Measurements in the Early Stage Software Start-ups: A Multiple Case Study in a Nascent Ecosystem

Grace Kamulegeya *, Raymond Mugwanya †, Regina Hebig ‡

Abstract. *Context:* Software measurement is crucial to stay competitive and deliver quality software products. *Problem:* While much research has been done on measurement in large companies in developed countries, there is limited research on measurement in start-ups. So far there are no studies on whether these results apply to nascent ecosystems, such as those in East Africa. *Goal:* The aim of this study is to understand the use and perceived benefits of measurement in software start-ups in East Africa.
Method: We performed a multi-case study on 19 software start-ups in hubs in Uganda and Kenya, through conducting semi-structured interviews. We transcribed and analyzed them using the content analysis technique. *Results:* We identified that start-ups are using a number of business and product-oriented metrics. Furthermore, we found no evidence on the use of design-oriented metrics. Nonetheless, start-ups have considerable expectations on the benefits of measuring. Finally, metrics found in this study partially differ from metrics used in start-ups in developed countries.
Conclusion: There is a need to create a more inclusive characterization for measurement as early start-ups in East Africa cannot yet be represented with known models.

Keywords: Software start-up, measurement, nascent ecosystem, lean start-up

1. Introduction

Measuring is a crosscutting activity within the software life cycle. As found by Staron et al. [26], measurements activities are affecting everything from the technical to business aspects of software companies.
Measurement activities provide results that give insight, support decision making or provide actionable alerts [26] into technical and business aspects of companies.

---

*College of Computing, Makerere University, Kampala, Uganda, gkamulegeya@cit.ac.ug
†College of Computing, Makerere University, Kampala, Uganda, rmugwanya@cit.ac.ug
‡Chalmers—Gothenburg University, Gothenburg, Sweden, regina.hebig@cse.gu.se

Especially in large software companies, established measurement programs can be found [26].

However, while most research on measurements is done in large companies, little is known about measurement in software start-ups [28].

Software start-ups are newly created firms developing software intensive products and/or services [25]. These start-ups are usually disruptive organizations that are operating under uncertain conditions and face numerous challenges to grow and succeed as mature businesses. Studies have shown that many start-up ecosystems are experiencing failure rates of up to 90% [25] of their software start-ups in the first two years [9]. Thus, it is crucial to help founders to understand and measure the state of their start-ups. As Croll and Yoskovitz [7] put it: *"If you measure better, you're more likely to succeed"* [7].

Providing better support for start-ups is especially crucial in emerging regions, such as East Africa.

According to the software start-up ecosystem maturity model proposed by Cukier et al.[9], the East African ecosystem is in the nascent maturity phase. This phase is characterized by already existing start-ups, few investment deals, optional government initiatives to spur the development of the ecosystem and no relevant output in terms of jobs or worldwide penetration.

However, to the best of our knowledge, the research that exists on measurement in start-ups was performed in ecosystems that can at least be considered evolving, mature, or self-sustaining[9].

In this paper, we investigate how and what software start-ups in East African measure and the benefits of measurement they perceive and experience. We compare the results to the work of Croll and Yoskovitz [7] who propose measurements for software start-ups in more mature ecosystems.

The remainder of this paper is organized as follows. In section 2 we discuss related work. Section 3 presents the methodology used in this study. Section 4 reports the results of the multi-case study. Section 5 presents the discussions that answer the research questions. We summarize our results and discuss future work in Section 6.

2. Related Work

Related work for this research stems from two partially overlapping research directions. On the one hand, there is research on measurement systems and frameworks within the wider area of software engineering. This research focuses on the question what measurements are and can be used in software industry. On the other hand there is research on software start-ups, which is concerned with the question how software start-ups are operating and can be successful. Our study is on the intersection of these two areas. In the following we summarize relevant related work in both areas, as well as in the intersection.

Measurements in Software Development Software measures or metrics are omnipresent in software development and come in vast varieties [1]. Companies adopt

varieties of measures, which might happen in a systematic or ad-hoc way. A number of studies, have explored measurements in the broader domain of software engineering. For example very early studies looked at technical aspects of software such as software size estimation and measurements [18] using software function, source lines of code [2], and software productivity measurements [15]. There is also recent work around software metrics [13] and mapping studies highlighting the work done in areas like software product size measurement methods [3]. While these measures are in theory applicable to start-ups it is not clear whether they are also used in them.

The term measurement program is often used to describe a systematic effort of a company to collect, analyze, and visualize measures. Staron and Meding [26], in their work assess the robustness of measurement programs in large software development organizations. They use the MeSRAM method, whose aim is as they put it, is to *"support the companies to optimize the value obtained from the measurement programs and their cost"*[26]. The MeSRAM method uses an aspect called "metrics used", that organizes the metrics found in software companies using five sub categories (business metrics, product metrics, design metrics, organizational performance metrics and project metrics). Other frameworks and methods aim mostly at the assessment of measurement programs and organizational infrastructures, such as Comer and Chard's framework [6], MIS-PyME [11], and the method by Daskalantonakis et al. [10]. However, in contrast to these works, we are taking a focus on software start-ups, focusing on the concrete measurements used and their benefits.

Software Start-ups   There has been considerable research on software start-ups relating to their success factors [5], failure factors [14, 8] and challenges [27]. Paternoster et al. [19] in their behavioral framework highlight how inconsistency in managerial strategies and execution can lead to failure [14]. Klotins et al. [16] carried out a literature review and identified gaps in practices that support start-ups in successfully transitioning through the start-up life cycle. However, these studies rarely focus on measurement and related practices.

Measurements in Software Start-ups   Some empirical studies have been done on measurement in start-ups, especially in operational performance measurement. For example, Rompho [23] in their study on performance measurements in start-ups highlight that *"there is a positive relationship between the perceived importance and the performance of each metric"*. However, they found not no significant relationship between importance and performance of each metric among the various stages of start-ups. Other studies focus on specific measurement approaches. For example, Paranjape et. al. [17] conducted a secondary study to evaluate the Balanced Scorecard as a known performance measurement system in business and further examine problems associated with designing and implementing performance measures. They find the Balanced Scorecard method to still be a dominant performance measurement system but its successful implementation within business are less prevalent. However, these studies focus on start-ups from a business perspective and do not consider software start-ups, consequently ignoring software measurements. A rare example of literature focusing on measurement in software start-ups is the work of

Croll and Yoskovitz [7]. In their work, Croll and Yoskovitz categorize start-ups into six types that include software as a service, e-commerce, free mobile app, media site, user-generated content and two-sided market place. They also identify fine-grained metrics that are to be tracked by the lean start-ups.

Most of the empirical and secondary studies around measurements in start-ups are still not focused on software start-ups. Similarly much of the research on software measurements happens in mature companies. Thus, little is known today about software measurements used in software start-ups. The few works existing focus on the developed world only, ignoring regions such as East Africa. .

3.   Research Methodology

To learn about the measurements in East African start-ups, we performed an exploratory case study [31] that involved conducting interviews for start-ups which are located in different hubs in Uganda and Kenya. The main motivation for this method was that case studies, unlike experiments and survey strategies, allow the study of phenomena with unclear or unknown extent and boundary in their context [31]. Performing an exploratory study allows us to learn more about these boundaries, i.e. what comprises measurements in East African software start-ups. Our research design follows the guidelines suggested by Runeson and Höst [24] for conducting and reporting case studies.

Research Questions   To investigate the assessment and measurements within start-ups, we raise and answer the following research questions.

- RQ1: What are software start-ups in East Africa (an nascent ecosystem) assessing and measuring?

- RQ2: What benefits of measurements do the start-ups perceive?

- RQ3: How do metrics used in East African start-ups compare to those in lean start-ups in developed countries?

We define an assessment as an activity that potentially leads to a measure and a measurement as a qualitative or quantitative value that can be used for insight, decision making or alert within a start-up.

Data Collection   We conducted semi-structured interviews with start-ups founders. This is because semi-structured interviews allowed us to steer the conversation in the direction of the study, while providing the freedom to explore upcoming topics during the interview. Thus, this method provided us with the potential to discover unknown and unexpected aspects of measurement within early stage software start-ups.
We grouped the interview questions into three themes, namely (1) start-up challenges, (2) start-up growth milestones, and (3) measurements in start-ups. All questions were

independently reviewed by the co-authors and refined in several iterations to ensure that they are understandable and cover the aspects relevant for this study. Only few of the questions needed to be refined during the first four interviews in a hub in Uganda. While we had three themes in the interviews, in this paper we focus, analyze and report on the measurements in early stage software start-ups in East Africa. Each interview lasted on average 45 minutes. We recorded and transcribed all of the interviews.

Case and Subjects Selection To get in contact with start-ups for the study, we approached hubs in Kampala, Uganda and Nairobi, Kenya using convenience sampling [20]. These cities have most of the active hubs in Uganda and Kenya - probably, because of the high concentration of universities (where young developers come from) as well as availability of quality amenities and internet infrastructure. We initiated contact with hub managers and later asked them to link us to the start-ups in their hubs. This ensured that the choice of start-ups was not driven by us, but by the hub managers. When a start-up agreed to participate, we scheduled an interview with one of the active founders. This was motivated by the wish to interview someone who was conversant with both,the start-up's past and current operations.

We approached three hubs in Kampala that had prior collaboration with Makerere University. We refer to the hubs as R, O, I to maintain their anonymity. The two hubs R and O are incubation hubs, R incubating any type of start-up and O mainly incubating software start-ups. Hub I, is hybrid (a co-working space and an incubation hub), but predominantly targets software businesses.

In Kenya, we made contact with hubs with the help of a local researcher. After initial contact with six hubs, three agreed to participate. We refer to these hubs as C, W, and N. Hub C, is an incubation hub for all types of start-ups while W and N are co-working spaces.

Thus, we visited six hubs in total, that included, 2 co-working spaces (W, N), 3 incubation spaces (R,O,C) and 1 hybrid hub (I). We initially carried out 23 interviews in all the six hubs. We interviewed 10 start-ups from hub I, because its is the largest hub in Kampala, Uganda, 2 from O, 4 from R, 4 from C, one from N and 2 from W. Of these interviews, we later excluded 4 (1 from hub I and 3 from hub R) for this study during the analysis phase, since it turned out that they were not actually developing software or using any software in their business. We interviewed fewer start-ups in Kenya compared to Uganda, since we were there for only one week. Table 1 characterizes 19 start-ups included in this study. We indicate the type of the hosting hub (incubation hub, co-working space, or hybrid), the age of the start-ups in months, the start-up type and whether the start-up is measuring. We use codes to represent the start-ups and hubs to maintain their anonymity. Table 2 summarizes the age structure of our start-ups.

Analysis We imported all interview transcripts into Excel spreadsheets. In each excel sheet, we organized the responses to each question under the corresponding question in the interview guide. This was done to allow for tracking of the emerging codes and quotations in the raw transcripts. We used coding to analyze the interview

Table 1. Characterization of Interviewed software start-ups (Uganda = UG, Kenya = KE)

No	Hub ID	Start-up ID	Host Country	Hosting Type	Hub Age (Months)	Start-up Type	State of assessment/ Measuring
1	**C**	CB	KE	Incubation Hub	24	Software as a Service	Measuring
2		CZ	KE	Incubation Hub	11	Software as a Service	Measuring
3		CES	KE	Incubation Hub	12	Software as Service	Measuring
4		CE	KE	Incubation Hub	10	Mobile App	Measuring
5	**R**	RI	UG	Incubation Hub	51	e-commerce	Measuring
6	**O**	OI	UG	Incubation Hub	48	Using software in their workflow	Measuring
7		OC	UG	Incubation Hub	15	Fintech	Measuring
8	**W**	WT	KE	Co-working space	10	Two-sided marketplace	Measuring
9		WE	KE	Co-working space	12	Two-sided marketplace	Measuring
10	**N**	NF	KE	Co-working space	43	Software as a Service	Measuring
11	**I**	IH	UG	Hybrid	7	e-commerce	Measuring
12		IY	UG	Hybrid	27	User-generated content	Measuring
13		IT	UG	Hybrid	14	Software development	Measuring
14		ID	UG	Hybrid	12	Media Site	Measuring
15		IA	UG	Hybrid	9	Software as a service	Measuring
16		IR	UG	Hybrid	8	Mobile App	Measuring
17		IN	UG	Hybrid	15	Fintech	Measuring
18		IP	UG	Hybrid	32	Two-sided marketplace	Measuring
19		IS	UG	Hybrid	39	Software as a service	Measuring

Table 2. Start-up age characterization

Age Bracket (Months)	Start-ups
0-6	None
7-12	CZ, CES, CE, IH, ID, IA, IR, WT, WE
13-18	IT, IN, OC
19-24	CB
25-30	IY, IP
31-36	IS
>36	RI, NF, OI

transcripts. This involved reading through each transcript, line by line to derive the emerging codes. Runeson and Höst [24] point out that this analysis is a suitable approach for analyzing software engineering case studies. After the first round of coding, we grouped the emerging codes from the 19 transcripts, resulting in high-level groups that included, "progress assessment", "what to measure", "benefits of measurement", "use of the measurement dashboard" and "benefits of using the dashboard". For each emerging code, we identified matching start-ups by examining each start-up transcript in related interview question. We reviewed the grouped codes to clarify the ambiguous ones, merge and /or split them.

Afterwards, we performed a second round of coding, using the merged codes. With the help of the codes, we were able to analyze and discuss the interviews to identify and characterize measurements used and benefits perceived by the start-ups. To analyze and discuss the results emerging from the coding, we make use of parts of

MeSRAM [26]. This method is one of the most complete and recent approach to assess measurement programs in large software firms. So far it has mostly been applied in large software companies. However, MeSRAM's aspect "metrics used" provides sub categories of metric types (business metrics, product metrics, design metrics, organizational performance metrics and project metrics), which are also software start-ups. We apply these sub categories, to systematize the measurements that emerged from the synthesis of the codes from the start-up interviews in this study.

To gain further insight into the use of measurements and their benefits in early stage start-ups, we performed a simple cross-case analysis [30] to identify, whether variables such as start-up age, hub type, and country have an effect on 1) how many metrics are used, 2) how many metrics are wished for, 3) how many benefits are experienced, and 4) how many benefits are expected by the start-ups. Since the number of data-points is limited, we decided to group the start-ups for each variable under study, so that the gained groups are meaningful and not too small for statistical testing. As a result we work with the following groups:

Variable 'start-up age' (3 groups): young (0-12 months, n = 9); medium (13-24 months, n = 4); old ($\geq$ 25 months, n = 6)

Variable 'hub type' (3 groups): Incubation hub (n = 7); Co-working space (n = 3); Hybrid (n = 9)

Variable 'country' (2 groups): Kenya (n = 7); Uganda (n = 12)

We do not consider the start-up type as a variable here, since we have for most types, only one or two start-ups, which is not enough to form a representative sample per type.
For the test, we decided to used the Wilcoxon Signed-Rank Test [4] that can be used to test the probability that two samples belong to the same population. Furthermore, the Wilcoxon test has the advantage that it can be applied to data that has a normal distribution and data that is not normally distributed. We focus on the following null hypotheses:

- $H0_1$: The studied variables have no impact on the number of measurements used by a start-up.

  - $H0_{1a}$: The start-up age has no impact on the number of measurements used by a start-up.
  - $H0_{1b}$: The hub type has no impact on the number of measurements used by a start-up.
  - $H0_{1c}$: The country has no impact on the number of measurements used by a start-up.

- $H0_2$: The studied variables have no impact on the number of measurements wished for by a start-up.

- $H0_{2a}$: The start-up age has no impact on the number of measurements wished for by a start-up.
- $H0_{2b}$: The hub type has no impact on the number of measurements wished for by a start-up.
- $H0_{2c}$: The country has no impact on the number of measurements wished for by a start-up.

- $H0_3$: The studied variables have no impact on the number of benefits experienced by a start-up.
  - $H0_{3a}$: The start-up age has no impact on the number of benefits experienced by a start-up.
  - $H0_{3b}$: The hub type has no impact on the number of benefits experienced by a start-up.
  - $H0_{3c}$: The country has no impact on the number of benefits experienced by a start-up.

- $H0_4$: The studied variables have no impact on the number of benefits expected by a start-up.
  - $H0_{4a}$: The start-up age has no impact on the number of benefits expected by a start-up.
  - $H0_{4b}$: The hub type has no impact on the number of benefits expected by a start-up.
  - $H0_{4c}$: The country has no impact on the number of benefits expected by a start-up.

In addition, we investigate whether the number of benefits experienced is related to the number of measurements used, the number of measurements wished for, or the number of benefits expected:

- $H0_5$: Whether benefits are experienced do not correlate with the number of benefits expected, measurements used, or measurements wished for by a start-up.
  - $H0_{5used}$: Whether benefits experienced do not correlate with the number of measurements used by a start-up.
  - $H0_{5wished}$: Whether benefits experienced do not correlate with the number of measurements wished for by a start-up.
  - $H0_{5expected}$: Whether benefits experienced do not correlate with the number of benefits expected by a start-up.

Finally, we compared the measurements in early stage software start-ups in East Africa to the lean metrics presented by Croll and Yoskovitz [7]. For that we use Croll's categories to group the East Africa start-ups and match the found metrics to the lean start-up metrics described by Croll and Yoskovitz. We discuss the results of the comparison of these metrics in the start-ups and existing lean start-ups in Section 4 and 5.

4. Results

In this section we show the results of our study and highlight observations with regards to what measurements were found in the East African software start-ups, what measurement benefits these start-ups perceive, and to what degree the found measurements match to those found in Croll and Yoskovitz [7] work.

4.1. Assessments and Measurements in East African Software Start-ups

In Tables 3 and 4 we summarized what assessments and measurements we found in the early stage start-ups in East Africa. Our interviewees mentioned metrics that they use and they would wish to use in future (probably starting to use them as they grow or as they get resources and capabilities).

Table 3. Measurements and assessments in start-ups categorized based on MeSRAM[26] , ["-" = No Start-ups, "Y" = Yes, "N" = No]

No	MeSRAM Metric Category	Found Measurements/ Assessments	Start-ups Using the metric	Start-ups Wishing for the metric	Used in Start-ups
1	Business Metrics	Customer analytic (Number of people using platform, customer behaviour)	OI, WT, IY, IR, IP, NF	WE, IH, CE, ID , IA	Y
2		Product delivery process time estimation (Delivery reliability measurement)	-	IH	N
3		Rate of customer/partner acquisition/growing customer base	CES, NF	IS	Y
4		Revenue growth/generated revenue/activities that generate revenue	CES, IN	RI	Y
5		Using a telemetry tool	IN	-	Y
6		Tracking market indicators/market events	-	ID	N
7		Ability to close a business deal	CES	-	Y
8		Set and review business targets	IH	-	Y
9		Product awareness/customer interest	IY, IP	-	Y
10		Using market as a benchmark	OC	-	Y
11		Customer feedback measurement	OC	IY	N
12		Reaching key business milestones (patents, tax registration, incorporation)	WE	-	Y

Altogether, we found 28 assessments and measurements (also called metrics) from the 19 start-ups. Each of the 19 start-ups had at least one way of assessing for progress and some additional measurements they use or wish to use.
To understand these metrics, we categorized them based on the sub-categories in the "metric used" aspect of the MeSRAM model [26]. This model targets the robustness of a companies measurement program and is often applied to large software company. There are five subcategories: business, product, design, organizational performance and project metrics (Staron and Meding [26]). Thus, categorizing the metrics found in

Table 4. Measurements and assessments in start-ups categorized based on MeSRAM[26] (contin.), [”-” = No Start-ups, ”Y” = Yes, ”N” = No]

No	MeSRAM Metric Category	Found Measurements/ Assessments	Start-ups Using the metric	Start-ups Wishing for the metric	Used in Start-ups
13	Product Metrics	Product/feature usage	-	IN, IY, CE, OC	N
14		Production process time estimation	IH	-	Y
15		System reliability	-	RI	N
16		Ability to build a complete product	CB, CES	-	Y
17		Feedback from friends about product features (peer endorsement)	CZ	-	Y
18		Product maintenance/support	CES	-	Y
19		Comparing product versions (added features)	CE	-	Y
20	Organizational	Set and evaluate Key Performance Indicators (KPIs)	IP, ID, WT	-	Y
21	Performance	Time-based task setting, tracking and review for progress of project/staff	NF, WT	RI	Y
22	Metrics	Time-based project performance appraisal	NF	-	Y
23	Project Metrics	Monetary value of time spent on task/activity	-	IT	N
24		Set and evaluate tasks	CZ	-	Y
25		Activity completion time	-	IT	N
26		Process adherence by the team	-	OC	N
27		Tools usage by team	-	WT	N
28		Documenting and reviewing activities for progress	IS, WE, WT	-	Y

start-ups using the ”metrics used” sub-categories provides a first idea on how start-up metrics may relate to metrics used in more mature companies.

Business metrics 12 of the 28 metrics are business oriented, and most of the start-ups (17 of 19) are using or wish to use them. 12 start-ups are already using at least one business metric. Furthermore, 8 start-ups wish to use business metrics. For example, customer analytic is a metric that consists of more fined-grained metrics like number of downloads, number of user accounts, number of people using the platform or customer behaviour. This metric is the most popular business metric within the interviewed start-ups. Six start-ups (OI, WT, IY, IR, IP, NF) are already using it and five start-ups (WE, IH, CE, ID, IA) are wishing to use it.
Of the 12 business metrics found, 10 are already used by different start-ups and only two metrics are just wished for (Product delivery process time estimation and tracking market indicators/market events).

Product metrics Product related metrics are the second most popular metrics with 7 out of 28. These metrics are used or wished for by 9 different start-ups. 5 of the 7 product metrics are also already in use (by 5 different start-ups). There are however, 2 metrics (Product/feature usage and System reliability) that are not used

at all but only wished by some start-ups in this category. The product/feature usage metric is wished for by 4 of the 9 start-ups, but not used.

Organizational performance metrics Organizational performance metrics category has 3 of the 28 metrics we found in the start-ups. It is the category with the least number of used and wished for metrics in the interviewed start-ups. However, all of three metrics in this group are used by at least one of the 4 start-ups (IP, ID, WT, NF). Only one start-up (RI) wishes to use a metric in this category in future ("Time-based task setting, tracking and review for progress of project/staff").

Project metrics We found 6 project oriented metrics from the start-ups. However, in contrast to organizational performance metrics, project metrics are mostly wished for by the start-ups. Only 2 of those 6 metrics are actually used by the start-ups. Of the 6 different start-ups using and /or wishing to use these project metrics, 4 are using at least one of them. The most popular used project metric is documenting and reviewing activities for progress (IS, WE, WT).

Design metrics Most interestingly, none of the start-ups uses or wishes to use design metrics.

Table 5. Perceived Sufficiency of Measuring in start-ups

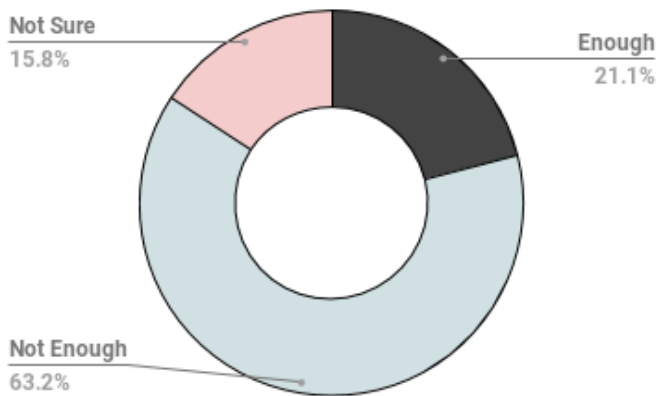| Perceived sufficiency of Measuring	Start-ups
Enough	NF, IN, IS, IP
Not Enough	CB, CZ, CES, CE, IO, OC, WT, WE, IY, IT, ID, IR
Not Sure	RI, IH, IA

Figure 1. Perceived Sufficiency of Measuring by start-ups

Perceived Sufficiency of Measuring by start-ups   We also asked whether the start-ups perceived their measuring as sufficient. Table 5 and Figure 1 summarize the answers. Of the 19 start-ups, 4 start-ups (NF, IP, IN, IS) affirmatively said that they were measuring enough, 12 start-ups do not think they are measuring enough, and 3 start-ups were not sure whether they measure enough.

Interestingly, 3 of start-ups were not using any measures at all (RI, IT, and IA). All of those claimed that they are not measuring enough or were at least not sure about whether they measure enough.

To sum up, the majority of the start-ups use at least one or more metrics and most of the start-ups conceded that they are not measuring enough. Our results have shown that there is wide variation in the used and wished measures. It is possible that this high variety is a symptom of missing best practices for measurement in start-ups. Future research will have to show whether common knowledge about best practices would reduce the visible variety.

4.2.   Perceived Benefits of Measurements

Given that some start-ups were measuring and some wishing to measure aspects of their businesses, we found it important to find out what they expected to benefit or experience from measuring. All, but one, of the start-ups clearly stated what they do or would like to use metrics (see Tables 3 and 4).

Table 6 lists the benefits that the start-ups are already experiencing or expecting when they measure.

Our results reveal 22 benefits experienced and/or expected by the 19 start-ups. The majority (17 of the 22 benefits), are expected and only 5, are actually experienced by the start-ups. The five experienced benefits are expressed by three start-ups (CZ, NF, IS), which are all from different hubs in Kenya and Uganda.

Figure 2, shows the number of expected or experienced benefits per start-up. A majority of the start-ups (16 of 19) have experienced or expect two or less benefits from measuring. There are however 3 start-ups (OC, IS, ID) that expect or have experienced 3 or more benefits of measuring. OC, a Financial Technology (fintech) start-up of 12 months, expects measuring to help a) detecting areas of success/strength and failure/weaknesses, b) indicate progress of its team members, and c) unveil its managerial/administrative ability. They had this to say:

"We shall be able to detect failure early in time meaning you can see them before they become catastrophic. If you are able to do that, then you are able to work on them. They also show you where you are making great success and within the team, they show you who is moving on well and who is not. It will show you, in terms of management, can someone handle a given area or a change is required in order to be able to match our competition."[OC].

Start-up, ID, a 12 months media site expects measuring to enable it a) to know what to focus on, b) to provide insights into the start-ups operations, and c) to make appropriate resource allocations:

Table 6. Perceived Benefits of Measuring in Start-ups

No	Benefits	Start-ups Expressing Benefit	Experienced or Expected
1	Proof that the product is being used versus just download-ing	CE	Expected
2	Reflect on how to earn(activities that generate rev-enue)/measuring revenue growth	CES, IN	Expected
3	Notice Change (growth)	CB, CES	Expected
4	**Gauging ability to deliver working product/service**	**CZ**	**Experienced**
5	**Profiling growth(areas of growth, comparing growth)**	**CZ, NF**	**Experienced**
6	Know what to focus on (based on success or failure)/where to focus(resources)/direction	OI, RI, IY, IS, IA, ID	Expected
7	Negotiating deals with reliable Information	OI, IS	Expected
8	Detect areas of success/strength and failure/weaknesses	OC, IY, IP	Expected
9	Indicate progress of team members	OC	Expected
10	Show managerial/administrative ability	OC	Expected
11	Provide a history of solutions that can aid solve current start-up problems	WT	Expected
12	It helps in providing evidence for securing start-up funding from investors	WE	Expected
13	**Show areas of positive growth/success**	**NF**	**Experienced**
14	Aids in timely delivery of projects and ensuring value for money	IT	Expected
15	**Determining the team size**	**IS**	**Experienced**
16	**Knowing how to grow and scale**	**IS**	**Experienced**
17	Helps in knowing revenue generating activities in a start-up	IN	Expected
18	Improving product based on measuring usage	IR	Expected
19	Can help in product feature-based pricing	IR	Expected
20	Understanding and optimizing product value chains	IH	Expected
21	Providing insights into operations	ID	Expected
22	Aids in making appropriate resource allocation	ID	Expected

"We can know where to focus our attention and also know how to plan for our time especially when attending events, you can know who is coming for an event or not from experience. [...] So tracking helps us know and give insights for certain events and their turn up. [...] There is a lot of mis-allocation of resources so my start-up will help allocate resources appropriately. Why would I talk about machine learning yet there is no one who has a start-up in that field?" [ID].

Finally, start-up IS, a 39 months old Software as a service (SaaS) start-up, has experienced 2 measuring benefits (enabling to determine the team size and know how to grow and scale) and expects two other benefits from measuring (negotiate deals using reliable information and to know what to focus on) :

"[...] The biggest benefit is that we can know how to grow and how to scale and where to focus our resources. For example, we are not focusing on development a lot now because what we want to first grow to a certain point and then come back and add features. Secondly, it helps us determine the size of our team. Thirdly to make business decisions like should we get someone to invest in us or should we use what is coming in for a longer time." [IS]
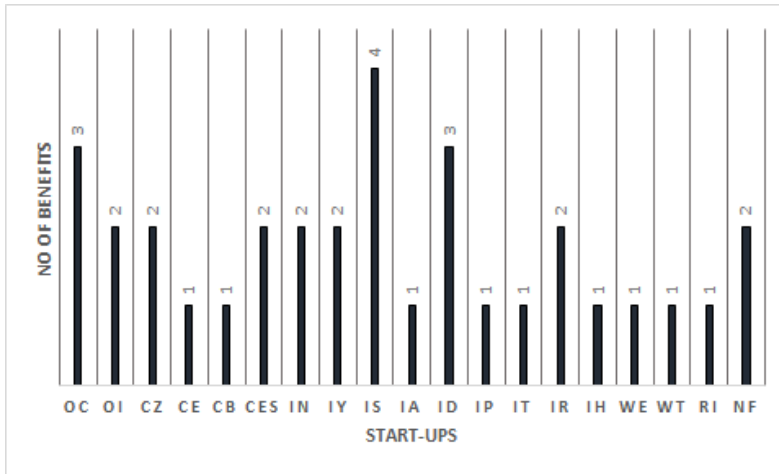
Figure 2. Number of Measuring Benefits Expected and/or Experienced per Start-up

The most popular expected benefit is knowing what to focus on (based on success or failure of operations of the start-up) or where to focus resources. It was mentioned by 6 start-ups (OI, RI, IY, IS, IA, ID). These start-ups come from the three hubs (O, I, R) we visited in Kampala. Following are some example quotes expressing this expectation:

"Strategy: we shall know what to focus on but also that is key in new markets. Because if you know what to focus on, you don't have to start from scratch. So, our measurements will help us understand what we need to focus on but also it will help us in our partnerships for example I can tell the courier that for me I will bring mostly this amount of goods so give me a deal and it helps us work with the stalls." [OI]

"Well because we are all over the place, you have to spend your efforts in doing the right thing and the only way you are doing the right thing is if you are measuring. If you have a goal like moving from here to Mbarara, you have to know that you are 50km in, moving at 20km/hr and you will arrive at 4:00 pm. So, it gives you a chance to know what the right directions are so that you don't go astray and speed towards like Mabira forest. Measuring is the key." [RI]

"We can know where to focus our attention and also know how to plan for our time especially when attending events, you can know who is coming for an event or not from experience." [ID]

Finally, our results also show that 16 of the 22 benefits where mentioned by a single start-up, only. Thus, only few benefits are shared among the start-ups.

Table 7. Data by variable groups

Variable	Group	n	average # metrics	average used wished metrics	# average # experienced benefits	average # expected benefits
Start-up age	young	9	1.89	1.00	1.22	0.33
	medium	4	1.25	1.25	1.75	0.00
	old	6	1.83	1.00	1.33	0.67
Hub type	Incub. hub	7	1.57	1.00	1.43	0.29
	Hybrid	9	1.33	1.22	1.56	0.33
	Co-work.	3	3.33	0.67	0.67	0.67
Country	Kenya	7	2.57	0.57	0.86	0.57
	Uganda	12	1.25	1.33	1.67	0.25
Benefits	Yes	4	2.00	0.75	-	0.50
Experienced	No	15	1.67	1.13	-	1.60

Table 8. A Wilcoxon Signed-Rank test for used measurements.

Variable	x	y	hypothesis (used_metrics(x) = used_metrics(y))	W	p-value	result
Start-up age	young	medium	$h01a_{ym}$	20.5	0.75	$h01a_{ym}$ not rejected
	medium	old	$h01a_{mo}$	9.5	0.66	$h01a_{mo}$ not rejected
	young	old	$h01a_{yo}$	27	1	$h01a_{yo}$ not rejected
Hub type	Incub. hub	Co-work.	$h01b_{ic}$	3.5	0.12	$h01b_{ic}$ not rejected
	Incub. hub	Hybrid	$h01b_{ih}$	30.5	0.96	$h01b_{ih}$ not rejected
	Co-work.	Hybrid	$h01b_{ch}$	2.5	0.045	$h01b_{ch}$ not rejected (Bonferroni corr.)
Country	Kenya	Uganda	$h01c_{ku}$	60	0.12	$h01c_{ku}$ not rejected
Benefits Experienced	Yes	No	$h05_{used}$	34.5	0.68	$h05_{used}$ not rejected

4.3. Cross-Case Analysis

Table 7 summarizes the data used as input for the cross-case analysis. The Tables 8, 9, 10, and 11 below show the results of the Wilcoxon signed-rank test for the different variable groups. We work with a threshold of 0.05 for the p-value.

Concerning hypothesis $H0_1$ and its sub-hypotheses, we compared each pair of groups within the 3 variables (start-up age, hub type, and country) as shown in Table 8. As there are no significant difference between the different start-up ages, we conclude that we cannot reject hypothesis $H0_{1a}$ (the start-up age has no impact on the number of used metrics). A similar result is found when comparing the different hub types. However, when comparing metrics used in co-working spaces and hybrid hubs, we have a p-value below 0.05. In itself this would be a significant result. However, we have to take into account that hypothesis $H0_{1b}$ is tested 3 times ($h01b_{ic}$, $h01b_{ih}$, and $h01b_{ch}$) causing a multiple comparison problem (the more often we test a hypothesis, the higher the probability to get a significant result). To avoid that problem, we perform a Bonferroni correction on the threshold, by dividing the threshold by the

Table 9. A Wilcoxon Signed-Rank test for wished measurements.

Variable	x	y	hypothesis (wished_metrics(x) = wished_metrics(y))	W	p-value	result
Start-up age	young	medium	$h02a_{ym}$	15	0.68	$h02a_{ym}$ not rejected
	medium	old	$h02a_{mo}$	14	0.74	$h02a_{mo}$ not rejected
	young	old	$h02a_{yo}$	28.5	0.90	$h02a_{yo}$ not rejected
Hub type	Incub. hub	Co-work.	$h02b_{ic}$	11	1	$h02b_{ic}$ not rejected
	Incub. hub	Hybrid	$h02b_{ih}$	27	0.65	$h02b_{ih}$ not rejected
	Co-work.	Hybrid	$h02b_{ch}$	19	0.32	$h02b_{ch}$ not rejected
Country	Kenya	Uganda	$h02c_{ku}$	23.5	0.11	$h02c_{ku}$ not rejected
Benefits Experienced	Yes	No	$h05_{wished}$	23.5	0.53	$h05_{wished}$ not rejected

Table 10. A Wilcoxon Signed-Rank test for experienced benefits.

Variable	x	y	hypothesis (experienced_benefits(x) = experienced_benefits(y))	W	p-value	result
Start-up age	young	medium	$h03a_{ym}$	14	0.39	$h03a_{ym}$ not rejected
	medium	old	$h03a_{mo}$	8	0.28	$h03a_{mo}$ not rejected
	young	old	$h03a_{yo}$	23	0.59	$h03a_{yo}$ not rejected
Hub type	Incub. hub	Co-work.	$h03b_{ic}$	8.5	0.62	$h03b_{ic}$ not rejected
	Incub. hub	Hybrid	$h03b_{ih}$	29.5	0.81	$h03b_{ih}$ not rejected
	Co-work.	Hybrid	$h03b_{ch}$	11.5	0.71	$h03b_{ch}$ not rejected
Country	Kenya	Uganda	$h03c_{ku}$	48	0.51	$h03c_{ku}$ not rejected

Table 11. A Wilcoxon Signed-Rank test for expected benefits.

Variable	x	y	hypothesis (expected_benefits(x) = expected_benefits(y))	W	p-value	result
Start-up age	young	medium	$h04a_{ym}$	12.5	0.41	$h04a_{ym}$ not rejected
	medium	old	$h04a_{mo}$	9.5	0.65	$h04a_{mo}$ not rejected
	young	old	$h04a_{yo}$	24	0.75	$h04a_{yo}$ not rejected
Hub type	Incub. hub	Co-work.	$h04b_{ic}$	15.5	0.27	$h04b_{ic}$ not rejected
	Incub. hub	Hybrid	$h04b_{ih}$	28.5	0.78	$h04b_{ih}$ not rejected
	Co-work.	Hybrid	$h04b_{ch}$	21.5	0.14	$h04b_{ch}$ not rejected
Country	Kenya	Uganda	$h04c_{ku}$	20.5	0.06	$h04c_{ku}$ not rejected
Benefits Experienced	Yes	No	$h05_{expected}$	10.5	0.04	$h05_{expected}$ rejected

number of tests, i.e. $0.05/3 = 0.016$. Given that adjusted threshold, $h01b_{ch}$ is not rejected as well. Thus, our data does not indicate any impact of the hub type on the number of used metrics ($H0_{1b}$ is not rejected). Finally, we do not find an impact of the country on the number of used metrics ($H0_{1c}$ is not rejected).

We approached hypothesis $H0_2$ and its sub hypotheses similarly to hypothesis $H0_1$. As Table 9 illustrates, neither of the sub-hypotheses can be rejected. Thus, our data does not show an impact of the variables (start-up age, hub type, and country) on the number of metrics wished for by a start-up.

We find the same results, when investigating the number of benefits experienced ($H0_3$) and the number of benefits expected ($H0_4$) by the start-ups. None of the sub-hypotheses can be rejected (as shown in Tables 10 and 11), indicating that the three variables have no impact on the number of benefits experienced or expected.

Finally, we tested hypothesis $H0_5$, by comparing start-ups that experience benefits with start-ups that experience no benefits, with regards to the number used metrics ($h05_{used}$), the number of wished-for metrics ($h05_{wished}$), and the number of expected benefits ($h05_{expected}$) (shown in Tables 8, 9, and 11). The results do not allow us to reject $H0_{5used}$ and $H0_{5wished}$. Thus, it seems that there is no relationship between the number of used or wished for metrics and the experienced benefit of metric use. However, we can reject $H0_{5expected}$, with an average number of expected benefits of 0.5 for start-ups already experiencing benefits and 1.6 for start-ups that do not yet experience benefits. Thus, our data indicates that there is some form of relation between the number of benefits expected and the number of benefits experienced.

4.4. Comparison to Lean Start-ups in Developed Countries

To better understand how the metrics found in the East African context relate to the other contexts we decided to compare them to the metrics for lean start-ups described by Croll and Yoskovitz [7] which were collected in context of developed countries. In Tables 12, 13, and 14, we present the comparison matrix. The lean metrics presented by Croll and Yoskovitz [7], are associated to start-up categories and are more fine grained/specific than our aggregated list in Tables 3 and 4.

Thus, we first mapped our start-ups to Croll and Yoskovitz [7] start-ups categories (shown in column 7 in Table 1). In a second step, we matched the metrics used/wished for by the start-ups to the metrics listed by Croll and Yoskovitz. Note, that we refined the description of the metrics here to reflect more specifically what happens in the start-ups and enable the mapping.

The goal of this matching was to establish the common metrics and unused metrics in East Africa start-ups compared to known lean start-up metrics in the developing world.

Start-up categories   The categories by Croll and Yoskovitz [7] include e-commerce, Software as a Service (SaaS), free mobile app, two-sided marketplaces, media site, and user generated content. We already categorized the start-ups in Table 1 based on these categories.

Table 12. A match of Croll's and Yoskovitz's Lean start-up metrics [7] to the metrics found in our study.

Type and Start-ups	Lean start-up Metrics	Metric Description	Matching/Related Lean and start-up metrics
Software as a service (SaaS): CB, CZ, CES, NF, IA, IS	Attention	How effectively the business attracts visitors.	-
	Enrollment	How many visitors become free or trial users.	3 (Rate of user acquisition new users and merchants - IS)
	Stickiness	How much the customers use the product.	1 (Web page usage tracking - NF)
	Conversion	How many users become paying customers and switch to a higher-paying tier.	-
	Revenue per customer	How much money a customer brings in a given time period.	4 (Amount of generated Revenue - CES)
	Customer acquisition cost	How much it costs to get a paying user.	3 (Customer growth- CES)
	Virality	How likely customers are to invite others and spread the word, and how long it takes them to do so.	-
	Up-selling	How often and why customers increase their spending.	-
	Up-time and reliability	Number of complaints, problem escalations, or outages.	18 (Product maintenance -CES)
	Churn	How many users and customers leave in a given time period.	3 (Customer Retention- NF),
	Lifetime value	How much customers are worth from cradle to grave.	-
E-commerce: RI, IH	Purchases per year	The number of purchases made by each customer per year.	4 (Transaction revenue volume - RI)
	Average shopping cart size	The amount of money spent on a purchase.	4 (Transaction revenue volume - RI)
	Abandonment	The percentage of people who don't finish a begun purchase.	-
	Cost of customer acquisition	The money spent to get someone to buy something.	-
	Revenue per customer	The lifetime value of each customer.	-
	Top keywords driving traffic to the site	Those terms that people are looking for, and associate with you—a clue to adjacent products or markets.	-
	Top search terms	Both those that lead to revenue, and those that don't have any results.	-
	Effectiveness of recommendation engines	How likely a visitor is to add a recommended product to the shopping cart.	-
	Virality	Word of mouth, and sharing per visitor.	1 (customer behavior analytic -IH)
	Conversion rate	The number of visitors who buy something.	-
	Mailing list effectiveness	Click-through rates and rates of buyers returning.	-

Table 13. A match of Croll's and Yoskovitz's Lean start-up metrics [7] to the metrics found in our study. (contin.)

Type and Start-ups	Lean start-up Metrics	Metric Description	Matching/Related Lean and start-up metrics
	Downloads	How many people have downloaded the application, as well as related metrics such as app store placement, and ratings.	1 (Number of downloads - CE),
Free Mobile App: CE, IR	Customer acquisition cost (CAC)	How much it costs to get a user and to get a paying customer.	-
	Launch rate	The percentage of people who download the app, actually launch it, and create an account.	1 (Number of new accounts - IR), 1 (Number of activations - IR),
	Percent of active users/players	The percentage of users who've launched the application and use it on a daily and monthly basis: these are your daily active users (DAU) and monthly active users (MAU).	1 (Number of daily active users - CE), 1 (number of users - CE), 1 (Number of application installations - CE, IR)
	Percentage of users who pay	How many of your users ever pay for anything.	1 (Frequency of platform usage - CE)
	Time to first purchase	Time between account activation and first purchase.	-
	Monthly average revenue per user (ARPU)	Revenue per user through purchases and watched ads.	1 (Frequency of platform usage - CE)
	Ratings click-through	Percentage of users adding ratings or reviews.	-
	Virality	On average, how many other users a user invites.	-
	Churn	How many customers have uninstalled the application, or haven't launched it in a certain time period.	1 (Number of uninstallations - IR)
	Customer life-time value	How much a user is worth from cradle to grave.	-
Media site: ID	Audience and churn	How many people visit the site and how loyal they are.	1 (Number of followers on social media platforms - ID)
	Ad inventory	The number of impressions that can be monetized.	-
	Ad rates	Sometimes measured in cost per engagement—essentially how much a site can make from those impressions based on the content it covers and the people who visit.	1 (Tracking articles - ID)
	Click-through rates	How many of the impressions actually turn into money.	-
	Content/advertising balance	The balance of ad inventory rates and content that maximizes overall performance.	1 (Key words -ID),

Table 14. A match of Croll's and Yoskovitz's Lean start-up metrics [7] to the metrics found in our study. (contin.)

| Type and Start-ups	Lean start-up Metrics	Metric Description	Matching/Related Lean and start-up metrics
User-Generated Content: IY	Number of engaged visitors	Frequency and length of customer visits.	1 (Number of returning customers - IY)
	Content creation	The percentage of visitors who interact with content in some way, from creating it to voting on it.	-
	Engagement funnel changes	How well the site moves people to more engaged levels of content over time.	1 (Number of people using the platform - IY)
	Value of created content	Business benefit of content, from donations to media clicks.	-
	Content sharing and virality	How content gets shared, and how this drives growth.	-
	Notification effectiveness	The percentage of users who, when told something by push, email, or another means, act on it.	9 (Product awareness by public - IY)
Two-sided Market Place: WT, WE, IP	Buyer and seller growth	The rate at which you're adding new buyers and sellers, as measured by return visitors.	1 (Number of organizations contacted in a quarter - IP), 1 (Number of walk-in clients - IP)
	Inventory growth	The rate at which sellers are adding inventory—such as new listings—as well as completeness of those listings.	20 (Performance of retailers and wholesalers - WT)
	Search effectiveness	What buyers are searching for, and whether it matches the inventory you're building.	1 (Number of site visits - IP)
	Conversion funnels	The conversion rates for items sold, and any segmentation that reveals what helps sell items	-
	Ratings and signs of fraud	The ratings for buyers and sellers, signs of fraud, and tone of the comments.	-
	Pricing metrics	If you have a bidding method in place (as eBay does),then you care whether sellers are setting prices too high or leaving money on the table.	-
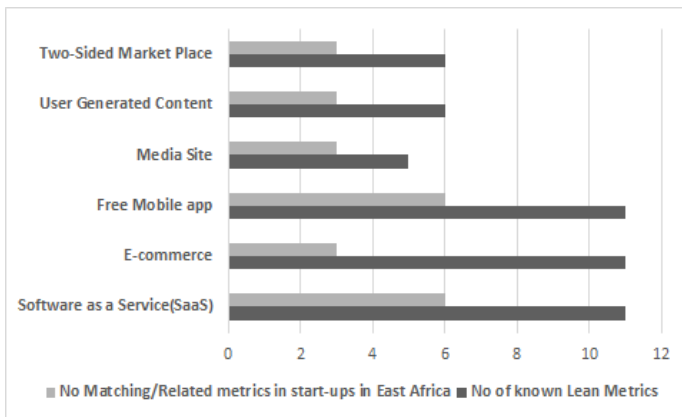
Figure 3. Number of Lean metrics and matching/related start-up metrics

However, it was difficult to fit some start-ups found in East Africa in these categories. IN and OC are better categorized as Fintechs (start-ups that provide payment and financing solutions to customers, using existing telecommunication and banking infrastructure). OI can be categorized as a start-up that uses software as part of its work-flow and IT can be categorized as a software development start-up (developing software solutions for other companies).

Metrics Comparison In Figure 3, we illustrate the number of matching/related metrics found in the start-ups in East Africa in comparison to those already known in Lean start-ups presented by Croll and Yoskovitz [7].

Overall, we could match 21 of the metrics found in East African start-ups to the 71 metrics listed by Croll and Yoskovitz (Software as a Service (SaaS)(6/11); E-commerce (3/11), Free mobile app (6/11), media site (3/5), user-generated content (3/6) and Two-sided market place (3/6)). For at least half of the Croll's and Yoskovitz's metrics we found matching metrics in the start-ups, with exception of E-commerce.

However, we also found metrics in the East African start-ups that could not be matched or related to the known lean start-up metrics. Table 15 summarizes these metrics. Note that the table shows the metrics found only in the listed East African start-ups. We provide the ID from the corresponding aggregated metric in Tables 3 and 4.

On matching with lean start-up metrics, we found 21 new metrics in the studied early stage start-ups. These include: 7 in the SaaS start-ups, 6 in e-commerce start-ups, one in free mobile apps, 2 in media site start-ups, 2 in user-generated content start-ups and 3 in two sided market places.

Table 15. Unique metrics found in East Africa start-ups in comparison to Lean start-ups metrics [7]

Start-up categorization	comparatively unique metrics found in East Africa start-ups
Software as a service (SaaS)	16 (Ability to build a complete product - CB and CES), 24 (Set and evaluate tasks - CZ), 17 (Feedback from friends about product features - CZ), 7 (Ability to close a business deal - CES), 21 (Time-based task setting, tracking and review for progress of project/staff - NF), 22 (Time-based project performance appraisal - NF), 1 (Track impact of advice-IA)
E-commerce	15 (System reliability - RI), 21 (Time-based task setting, tracking and review for progress of project/staff - RI), 6 (Tracking market indicators/market events - RI), 8 (Set and review business targets - IH), 14 (Production process time estimation - IH), 2 (Product delivery process time estimation - IH)
Free Mobile App	13 (Usage per product feature - IR)
Media site	6 (Number of events attended - ID), 1 (Posting time by customers - ID)
User-Generated Content	1 (call given number of customers - IY), 11 (get product feedback-IY)
Two-sided Market Place	27 (Document and tracking progress of team members - WT), 12 (Reaching key business milestones,like tax registration, incorporation, patents -WE), 9 (Number of customers aware of the product - IP)

5. Discussion

In this section, we discuss the results in regard to our main research questions RQ1, RQ2, and RQ3.

5.1. RQ1: What are software start-ups in East Africa (a nascent ecosystem) assessing and measuring?

Our results indicate that even early stage start-ups in East Africa see value in measuring. Some are experiencing and /or expecting benefits. However, we did not find any formal measurement programs in these start-ups.

Business metrics Our results showed that business metrics were the most popular within the start-ups in East Africa. Of the business metrics, the customer analytic metric is the most used and wished for by the start-ups. It is an aggregated metric with a number of sub metrics that can be specific to a niche or particular business. One of the possible reasons why this metric is popular could be that it is widely supported by existing expensive and affordable tools. For example Facebook and Google offer affordable customer related analytic tools. The other could be that customer analytic metric also provides quantifiable values (like number of customers, number of active accounts, number of followers etc.) that start-ups may use in decision making or to get insight into their young businesses. Early stage start-ups are trying to get the product/market fit[22] right. This is one of the ways to prove that they are providing value to customers. For them to have confidence that they will commercialize, they

need to analyze customer related aspects of their businesses. Finally, the popularity of business metrics in general may be caused by the start-ups' need to become profitable in a short time.

Product metrics Product related metrics were the second most popular. This may not be surprising given the fact that most of the start-ups (13/19) are 2 years old and below. During the early stages, a number of start-ups are trying to develop the right product to fit the market (product/market fitting[22]).
However, it is interesting to see that these metrics are mostly in use by start-ups of the hub C, while start-ups from other hubs are rather wishing to use these metrics. This might hint at a potential hub C it's start-ups probably a common mentor.
The product/feature usage metric is a popular wished for metric by the start-ups (IN, IY, CE, OC). It is rather surprising that no start-up actually uses this metric. This could be because, the metric requires information from customers that the start-ups may not have in their early stages. This could also be the reason start-up CZ (11 months old) is using the feedback from friends about product feature (peer endorsement) metric instead. Also three of the four start-ups (IN, CE, OC) that wish to use this metric are below 2 years. It is harder to say why older start-ups are not interested in this metric. A probable reason could be a shift in focus, e.g. towards business aspects, favoring metrics such as customer analytic. In contrast to the product/feature usage metric, those product metrics that are in use appear to need low staff effort, require little or no tooling and are convenient. This might be a reason why start-ups opt for them.
Finally, system reliability is mentioned by RI the oldest start-up (51 months). A reason for that might be that this metric becomes only meaningful once there is a working deployed system. Again, we do not know why this start-up is not actually using the metric. However a possible reason could be that it is difficult to measure, requiring expertise that the start-up may not have.

Organizational Metrics Organizational metrics are the least mentioned category by the start-ups. Only 3 metrics (Key Performance Indicators (KPIs), time-based task setting, tracking and review for progress of project/staff and time-based project performance appraisal) were identified from the start-ups. These metrics are used by 4 start-ups, two relatively young start-ups ID (12 months), WT (10 months) and two older start-ups IP (32 months), NF (43 months) and wished for by another older start-up RI (51 months).
The motivation for the older start-ups could be that they have started to look at bigger solutions other than the Minimal Viable Product (MVP)[21] [12]. They are thus managing their solutions as projects with project teams unlike the younger start-ups. Start-up ID, is a younger start-up using metrics in this category. However, this start-up has some experienced partners, as clarified during the interview:

> *"One of my partners is an ardent investor, he has made past mistakes in investments, raising money and also setting KPIs making priorities focus and brings such expertise in the team [...] so we have a team of advisers one of them is the C.E.O of company x"*

The youngest start-up in this metrics category (WT), had just successfully gone through its funding round. Its preparation for funding may be the reason to use these organizational metrics so they could convince the investors. It would be interesting to study the role of investors when it comes to metrics that start-ups adopt in future research.

Project metrics Project metrics are mainly wished for by (4/6) start-ups. The most popular used metric in this category is documenting and reviewing activities for progress.It is used by three start-ups (IS, WE, WT). When comparing used and wished for metrics in this category, it can be seen that used metrics can be assessed manually and without special tooling. In contrast, the wished for metrics, such as process adherence, seem more difficult to assess,

Design metrics Our results showed that there are no metrics targeted at product design by any start-up. This could indicate a (1) lack of knowledge on what to measure during design or (2) the lack of a formal measurement program (like MeSRAM) that explicitly states what aspects of a product design to measure. It is alternatively possible that the start-ups are aware of such metrics, but consider their code-base to be too small to have a reason for such metrics.

Perceived Sufficiency The results showed that four start-ups NF (43 months), IN(15 months), IP (32 months) and IS (39 months) perceive their measuring to be enough. It is interesting to observe that three of the four start-ups (NF, IP, IS), are over two years old.
Three of these start-ups (IN, IP, IS) come from the same hub (I). However they are using different measurements, possibly, because they operate different start-up types: IN is a Fintech, IP is a two-sided marketplace, IS is a Software as a Service(SaaS). Thus, we see again a hint that hubs are having an impact on what start-ups measure. However, the general trend shows that start-ups would like to have better measurement programs. This is visible in the high number of start-ups reporting that they don't perceive their measurement as sufficient as well as the high number of wished for measurements reported.

Future Research Finally, it would be interesting to know why start-ups decide the metrics they use. The cross-analysis could not prove a relation of metric choice and factors such as hub-type or country. We suspect that there are more external factors involved, such as the start-ups relationship with other entities, e.g. mother companies or investors, or also factors such as costs (some metrics are more expensive to implement than others). We plan to address this question in future work.

5.2. RQ2: What are perceived benefits of measurements to start-ups?

Results have shown that the majority of the start-ups have not experienced real benefits of measuring but have expectations of benefits in case they measure enough. This is in tandem with the fact that although all the start-ups are measuring, 84% of the start-ups perceive their measuring as not enough or some are not even sure of measuring. This compares well with 82% of the 22 benefits that are just expected by the start-ups rather than experienced. Three start-ups (IS, NF, CZ) have experienced the benefits of measuring. For the two older start-ups IS (39 months) and NF (43 months), we might expect that because they had enough time to use the measurements and also experience their effects. It is also possible that these companies already learned what measures work best for them. Both companies also consider themselves to measure enough.

CZ (11 months) is young but experiencing some benefits of measuring. This could be because it is measuring actively the important aspect of its business. This enables it to make decision or gives it insight, as exemplified by the following quote:

"For a given month everyone is given tasks and we evaluate these tasks and if by the end of the month the tasks haven't been completed then we realize we are stuck."[CZ]

It has been shown that start-ups have considerable hope in measuring since the majority expressed their expectations from measuring. There seems to be a relationship between perceived sufficiency of measuring and expected and experienced benefits in start-ups. Start-ups that are not measuring enough or not measuring at all may not experience the benefits of measuring although they may have expectations of the positive benefits of measuring. Start-ups have shown hope in the benefits of measuring and they generally appreciate that they are not measuring enough.

One interesting question is why start-ups are not experiencing more benefits of measuring. The cross-case analysis revealed no significant differences in the characteristics of start-ups that experience benefits and the ones that don't. It is possible that the lack of benefits is due to ill-fitting measurements. However, it is also possible that some start-ups are applying good measures, but do not have the capacity to react to the measurement results. Future work will have to further investigate the reasons for that.

5.3. RQ3: How do metrics used in East African start-ups compare to those in lean start-ups in developed countries?

Start-up Categories   The nascent East Africa start-up ecosystem has all types of the start-ups found in lean start-ups summarized by Croll and Yoskovitz [7]. Software as a service is a popular start-up model in the East African ecosystem. This could be because of its ease of scalability and its potential to be self sustain (fewer overheads) after the initial investment.

We observed that not all the start-ups interviewed in East Africa could be categorized. This limitation in the lean start-up categorization may indicate the need for a wider categorization of the start-ups in East Africa, such as fintech. It also directly indicates the need to investigate metrics that are important for these types of start-ups in the nascent East Africa ecosystem.

Metrics Taken together, the East Africa start-ups used or wished to use at least half of the known lean start-up metrics in each category. However, single start-ups have a much lower coverage of those metrics. The choice of the lean metrics to use may be influenced by the start-up growth stage, as our earlier results showed that most of these interviewed start-ups are 2 years and younger. Another reason for the low coverage is that start-ups may not know all the important metrics to capture for their particular type of start-up.

Furthermore, the start-ups are also using or wishing for some unique metrics which have not been listed by Croll and Yoskovitz [7]. Future work will have to show whether this is due to the East Africa ecosystem or due to the relatively young age of start-ups interviewed in this study.

Summary Our results indicate the need for additional lean start-up categories that include fintech, software development, and software as part of work-flow. These new proposed categories mean that more lean start-up metrics need to be investigated to enable these types of start-ups to measure.

The fact that we found new metrics when interviewing the start-ups in East Africa may be an indicator of contextual influences on the start-ups. This indicates the need for more contextualized research on lean metrics or derivation of more inclusive metrics for start-ups in East Africa.

5.4.   Cross-Case Analysis

The cross-case analysis delivered surprisingly few results. It is very interesting to see that variables, such as start-ups age, country, or hub-type have no significant impact on the used metrics or experienced benefits of measurement. Especially, we would have expected an increase of metrics use with age, since large companies use much more metrics than the start-ups. There are alternative possible explanation. One is that our data-set is simply too small to prove such effects. An alternative interpretation could be that difference in metric use are visible only after much longer time, as opposed to the 1 to 3 years of age difference in the studied start-ups. It is possible that the differences are not in the number of metrics used or benefits experienced, but in the type of metrics used/benefits experienced. Future studies will have to investigate this possibility more in detail.

Finally, the found significant result of $H0_{5expected}$ seems logical, as benefits that are already experienced are not counted as expected anymore. Again, future studies will need to confirm or refute this finding.

5.5. Threats to Validity

For the discussion of the study validity, we follow a classification scheme used by Runeson [24] and Yin [29].

Construct validity   The main threat to construct validity is the risk that we have asked the wrong questions to assess what is really measured. To mitigate this threat we created the interview guide in several iterations to make sure that all three authors had the same understanding of the questions. Furthermore, we clarified and merged some questions from the interview guide after interviewing the first four start-ups (RD, RK, RW, and R1). To further increase the chances to get a complete picture of the measurements used, the interviewer used reformulated and repeated questions when the interviewee indicated that they didn't fully understand a question, or when the discussion deviated from the original topic.

Internal validity: A common risk during the analysis of data is a misinterpretation of what the interviewees said. We recorded and transcribed all the interviews and listened to them again in cases of doubt. In addition, we iterated together over the used codes and observations to make sure all measures were identified. These mitigation steps turned out crucial, as we initially omitted many of the unusual measures as they were only mentioned by interviewees, when asked how they assess their progress, but not when asked for what they measure.

External validity   Regarding external validity, our data has some limitations. First of all we studied start-ups in older than 6 months and younger than 4,5 years. While we believe that start-ups which are slightly younger or older than that might have a similar use of metrics, we do not expect the metric used to be the same across all company ages and maturities. Similarly our study focuses on a specific emerging ecosystem, namely East Africa. It is difficult to predict to what degree our findings can hold for software start-ups outside this regional context, e.g. in Europe.
A final threat to generalizability stems from the fact that we only interviewed start-ups that are working within hubs. However, to the best of our knowledge, there are only few early stage software start-ups in East Africa that are working independent of hubs. Therefore, we think that the 19 interviewed early stage start-ups allow a representative assessment of measurement in software start-ups in East Africa.

6. Conclusions and Future work

Early stage software start-ups in the nascent East Africa ecosystem are measuring some of their technical and business aspects. Using the MeSRAM[26] groups for "metrics used", the start-ups have been seen to preferably use business and product metrics and to a less extent organizational performance metrics. There has been no evidence of use or wish for design metrics (which exist in large software organizations)

in these start-ups. The older start-ups have been shown to also wish for project-oriented metrics. Organizational metrics are the least used metrics in the studied start-ups.

Start-ups have also shown considerable expectations in the benefits of measuring although a number of them perceive their measuring activities as not enough. A number start-ups in East Africa can be categorized using the lean start-up framework[7]. They are also using or wishing to use some of the known lean start-ups metrics, but there are also metrics that are not captured by the lean start-up framework. The lean analytic framework also fails to categorize some types of software start-ups in East Africa, although they exist and use or wish to use some metrics.

In the future, we will like to investigate the mapping between the used and wished for metrics and specific known growth stages of early start-ups. We would also like to examine, the suitability of these metrics in measuring certain aspects of a start-up. Using the metrics is one thing, but having a quality metric is another. We will therefore want to investigate the various properties of these metrics so that they can be qualitatively compared to metrics known elsewhere. We will also want to ascertain the extent of use of metrics within start-ups. Finally, we will like to propose an extension to the Lean Analytic framework, that will incorporate the currently unclassified start-up types in East Africa and common metrics they must measure as they grow.

Acknowledgment

References

[1] Abran A. *Software metrics and software metrology.* John Wiley & Sons, 2010.

[2] Albrecht A. J. and Gaffney J. E. Software function, source lines of code, and development effort prediction: a software science validation. *IEEE transactions on software engineering*, (6):639–648, 1983.

[3] Bajwa S. S., Gencel C., and Abrahamsson P. Software product size measurement methods. In *Proceedings of International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement, IWSM-Mensura*, volume 14, pages 176–190.

[4] Bhattacharyya G. K. and Johnson R. A. Statistical methods and concepts, 1977.

[5] Chorev S. and Anderson A. R. Success in israeli high-tech start-ups; critical factors and process. *Technovation*, 26(2):162–174, 2006.

[6] Comer P. and Chard J. A measurement maturity model. *Software Quality Journal*, 2(4):277–289, 1993.

[7] Croll A. and Yoskovitz B. *Lean analytics: Use data to build a better startup faster.* ” O’Reilly Media, Inc.”, 2013.

[8] Crowne M. Why software product startups fail and what to do about it. evolution of software product development in startup companies. In *Engineering Management Conference, 2002. IEMC’02. 2002 IEEE International*, volume 1, pages 338–343. IEEE, 2002.

[9] Cukier D., Kon F., and Krueger N. Designing a maturity model for software startup ecosystems. In *International Conference on Product-Focused Software Process Improvement*, pages 600–606. Springer, 2015.

[10] Daskalantonakis M. K., Yacobellis R. H., and Basili V. R. A method for assessing software measurement technology. *Quality Engineering*, 3(1):27–40, 1990.

[11] Díaz-Ley M., García F., and Piattini M. Mis-pyme software measurement capability maturity model–supporting the definition of software measurement programs and capability determination. *Advances in Engineering Software*, 41(10-11):1223–1237, 2010.

[12] Duc A. N. and Abrahamsson P. Minimum viable product or multiple facet product? the role of mvp in software startups. In *International Conference on Agile Software Development*, pages 118–130. Springer, 2016.

[13] Fenton N. and Bieman J. *Software metrics: a rigorous and practical approach.* CRC press, 2014.

[14] Giardino C., Wang X., and Abrahamsson P. Why early-stage software startups fail: a behavioral framework. In *International Conference of Software Business*, pages 27–41. Springer, 2014.

[15] Kitchenham B. and Mendes E. Software productivity measurement using multiple size measures. *IEEE Transactions on Software Engineering*, 30(12):1023–1035, 2004.

[16] Klotins E., Unterkalmsteiner M., and Gorschek T. Software engineering knowledge areas in startup companies: a mapping study. In *International Conference of Software Business*, pages 245–257. Springer, 2015.

[17] Paranjape B., Rossiter M., and Pantano V. Performance measurement systems: successes, failures and future–a review. *Measuring Business Excellence*, 10(3):4–14, 2006.

[18] Park R. E. Software size measurement: A framework for counting source statements. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST, 1992.

[19] Paternoster N., Giardino C., Unterkalmsteiner M., Gorschek T., and Abrahamsson P. Software development in startup companies: A systematic mapping study. *Information and Software Technology*, 56(10):1200–1218, 2014.

[20] Quinn Patton M. Qualitative research and evaluation methods, 2002.

[21] Ries E. Minimum viable product: a guide. *Startup lessons learned*, 2009.

[22] Ries E. *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Books, 2011.

[23] Rompho N. Operational performance measures for startups. *Measuring Business Excellence*, 22(1):31–41, 2018.

[24] Runeson P. and Höst M. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2):131, Dec 2008.

[25] Shi Y., Xu D., and Vessey I. Early-stage software start-up survival: the effects of managerial actions on firm performance. In *International Conference on HCI in Business*, pages 761–771. Springer, 2015.

[26] Staron M. and Meding W. Mesram–a method for assessing robustness of measurement programs in large software development organizations and its industrial evaluation. *Journal of Systems and Software*, 113:76–100, 2016.

[27] Sutton S. M. The role of process in software start-up. *IEEE Software*, 17(4):33–39, 2000.

[28] Unterkalmsteiner M., Abrahamsson P., Wang X., Nguyen-Duc A., Shah S., Bajwa S. S., Baltes G. H., Conboy K., Cullina E., Dennehy D., et al. Software startups–a research agenda. *e-Informatica Software Engineering Journal*, 10(1), 2016.

[29] Yin R. *Case Study Research: Design and Methods*. Applied Social Research Methods. SAGE Publications, 2003.

[30] Yin R. K. *Applications of case study research*. Sage, 2011.

[31] Yin R. K. *Case study research and applications: Design and methods*. Sage publications, 2017.