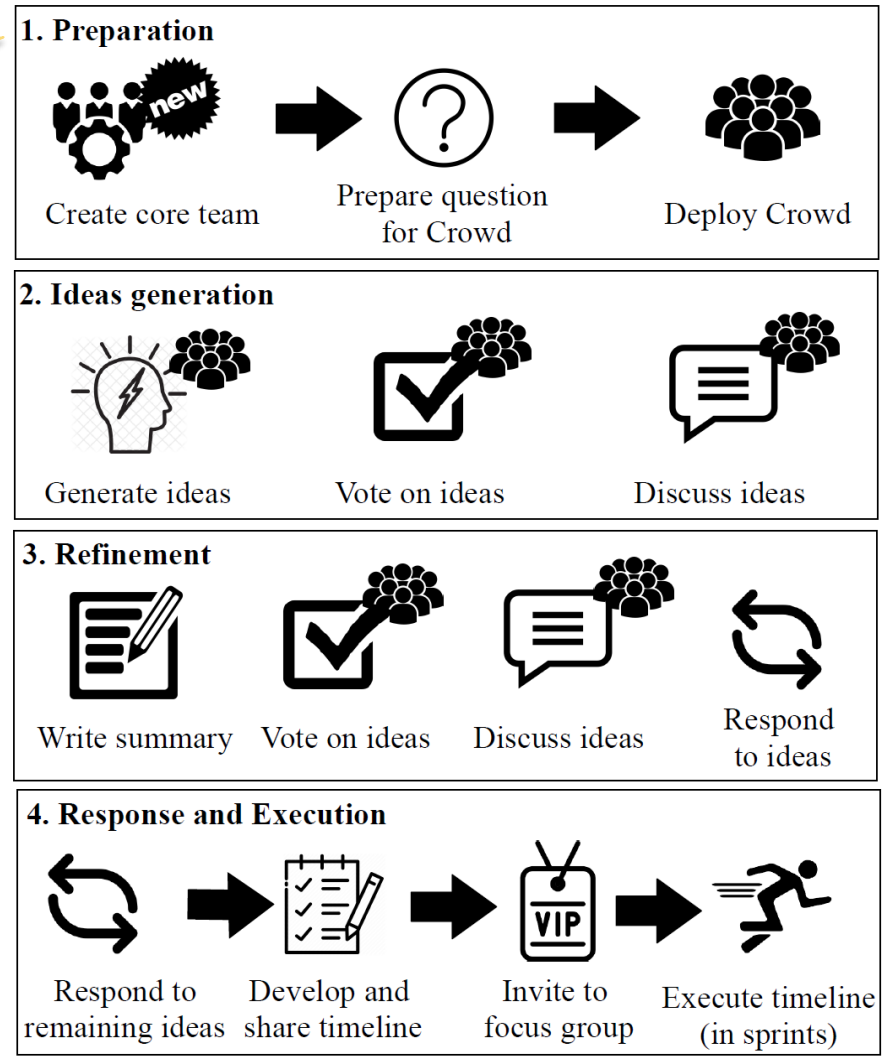


KMar Crowd – **CrowdRE** method



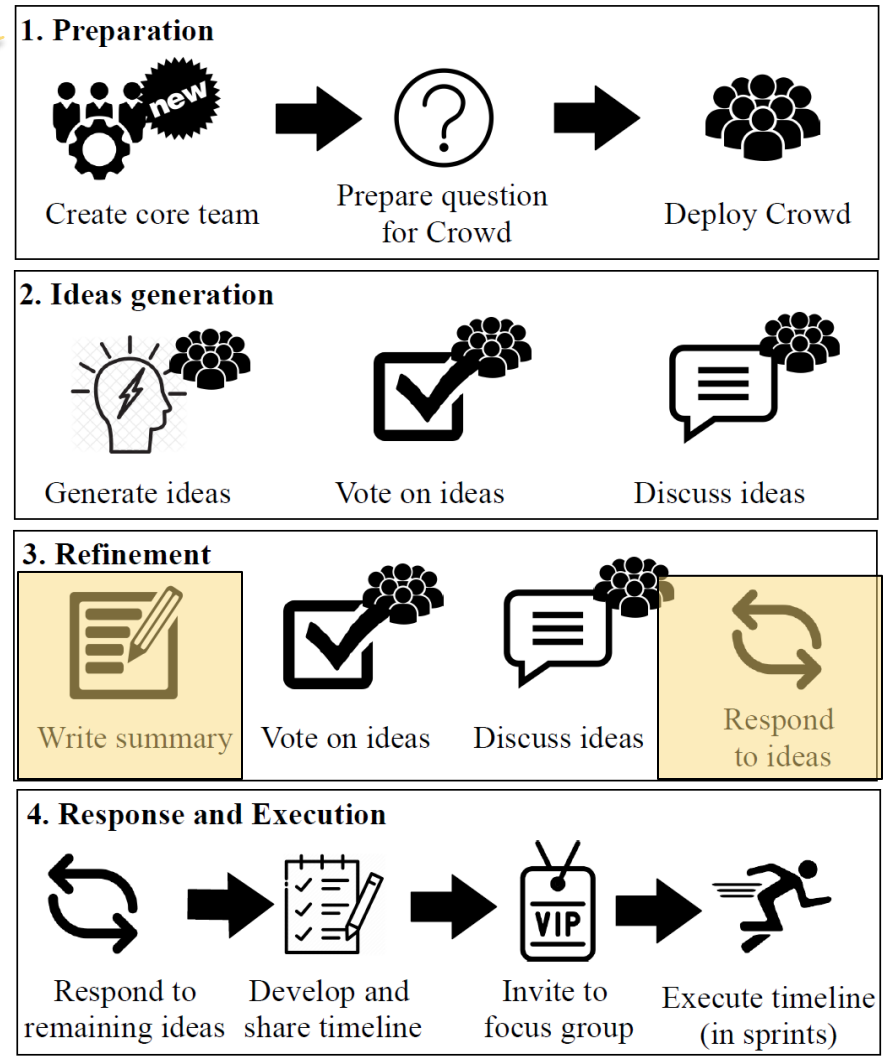
KMar Crowd – **CrowdRE** method

Extends the ideas of REfine



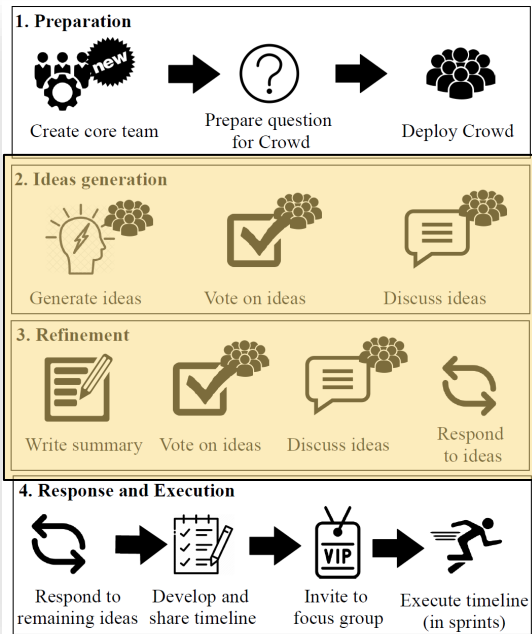
KMar Crowd – **CrowdRE** method

Extends the ideas of REfine

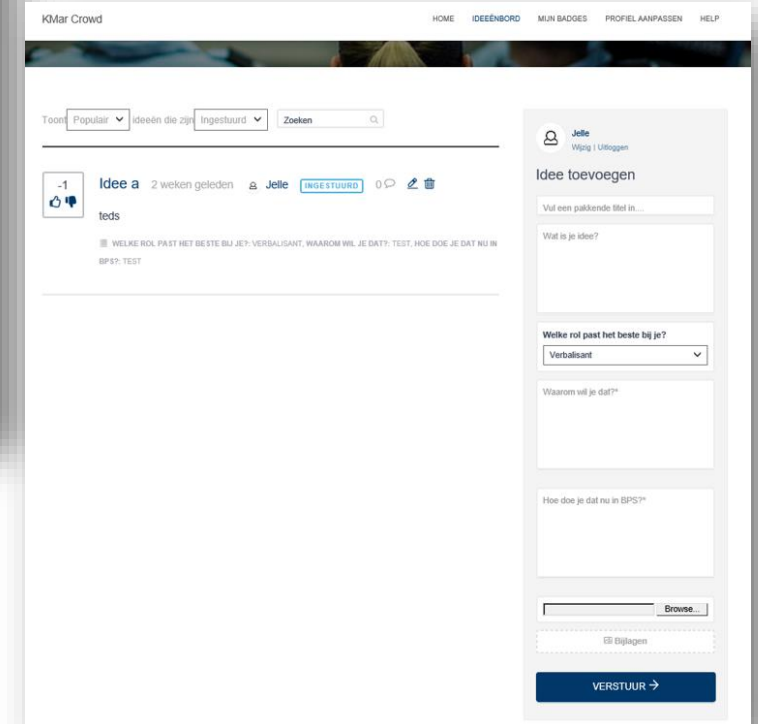
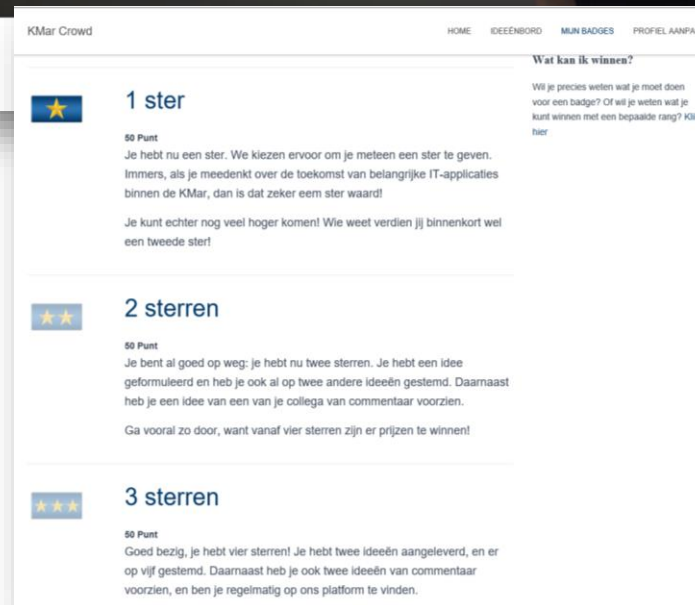
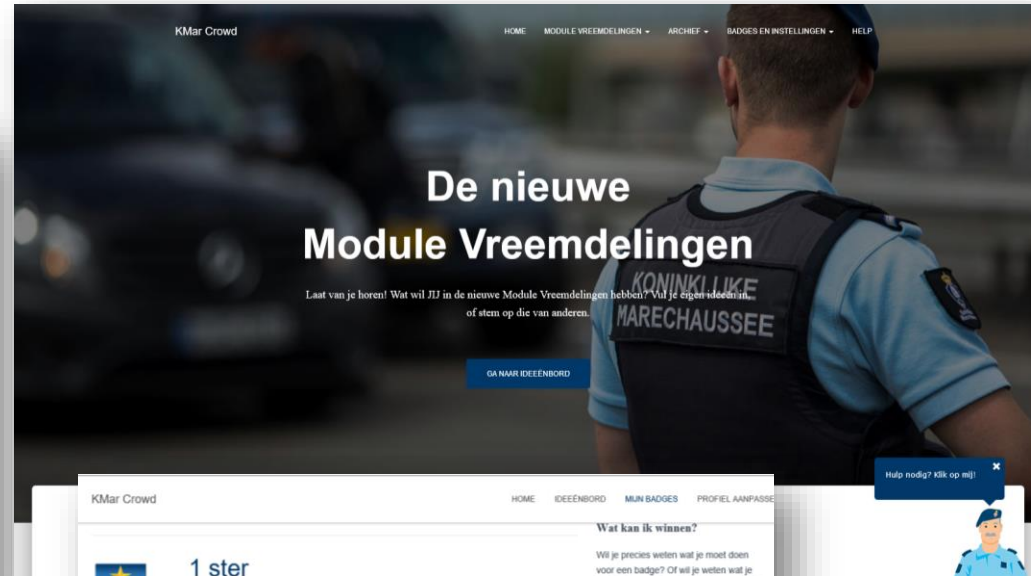
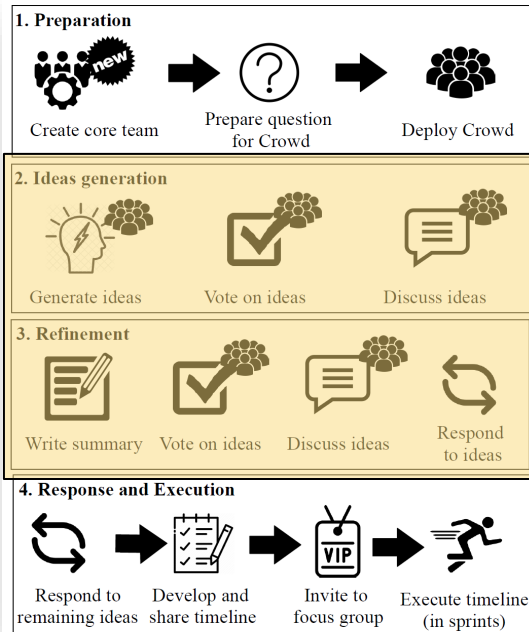


The moderator's role is crucial

KMar Crowd – **CrowdRE** platform



KMar Crowd – CrowdRE platform



KMar Crowd – Results – largeness



Measurement	S-Sys	V-Sys	Menkveld	REfine	GARUSO
Duration in days	33	56	35	35	92
Participants:					
- Invited	478	2,393	unk.	37	unk.
- Accessed	135	385	157	19	726
- Active	60	130	39	19	32
Ideas	32	78	57	21	56
Logins	240	623	247	unk.	unk.
Votes	316	531	89	130	160
Comments	28	78	14	37	unk.
Ideas / Accessed	0.24	0.20	0.36	1.11	0.08

KMar Crowd – Results – largeness



Measurement	S-Sys	V-Sys	Menkveld	REfine	GARUSO
Duration in days	33	56	35	35	92
Participants:					
- Invited	478	2,393	unk.	37	unk.
- Accessed	135	385	157	19	726
- Active	60	130	39	19	32
Ideas	32	78	57	21	56
Logins	240	623	247	unk.	unk.
Votes	316	531	89	130	160
Comments	28	78	14	37	unk.
Ideas / Accessed	0.24	0.20	0.36	1.11	0.08

KMar Crowd – Results – largeness



Measurement	S-Sys	V-Sys	Menkveld	REfine	GARUSO
Duration in days	33	56	35	35	92
Participants:					
- Invited	478	2,393	unk.	37	unk.
- Accessed	135	385	157	19	726
- Active	60	130	39	19	32
Ideas	32	78	57	21	56
Logins	240	623	247	unk.	unk.
Votes	316	531	89	130	160
Comments	28	78	14	37	unk.
Ideas / Accessed	0.24	0.20	0.36	1.11	0.08



KMar Crowd – Results – origin



Can we engage the *right* crowd members?

TABLE IV
ACTIVITY PER USER TYPE IN THE S-SYS CASE STUDY (N=135).

Origin	% of total	Per user activity		
		Ideas	Votes	Logins
Operational employee	58.52%	0.23	2.89	1.84
Middle management	8.15%	0.82	4.00	2.55
Non-targeted employee	33.34%	0.11	0.88	1.55

TABLE VII
ACTIVITY PER USER RANK IN THE V-SYS CASE STUDY (N=385).

Origin	% of total	Per user activity		
		Ideas	Votes	Logins
Soldiers	0.00%	-	-	-
Corporals	3.64%	0.36	3.43	2.43
NCOs	76.62%	0.22	1.41	1.63
Sub-altern officers	11.43%	0.10	1.45	1.68
Head officers	2.34%	0.00	0.00	1.22
Citizens	5.97%	0.09	0.13	1.30

KMar Crowd – Results – origin



Can we engage the *right* crowd members?

TABLE IV
ACTIVITY PER USER TYPE IN THE S-SYS CASE STUDY (N=135).

Origin	% of total	Per user activity		
		Ideas	Votes	Logins
Operational employee	58.52%	0.23	2.89	1.84
Middle management	8.15%	0.82	4.00	2.55
Non-targeted employee	33.34%	0.11	0.88	1.55

TABLE VII
ACTIVITY PER USER RANK IN THE V-SYS CASE STUDY (N=385).

Origin	% of total	Per user activity		
		Ideas	Votes	Logins
Soldiers	0.00%	-	-	-
Corporals	3.64%	0.36	3.43	2.43
NCOs	76.62%	0.22	1.41	1.63
Sub-altern officers	11.43%	0.10	1.45	1.68
Head officers	2.34%	0.00	0.00	1.22
Citizens	5.97%	0.09	0.13	1.30

KMar Crowd – Results – origin



Can we engage the *right* crowd members?



Operational employees were reached

TABLE IV
ACTIVITY PER USER TYPE IN THE S-SYS CASE STUDY (N=135).

Origin	% of total	Per user activity		
		Ideas	Votes	Logins
Operational employee	58.52%	0.23	2.89	1.84
Middle management	8.15%	0.82	4.00	2.55
Non-targeted employee	33.34%	0.11	0.88	1.55

TABLE VII
ACTIVITY PER USER RANK IN THE V-SYS CASE STUDY (N=385).

Origin	% of total	Per user activity		
		Ideas	Votes	Logins
Soldiers	0.00%	-	-	-
Corporals	3.64%	0.36	3.43	2.43
Non-commissioned officers	76.62%	0.22	1.41	1.63
Commissioned officers	11.43%	0.10	1.45	1.68
Warrant officers	2.34%	0.00	0.00	1.22
Other ranks	5.97%	0.09	0.13	1.30

KMar Crowd – Results – usefulness



Are the ideas from CrowdRE useful?

USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19

V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
Enough for MVP		59.5
Enough for product		27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

KMar Crowd – Results – usefulness



Are the ideas from CrowdRE useful?

USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19

V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
Enough for MVP		59.5
Enough for product		27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

KMar Crowd – Results – usefulness



Are the ideas from CrowdRE useful?

USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19

V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
Enough for MVP		59.5
Enough for product		27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

KMar Crowd – Results – usefulness



Are the ideas from CrowdRE useful?

USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19

V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
Enough for MVP		59.5
Enough for product		27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

KMar Crowd – Results – usefulness



Are the ideas from CrowdRE useful?

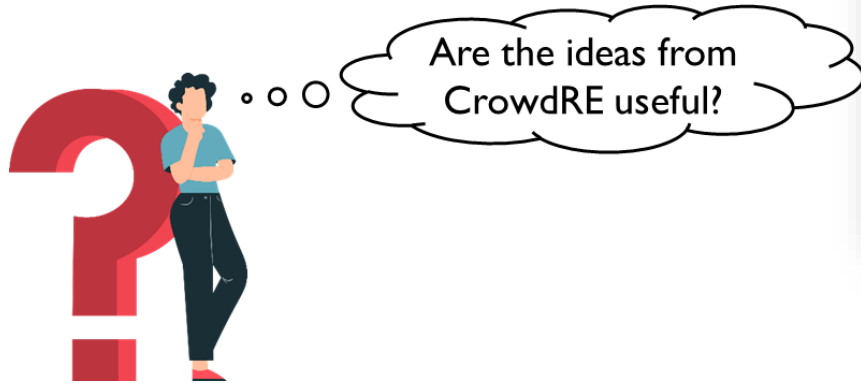
USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19

V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
	Enough for MVP	59.5
	Enough for product	27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

KMar Crowd – Results – usefulness



USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

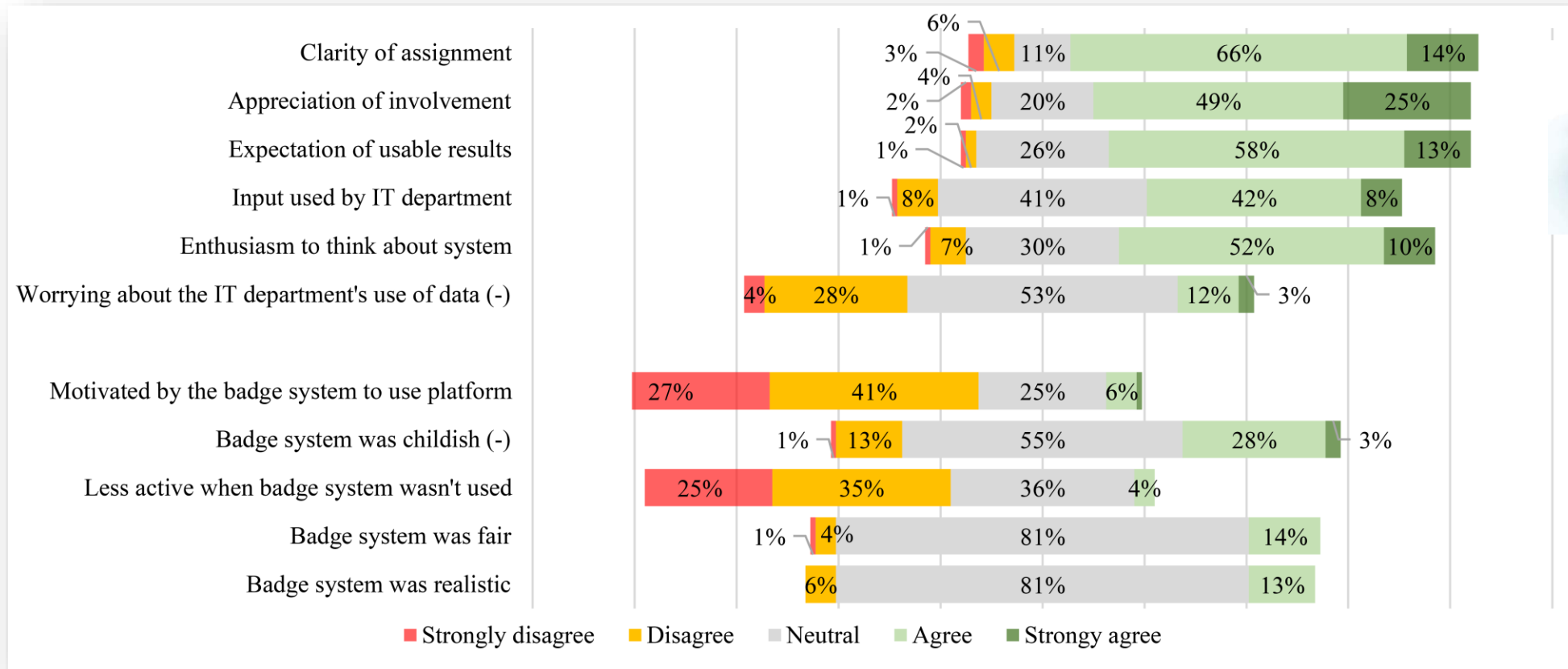
Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19



V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

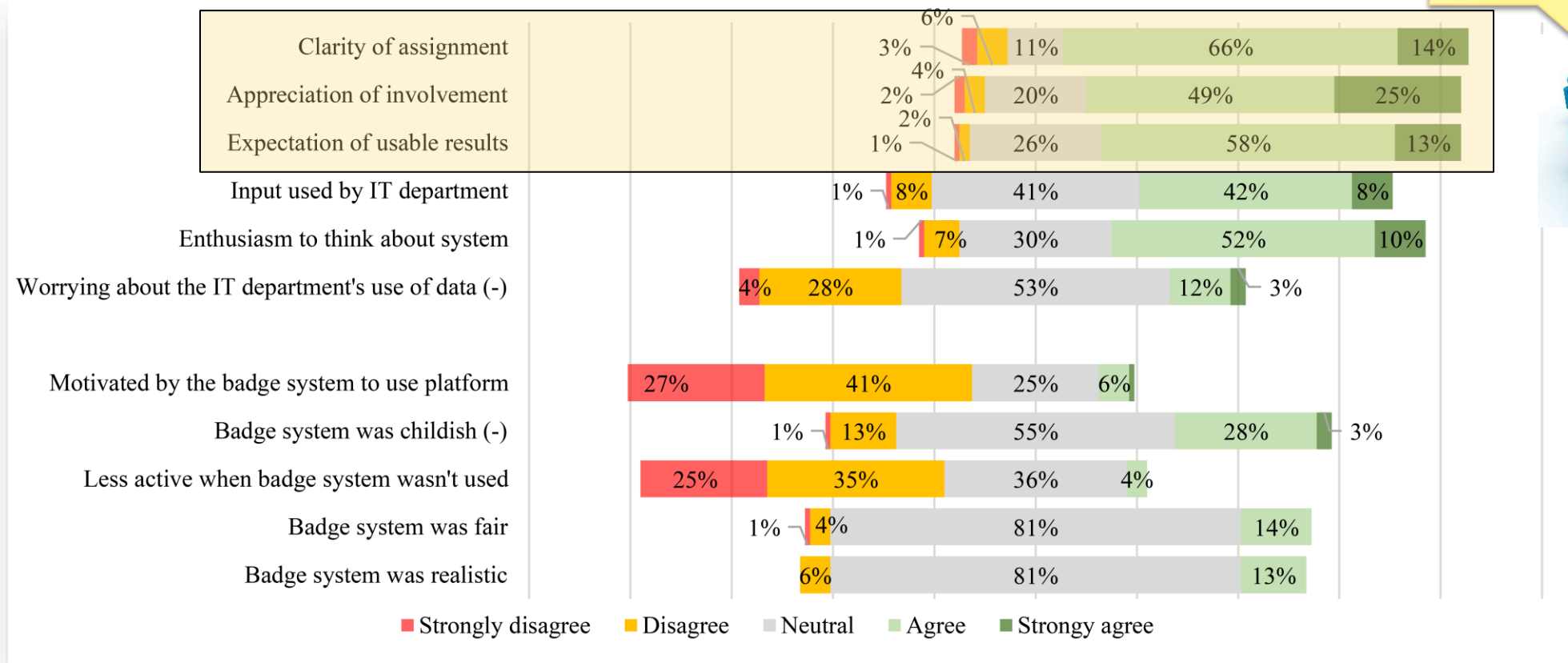
Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
	Enough for MVP	59.5
	Enough for product	27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

KMar Crowd – Results – appreciation

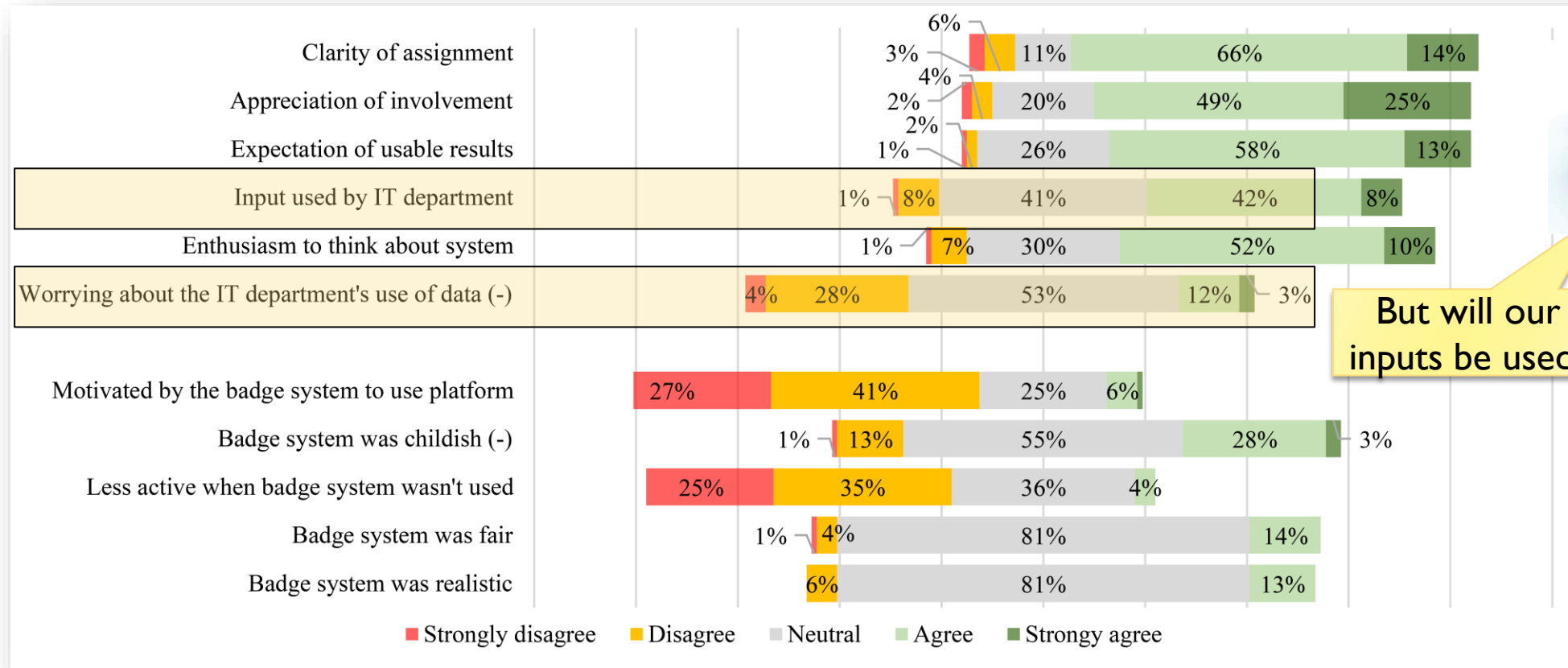


KMar Crowd – Results – appreciation

Positively involved!

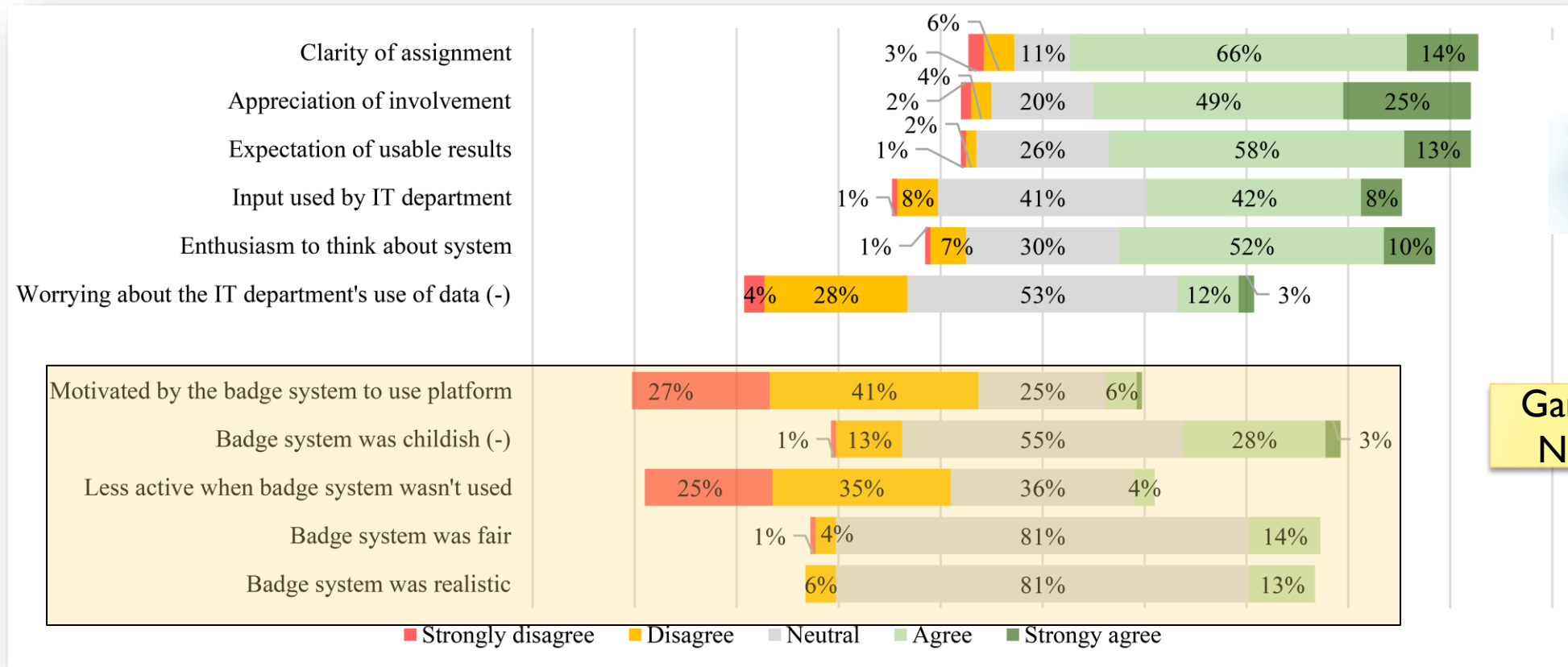


KMar Crowd – Results – appreciation



But will our inputs be used?

KMar Crowd – Results – appreciation



**Gamification?
Not for us!**

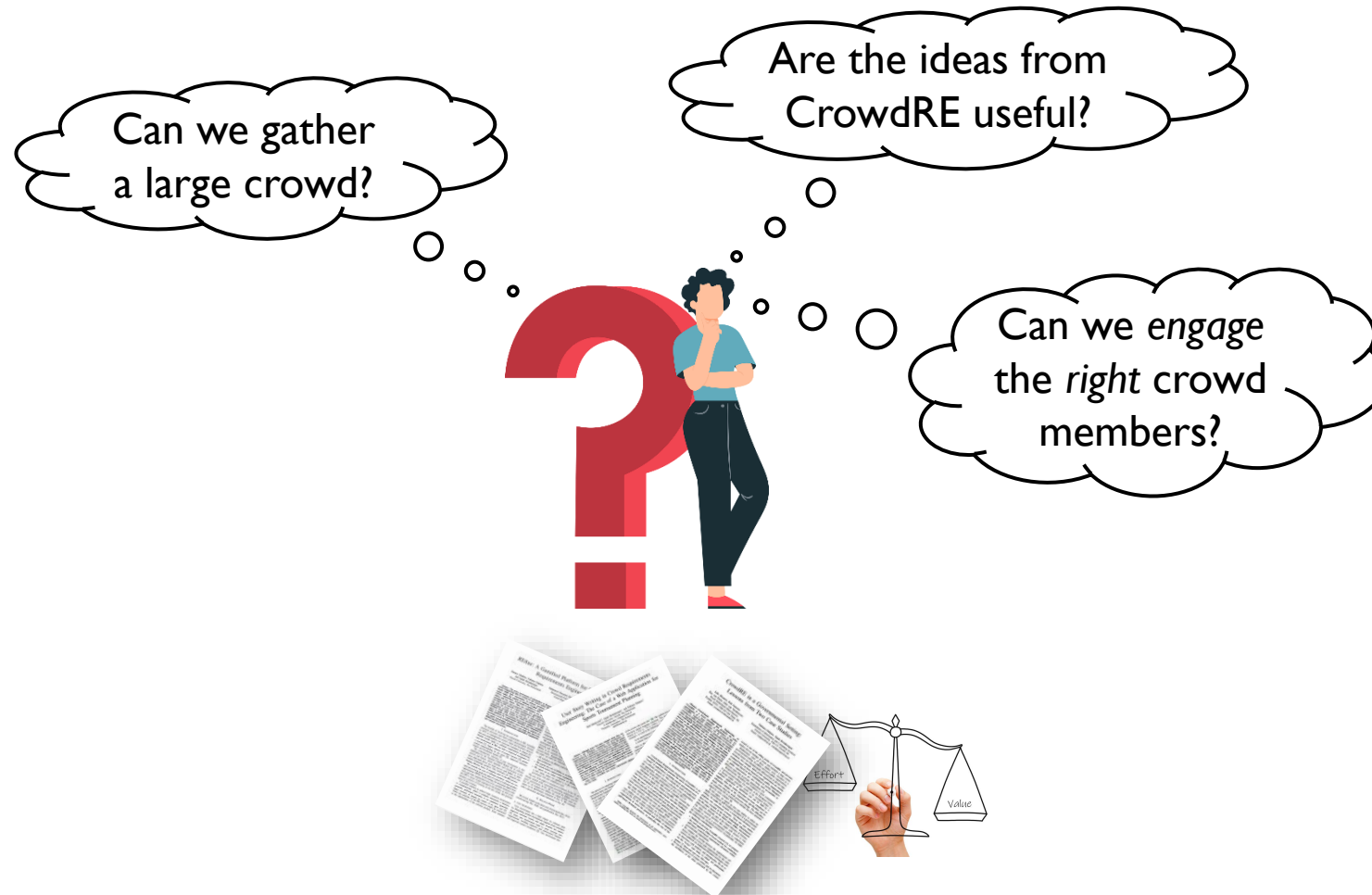
4. Implications and Outlook



Implications for Practice – **pull** feedback



Implications for Practice – **pull** feedback



Pull feedback – **crowd largeness**

Measurement	S-Sys	V-Sys	Menkveld	REfine	GARUSO
Duration in days	33	56	35	35	92
Participants:					
- Invited	478	2,393	unk.	37	unk.
- Accessed	135	385	157	19	726
- Active	60	130	39	19	32

Pull feedback – **crowd largeness**

Measurement	S-Sys	V-Sys	Menkveld	REfine	GARUSO
Duration in days	33	56	35	35	92
Participants:					
- Invited	478	2,393	unk.	37	unk.
- Accessed	135	385	157	19	726
- Active	60	130	39	19	32

Attracting a large crowd is difficult

Pull feedback – **crowd largeness**

Do we even know how to measure a crowd's size ?

Measurement	S-Sys	V-Sys	Menkveld	REfine	GARUSO
Duration in days	33	56	35	35	92
Participants:					
- Invited	478	2,393	unk.	37	unk.
- Accessed	135	385	157	19	726
- Active	60	130	39	19	32

Attracting a large crowd is difficult

Pull feedback – **crowd largeness**

Do we even know how to measure a crowd's size ?

Measurement	S-Sys	V-Sys	Menkveld	REfine	GARUSO
Duration in days	33	56	35	35	92
Participants:					
- Invited	478	2,393	unk.	37	unk.
- Accessed	135	385	157	19	726
- Active	60	130	39	19	32



↑people → ↑effort
...but...
↑people → ↑value ???

Attracting a large crowd is difficult

Implications for Practice – **ideas' usefulness**

USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19

V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
Enough for MVP		59.5
Enough for product		27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

Implications for Practice – **ideas' usefulness**

USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19

Useful as **novel**:
Difficult to validate, we need existing specifications

V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
Enough for MVP		59.5
Enough for product		27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

Implications for Practice – **ideas' usefulness**

USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19

V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
Enough for MVP		59.5
Enough for product		27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

Useful as **delighters**:
some ideas are likely to be
attractive requirements

Implications for Practice – **ideas' usefulness**

USEFULNESS OF THE IDEAS IN THE S-SYS CASE STUDY, ASSESSED BY THE ANALYSTS WHO CONDUCTED ELICITATION WITHOUT *KMar-Crowd*.

Measurement	Value	# Ideas
KANO model	Must-be	13
	One-dimensional	10
	Attractive	7
Gathered earlier	Completely	19
	Partly	6
	Not at all	5
Complete for dev teams	Yes	11
	No	19

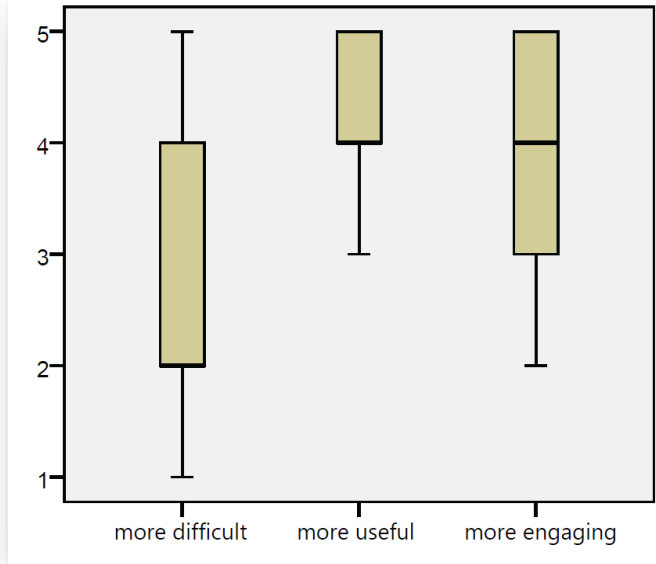
V-SYS: USEFULNESS OF THE IDEAS, ASSESSED BY A POOL OF ANALYSTS.

Measurement	Value	% Ideas
KANO model	Must-be	50.6
	One-dimensional	36.7
	Attractive	12.7
Enough for MVP		59.5
Enough for product		27.8
Granularity	Epic	40.5
	User Story	54.4
	Not applicable	5.1

enables developing the right software [5]. Furthermore, we have found that 95% of the crowdsourced USs are both easy to estimate and easily implementable based on our quality analysis and hour estimation as done by the main developer. Almost 90% of the feature requests can even be implemented within one workday.

Useful as **specific for dev teams**: many ideas will be, not all

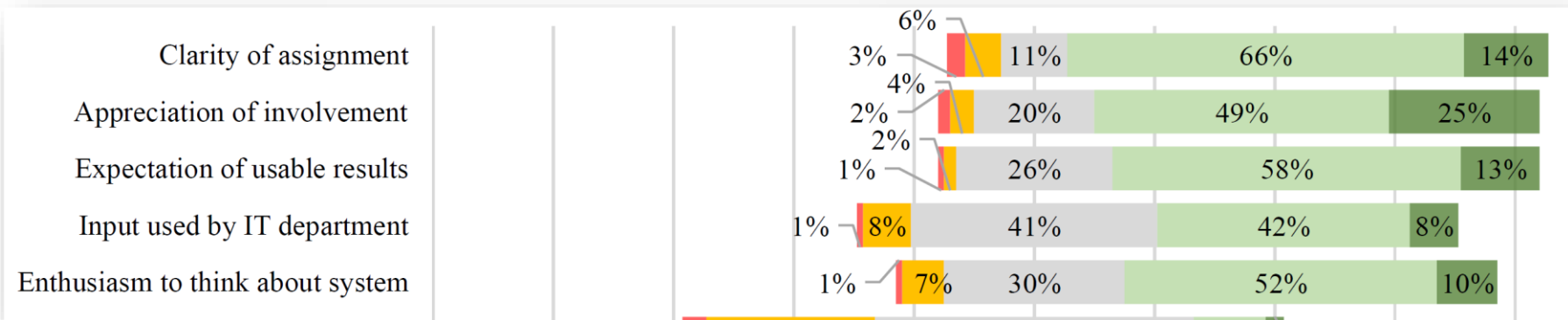
Implications for Practice – **user engagement**



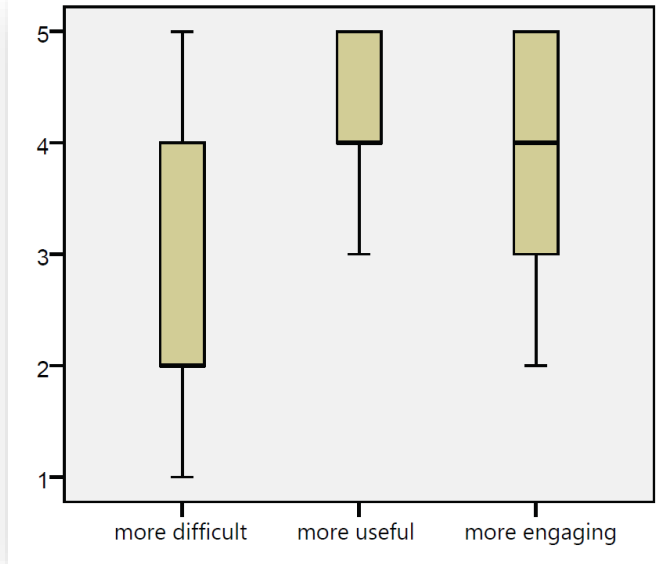
a feature themselves, while the other three respondents only voted for a feature. They perceived the platform as very useful, regarding all four possible interactions when rated on a five-point Likert scale: requesting (M = 4.9; SD = 0.28), viewing (M = 4.8; SD = 0.38), voting (M = 4.5; SD = 0.88), and commenting (M = 4.5; SD = 0.66). One user who requested a feature, voted for and commented on an idea and had previous experience in writing USs commented:

“You implemented the agile methodology in a very fun way. In such a manner the users get involved better and at least have the feeling their opinion matters”

Others found it “a fantastic way to improve the application” “very useful to allow users to submit ideas” and see it



Implications for Practice – **user engagement**

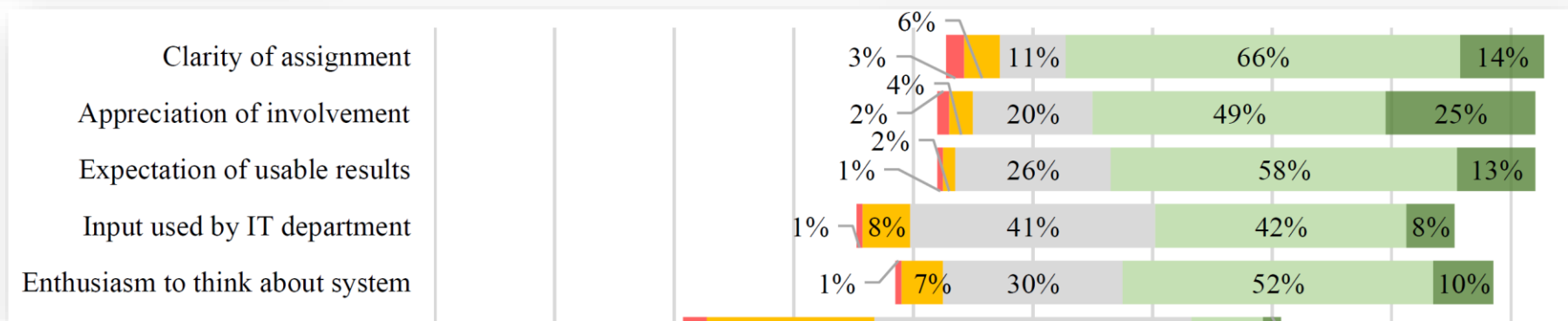


Perceived engagement is generally high

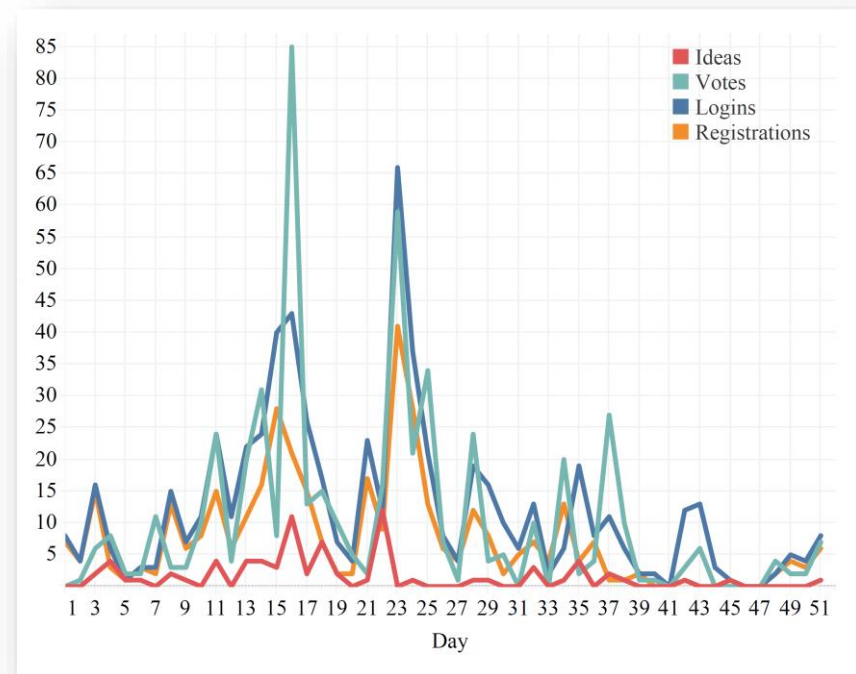
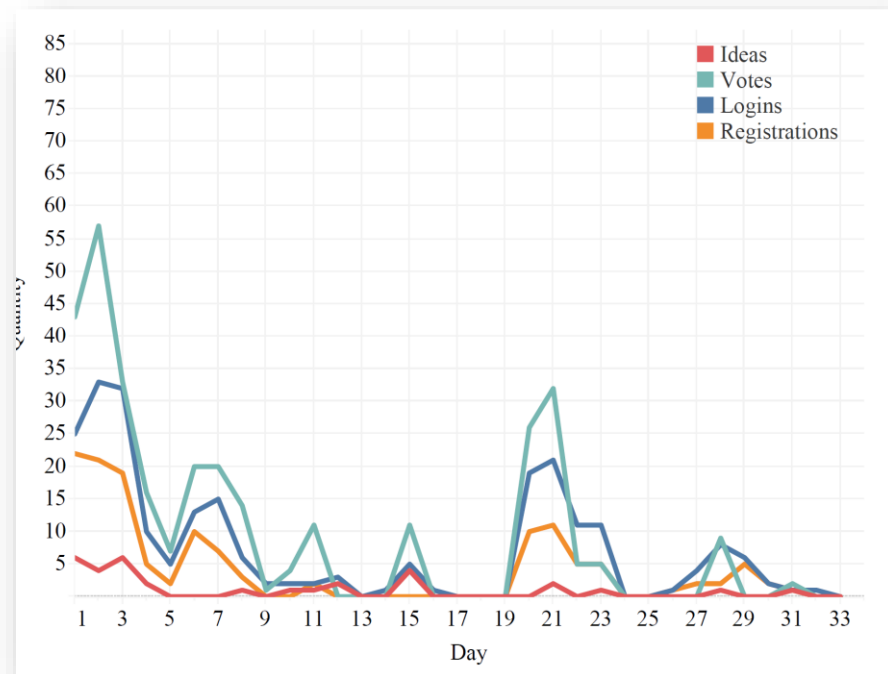
a feature themselves, while the other three respondents only voted for a feature. They perceived the platform as very useful, regarding all four possible interactions when rated on a five-point Likert scale: requesting (M = 4.9; SD = 0.28), viewing (M = 4.8; SD = 0.38), voting (M = 4.5; SD = 0.88), and commenting (M = 4.5; SD = 0.66). One user who requested a feature, voted for and commented on an idea and had previous experience in writing USs commented:

“You implemented the agile methodology in a very fun way. In such a manner the users get involved better and at least have the feeling their opinion matters”

Others found it “a fantastic way to improve the application” “very useful to allow users to submit ideas” and see it

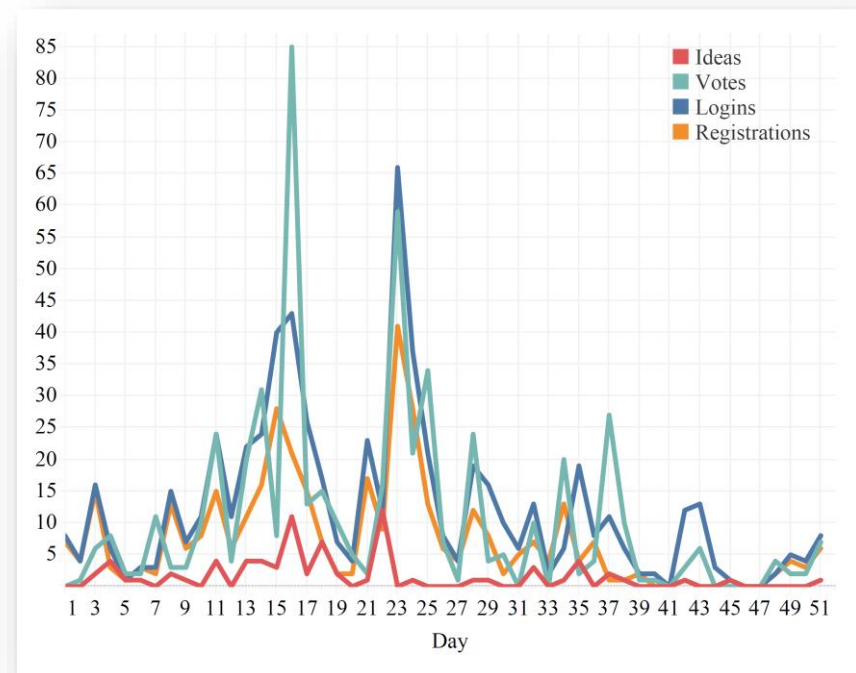
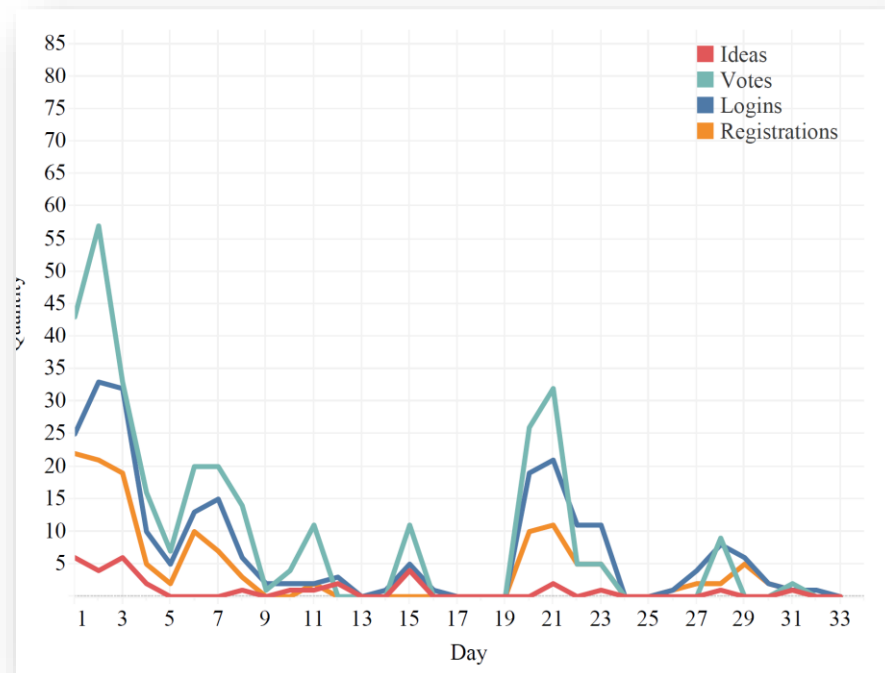


Implications for Practice – **user engagement**



Implications for Practice – **user engagement**

But engagement is not self-sustained



Implications for Practice – **user engagement**

Measurement	S-Sys	V-Sys	Menkveld	REfine	GARUSO
Duration in days	33	56	35	35	92
Participants:					
- Invited	478	2,393	unk.	37	unk.
- Accessed	135	385	157	19	726
- Active	60	130	39	19	32
Ideas	32	78	57	21	56
Logins	240	623	247	unk.	unk.
Votes	316	531	89	130	160
Comments	28	78	14	37	unk.
Ideas / Accessed	0.24	0.20	0.36	1.11	0.08

And expect many more users than ideas!

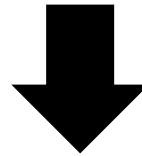
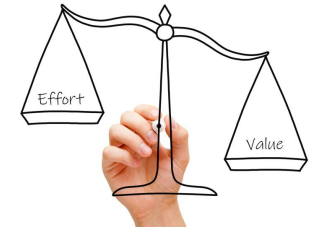
Research Question – **pull** feedback answer

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



Research Question – **pull** feedback answer

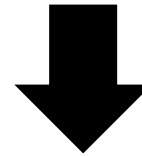
RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



- Extensive human effort is **unavoidable** with pull feedback

Research Question – **pull** feedback answer

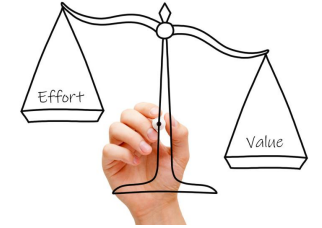
RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



- Extensive human effort is **unavoidable** with pull feedback
- Rather, focus on **increasing value**
 - Make use of the ideas
 - Update the users on how the ideas are used
 - Reward the most active participants

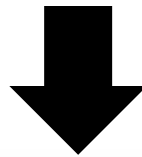
Research Question – **push** feedback answer

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



Research Question – push feedback answer

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



Automation = Value for Research

High citation figures

How do users like this feature? a fine grained sentiment analysis of app reviews	541	2014
Bug report, feature request, or simply praise? on automatically classifying app reviews	387	2015
How Can I Improve My App? Classifying User Reviews for Software Maintenance and Evolution	356	2015

Extensive number of papers

A Survey of App Store Analysis for Software Engineering
An Overview of User Feedback Classification Approaches

Standard metrics to show success

$$R = \frac{[fnd \cap rel]}{[rel]} = \frac{[TP]}{[TP] + [FN]}$$
$$P = \frac{[fnd \cap rel]}{[fnd]} = \frac{[TP]}{[FP] + [TP]} \quad (2)$$

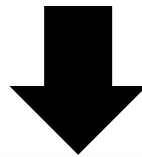
osite of recall and precision that is often called " is captured by the F -measure:

$$F = 2 \times \frac{P \times R}{P + R} \quad (3)$$

8 @2021 Fabiano Dalpiaz

Research Question – push feedback answer

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



Automation = Value for Research

High citation figures

- How do users like this feature? a fine grained sentiment analysis of app reviews. 541. 2014. E. Guzman, W. Maalej. 2014 IEEE 22nd international requirements engineering conference (RE), 153-162.
- Bug report, feature request, or simply praise? on automatically classifying app reviews. 387. 2015. W. Maalej, H. Nabil. 2015 IEEE 23rd international requirements engineering conference (RE), 116-125.
- How Can I Improve My App? Classifying User Reviews for Software Maintenance and Evolution. 356. 2015. S. Panichella, AD. Sorbo, E. Guzman, A. Visaggio, G. Canfora, H. Gall. Software Maintenance and Evolution (ICSME), 2015 IEEE International ...

Extensive number of papers

A Survey of App Store Analysis for Software Engineering
 William Martin, Federica Sarro, Yue Jia, Yuanzhan Zhang, and Mark Haman

An Overview of User Feedback Classification Approaches
 Robert Santos, Edward C. Groom, Karina Villala
 Pilsenerstr. 115SE, Kaiserslautern, Germany

Standard metrics to show success

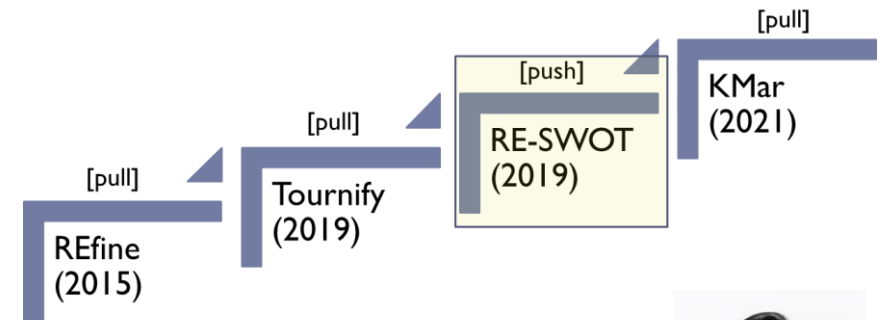
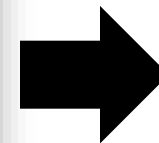
$$R = \frac{[fnd \cap rel]}{[rel]} = \frac{[TP]}{[TP] + [FN]}$$

$$P = \frac{[fnd \cap rel]}{[fnd]} = \frac{[TP]}{[FP] + [TP]}$$

osite of recall and precision that is often called " is captured by the *F*-measure:

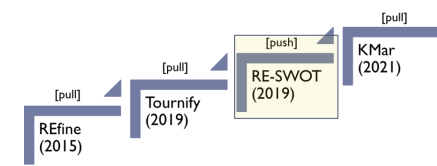
$$F = 2 \times \frac{P \times R}{P + R}$$

8 @2021 Fabiano Dalpiaz

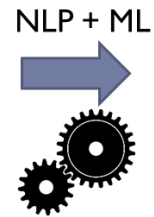
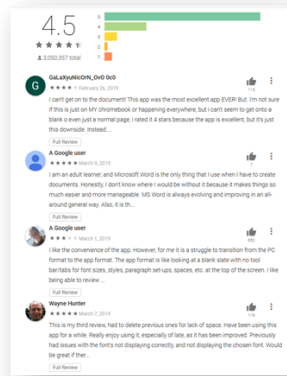


Fabiano Dalpiaz, Micaela Parente: RE-SWOT: From User Feedback to Requirements via Competitor Analysis. REFSQ 2019: 55-70





Which value in **push** feedback?



Topic	Senti. score
upload_photo, load_photo, photo_take, photo_want, upload_want, download_photo, upload_feature, itive move_photo, keep_upload, keep_try	1.51
file_name, folder_file, rename_file, file_add, Pos-folde	1.49
Sentence	Category
They just need to update the layout I fill like everything is hidden I want a better task bar.	feature request
Please restore a way to open pin links in external browser or let us save photos.	feature request
App crashes when new power up notice pops up.	problem discovery
Please fix the syncing issues with the iPad app.	problem discovery
It's already possible to rearrange boards why not the pins on a single board?	Information seeking
Overall it is fun and provides a lot of good info.	Information giving
This app runs so smoothly and I rarely have issues with it anymore.	Information giving

Standard metrics to show success

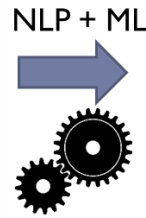
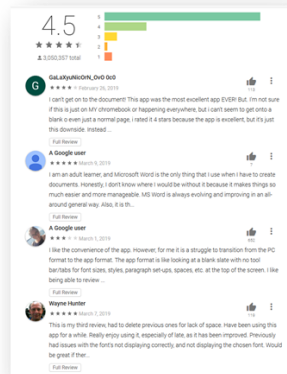
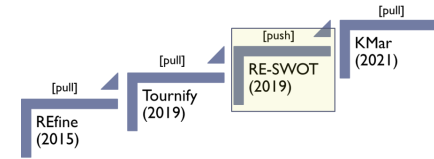
$$R = \frac{|\text{fnd} \cap \text{rel}|}{|\text{rel}|} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FN}|}$$

$$P = \frac{|\text{fnd} \cap \text{rel}|}{|\text{fnd}|} = \frac{|\text{TP}|}{|\text{FP}| + |\text{TP}|} \quad (2)$$

osite of recall and precision that is often called ' is captured by the *F*-measure:

$$F = 2 \times \frac{P \times R}{P + R} \quad (3)$$


Which value in **push** feedback?



Topic	Senti. score
upload_photo, load_photo, photo_take, photo_want, upload_want, download_photo, upload_feature, iiive	1.51 Pos-
move_photo, keep_upload, keep_try	
file_name, folder_file, rename_file, file_add, 1.49 Pos-	
Sentence	Category
file_name	feature request
chan	feature request
file_	feature request
down	feature request
file_	feature request
App crashes when new power up notice pops up.	problem discovery
Please fix the syncing issues with the iPad app.	problem discovery
It's already possible to rearrange boards why not the pins on a single board?	Information seeking
Overall it is fun and provides a lot of good info.	Information giving
This app runs so smoothly and I rarely have issues with it anymore.	Information giving

Standard metrics to show success

$$R = \frac{|\text{fnd} \cap \text{rel}|}{|\text{rel}|} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FN}|}$$

$$P = \frac{|\text{fnd} \cap \text{rel}|}{|\text{fnd}|} = \frac{|\text{TP}|}{|\text{FP}| + |\text{TP}|} \quad (2)$$

osite of recall and precision that is often called ' is captured by the F -measure:

$$F = 2 \times \frac{P \times R}{P + R} \quad (3)$$



Pillars of RE-SWOT



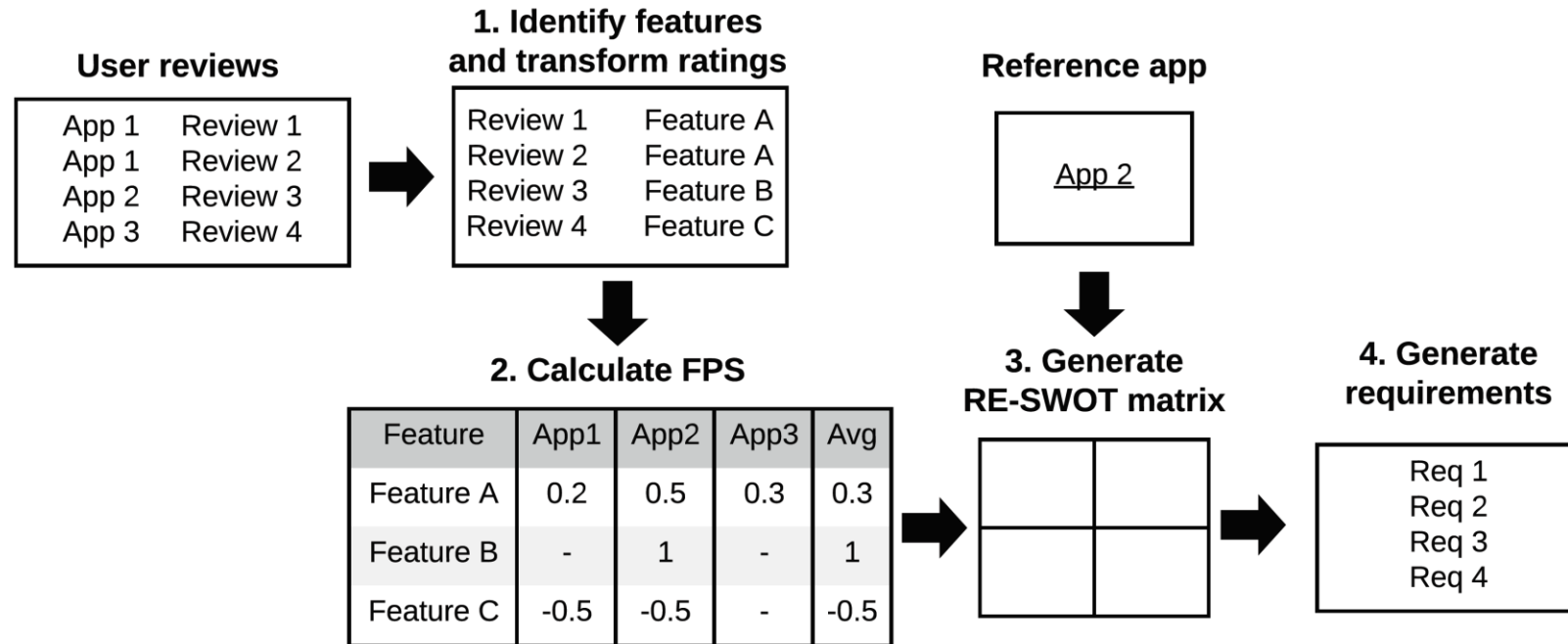
- ▶ Practitioner-focused setting
 - ▶ A **product manager** wants to analyze their **app's reviews**
 - ▶ Idea: analyze with respect to **competitors**
 - ▶ Through an **automatically created dashboard**

Pillars of RE-SWOT

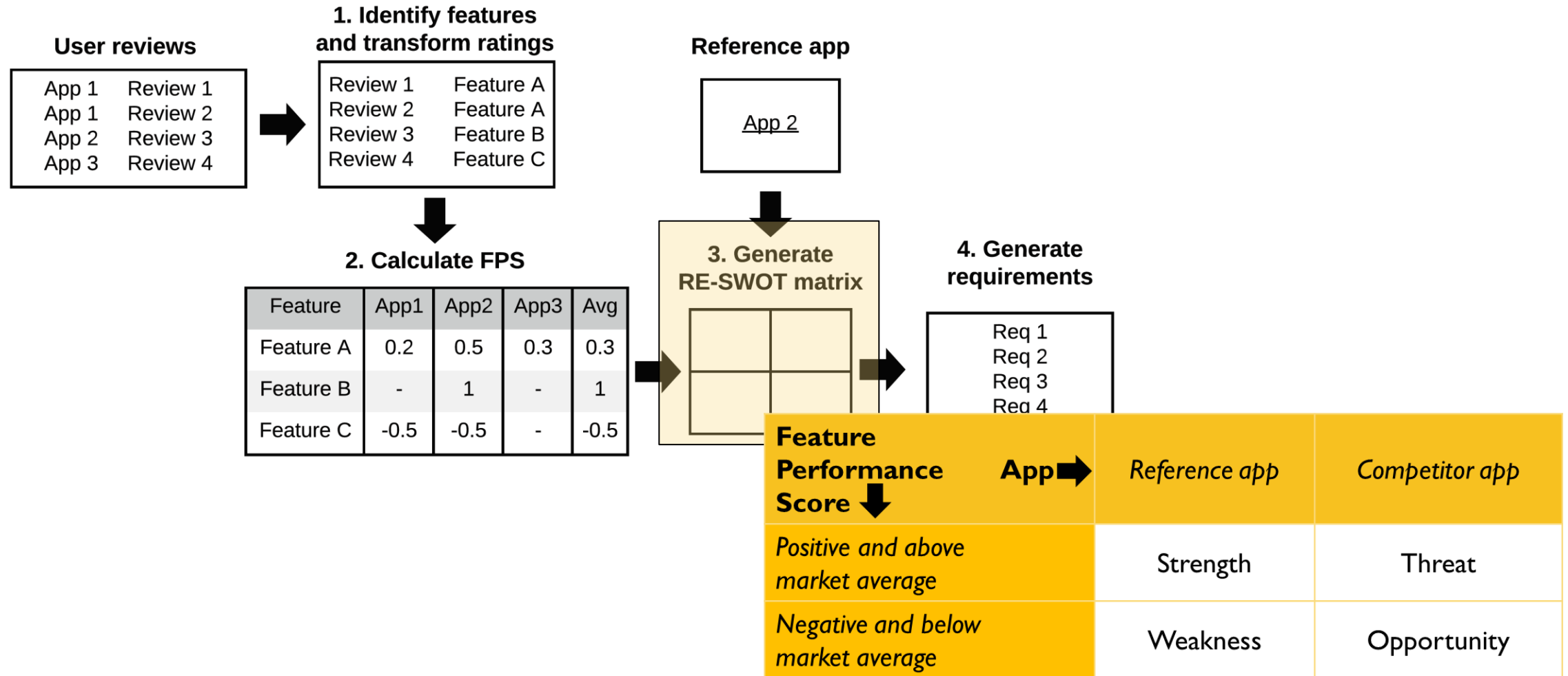


- ▶ Practitioner-focused setting
 - ▶ A **product manager** wants to analyze their **app's reviews**
 - ▶ Idea: analyze with respect to **competitors**
 - ▶ Through an **automatically created dashboard**
- ▶ In-vivo validation
 - ▶ **Ignore precision, recall, F1**
 - ▶ Consider product manager's inputs

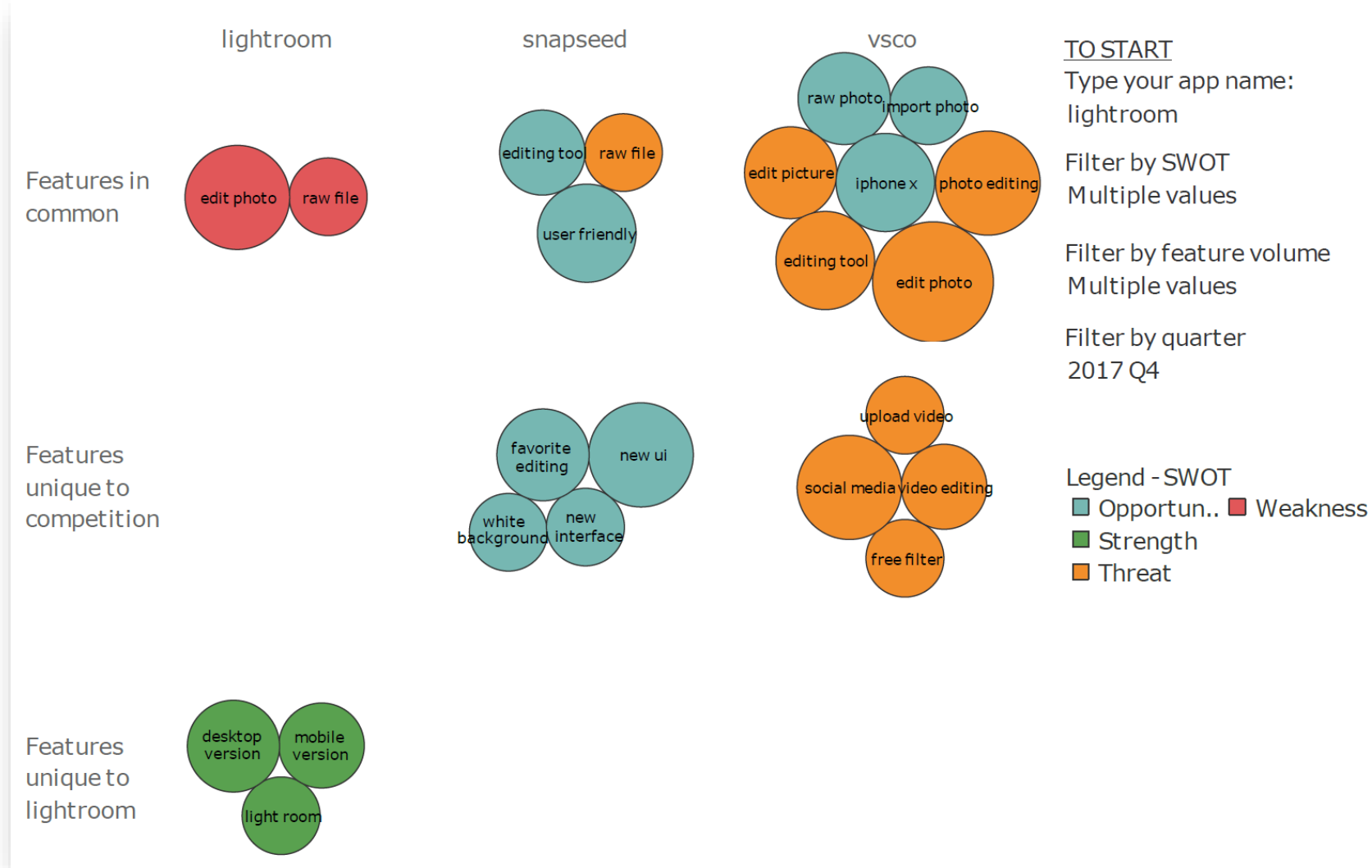
RE-SWOT illustrated



RE-SWOT illustrated



The RE-SWOT Matrix, in practice



Three case studies

▶ **Dating apps**

- ▶ Canadian company with 150M users
- ▶ Competitors: market leader, fast growing company

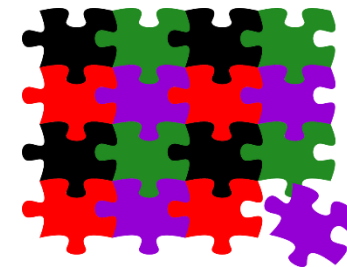


▶ **Travel apps for tourists**

- ▶ 400 employees company
- ▶ Competitors: app for Asia, market leader

▶ **Puzzle games**

- ▶ App with 135 puzzle types, small company
- ▶ Competitors: similar apps



Three case studies

▶ Dating apps

- ▶ Canadian company with 150M users
- ▶ Competitors: market leader, fast growing company



Interviews with product managers

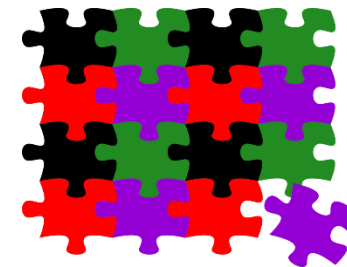
1. Introduction
2. Your current practices
3. Demo
4. Tool use
5. Questions

sts

asia, market leader

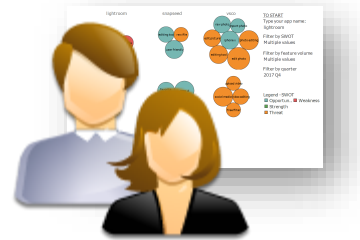
▶ Puzzle games

- ▶ App with 135 puzzle types, small company
- ▶ Competitors: similar apps



Main results from RE-SWOT

- + Visual/interactive aspect
- + Allows to handle many reviews
- + Some competitors' features could be found

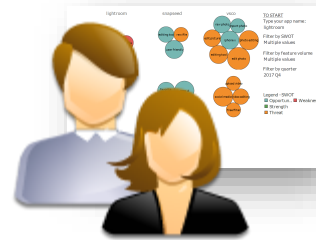


Product managers



Main results from RE-SWOT

- + Visual/interactive aspect
- + Allows to handle many reviews
- + Some competitors' features could be found



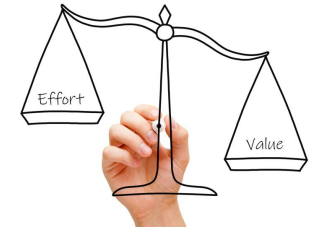
Product managers

- Feature extraction is inaccurate
- SWOT classification too
- We need trend analysis



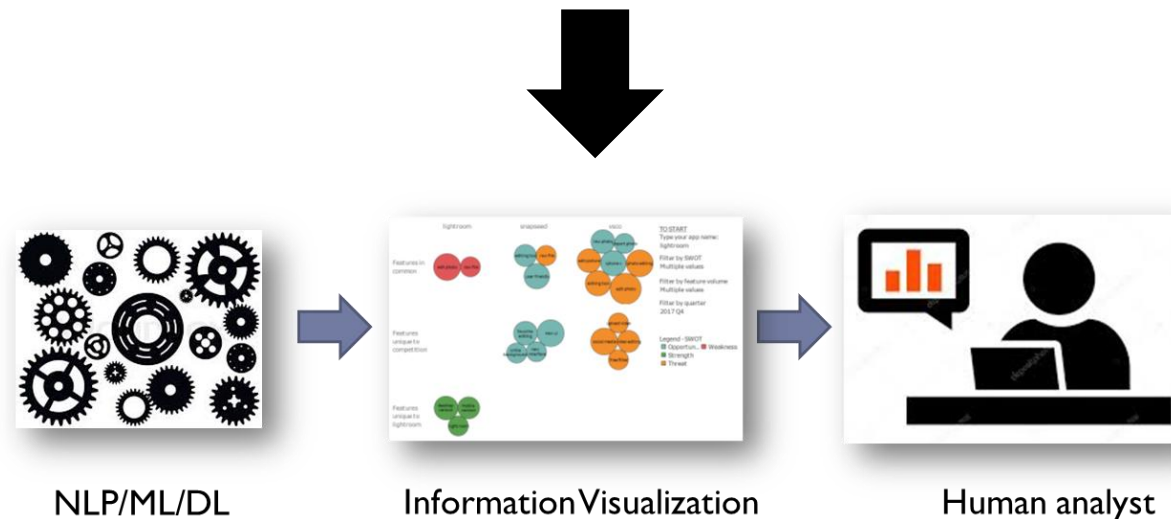
Research Question – **push** feedback answer

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



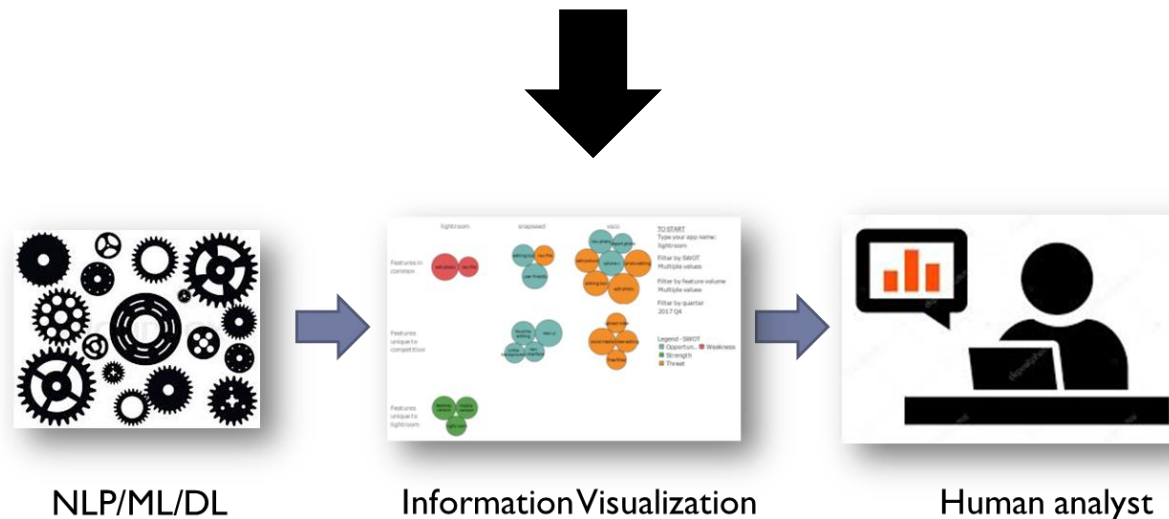
Research Question – **push** feedback answer

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



Research Question – push feedback answer

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



Requirements Eng (2014) 19:257–279
DOI 10.1007/s00766-013-0194-3

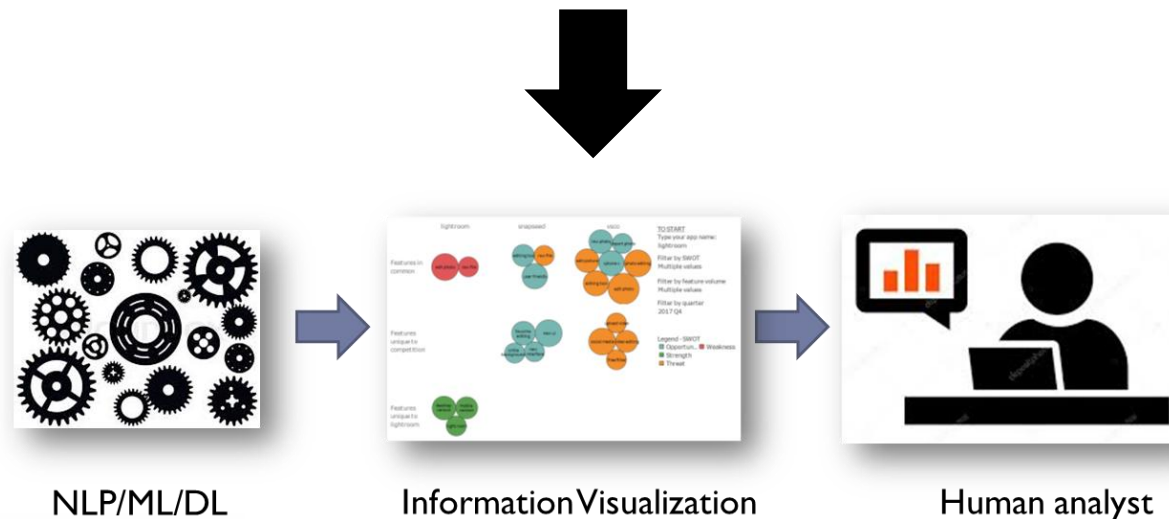
RE 2013

Visual requirements analytics: a framework and case study

Sandeep Reddivari · Shirin Rad · Tanmay Bhowmik ·
Nisreen Cain · Nan Niu

Research Question – push feedback answer

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



Requirements Eng (2014) 19:257–279
DOI 10.1007/s00766-013-0194-3

RE 2013

Visual requirements analytics: a framework and case study

Sandeep Reddivari · Shirin Rad · Tanmay Bhowmik ·
Nisreen Cain · Nan Niu

- Don't focus solely on ready-to-use metrics
- Yet, the metrics are useful to improve the algorithms
- Involve practitioners!

Conclusion

- ▶ It is time for CrowdRE to assess its **value in vivo**
 - ▶ Collaboration with industry is crucial
 - ▶ Don't focus only on academic metrics of success



Conclusion

- ▶ It is time for CrowdRE to assess its **value in vivo**
 - ▶ Collaboration with industry is crucial
 - ▶ Don't focus only on academic metrics of success
- ▶ **Human effort is unavoidable**
 - ▶ AI and visual analytics can help provide a summary
 - ▶ But you can't replace the human



A new research question for the CrowdRE community?

RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



A new research question for the CrowdRE community?

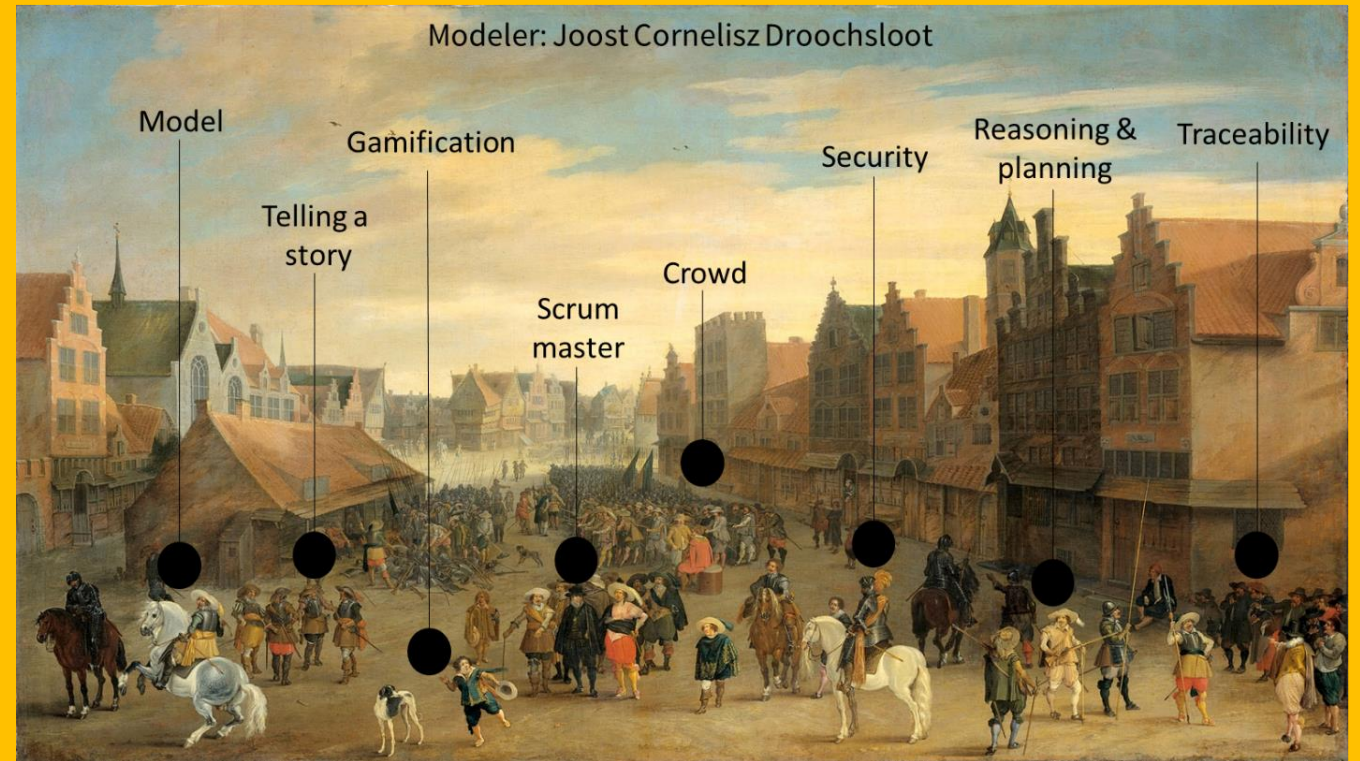
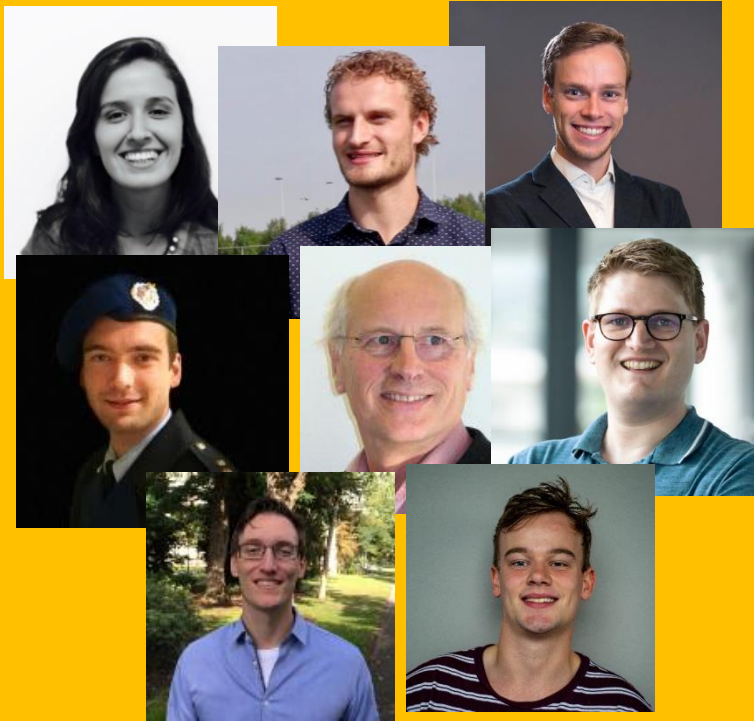
RQ: How to **reduce** the required **human effort** for managing user feedback so to **increase** CrowdRE's **value for practice**?



RQ': How to assess the delivered value of CrowdRE so to increase CrowdRE's adoption in practice?

Thanks from the **Requirements Engineering Lab** at **Utrecht University!**

Our CrowdRE crowd 😊



 f.dalpiaz@uu.nl

 [@FabianoDalpiaz](https://twitter.com/FabianoDalpiaz)