

Development and pilot testing of a maturity assessment instrument for testbeds in health care

Eva Collanus¹, Emilia Kiello-Viljamaa², Janne Lahtiranta¹, Antti Tuomisto¹

¹ Work Informatics, School of Economics, University of Turku, Turku, Finland; ² Faculty of Health and Welfare, Novia University of Applied Sciences, Turku, Finland

Emilia Kiello-Viljamaa, PhD, RN, Senior Lecturer, Faculty of Health and Welfare, Novia University of Applied Sciences, Henrikinkatu 7, FI-20500, Turku, Finland. Email: emilia.kiello-viljamaa@novia.fi

Abstract

High-quality testbed services play a key role in the development and testing of new innovations for health care. To ensure that the development and testing process is valid and reliable, the testbed itself should fulfil a certain maturity level. The objectives of this study were to develop and pilot a new maturity assessment instrument for health care testbeds and to assess the maturity level of participating testbeds using self-assessment auditing. The data were collected in 2022 from 11 health care testbeds in Finland using the developed Health Care Testbed Maturity Assessment (HCTM) instrument in the form of auditing. The data were analysed using descriptive statistical analysis. According to the study results, the average maturity level of the participating testbeds was 51%. The developed instrument indicated preliminary feasibility for assessing the maturity of health care testbeds. Further development and validation of the instrument are warranted.

Keywords: health care, maturity, testbed, instrument development

Introduction

During the last decade, testbeds have received a lot of attention in public health care – especially in the Nordic countries. The primary reason for this is that testbeds are regarded as a “tool of trade” for conducting development activities with the business sector. Formally, testbeds are physical or virtual environments where companies (or other organisations) can test, develop and validate function of their solutions [1–3]. In the literature, testbeds are also known as “living labs” or “Test Before Invest (TBI) facilities”; however, the common factor in all these characterizations is the co-creative innovation process that creates value for end users [4].

The use and development of solutions that are intended to be used in the field of healthcare, for example in patient care, is regulated [5]. The most mature healthcare testbeds can support this kind of development, by offering facilities, personnel, and other resources, that meet the stringent needs of regulation [1,3]. It follows from this, that maturity can be defined as an indication on how well defined and controlled the operations of a testbed are, and how well these operations support the conformity demands born from regulation.

As such, maturity of a testbed, does not refer to time in market or established position in it as the term could suggest. Maturity is a more profound

Published under a CC BY 4.0 license (<https://creativecommons.org/licenses/by/4.0/>).

measure of operations, that takes account resourcing, facilities, etc. In this, it is an indication of quality as well – something that can't be taken at a face value. For example, the mere location of a healthcare testbed (such as, within hospital or university campus) does not automatically imply that they share resources, or that the offered testbed services are of high quality.

The maturity assessment of testbeds is a prevalent need in the field of health care. The actual testbed service providers are diverse, and they may represent different organisations, such as hospitals or higher education institutes [6]. The testbeds, and the associated services themselves, may also vary significantly, as they may be real, simulated or laboratory-based [7]. In a situation as diverse as this, the maturity of testbeds can and do vary significantly, as the providers' operations and resources differ from each other [6].

Other aspects that underline the variety of testbed services from the viewpoint of maturity are lack of a) education, and b) related education. In Finland, testbed operations are not regulated per se. In principle, anyone can set up a healthcare testbed, and offer their services to companies if the offered services are consumer studies, or in similar vein.

In the development of the maturity instrument previous work, and experiences from other fields of operation, has provided a valuable base to build on. Findings from these, such as focus on facilities, resources, certification and standards compliance [6,8], has been recorded, and considered.

Background

Existing maturity assessment instruments or other assessment methods or models for health care testbeds reported in previous studies are scarce. Santonen et al. [9] proposed a maturity model that

defines the degree of collaboration maturity between living labs using four levels (Network, Coordinated network, Co-operation and Collaboration) and six factors (Environment, Membership, Structure-process, Communication, Purpose: Goal-Vision and Resources). Veeckman et al. [10] assessed living lab characteristics and outcomes using a conceptual framework consisting of 12 building blocks, each assessed using a four-point scale. Jukić et al. [11] created a model to assess organizational maturity for co-creation in public organizations, including health care. Osorio et al. [12] designed a tool to assess the maturity of innovation laboratories representing different domains, including Health and e-health.

Of these studies, the study by Santonen et al. [9] was the only one that was conducted in the health care domain exclusively. The other studies examined maturity in multiple domains but included the health care domain at some level. Most of the studies presented a maturity model or framework, but the study by Osorio et al. [12] presented a further developed tool to assess maturity. However, the tool is generic and does not consider characteristics and needs of health care sector, and it focuses on innovation laboratories, which are more focused on the earlier stages of new innovations, such as brainstorming and visualising of new innovations compared to testbeds. Furthermore, the tool is only operationalised based on a framework but not validated as a tool. Based on this literature search, no instruments developed and validated particularly to assess the maturity of testbeds in the health care domain were found.

The objectives of this study were to develop and pilot the maturity assessment instrument and to assess the maturity level of participating health care testbeds using self-assessment. The goal of this study is to develop an instrument that can be used

to assess the maturity of testbeds in health care. The research questions for this study are: 1) How valid and usable is the developed maturity assessment instrument based on the pilot testing? and 2) What is the self-assessed maturity level of participating health care testbeds?

Materials and methods

Design and setting

The study was divided into two phases: 1) instrument development and 2) testing. The setting of the study was testbed service provider organisations in Finland which offer testbed services in health care settings. The testbed service providers (Figure 1) in this study represented a) public health service providers, b) higher education institutions (universities and universities of applied sciences), and c) research organizations.

Instrument development

The instrument development process (Figure 2) followed the steps set out by DeVellis [13]: item generation, face validity, content validity with expert panels and testing. Item generation focused initially on factors that describe the essential characteristic

of the subject [14] or, in this case, the maturity of the testbed under evaluation. In generation, the starting point was the seven previously identified maturity factors: resources, facilities, marketing and communications, repeatability, contract models, certification and standards compliance, and time at the market area [6]. The final factors were formulated based on the generated items to better describe the content of each section of the instrument. The maturity factors (n=7) of the instrument were formulated as: operations (7 items), ecosystems (3 items + 3 specifying items), facilities (3 items), personnel (4 items + 2 specifying items), funding (3 items), quality (7 items + 5 specifying items) and marketing and dissemination (5 items + 1 specifying item). The rating of the instrument was based on the testbed organisation's self-assessment of its operations. The items were multiple-choice or yes/no items. Certain items also had specified items if the respondent had answered "yes" to the main item. The scoring for each main item ranged from 0 to 3. In specifying items, the scores ranged from 0.5 to 6. This means that the main questions were worth 0-3 points and if the question contained specifying questions, these could be worth an additional 0.5-6 points per answer.

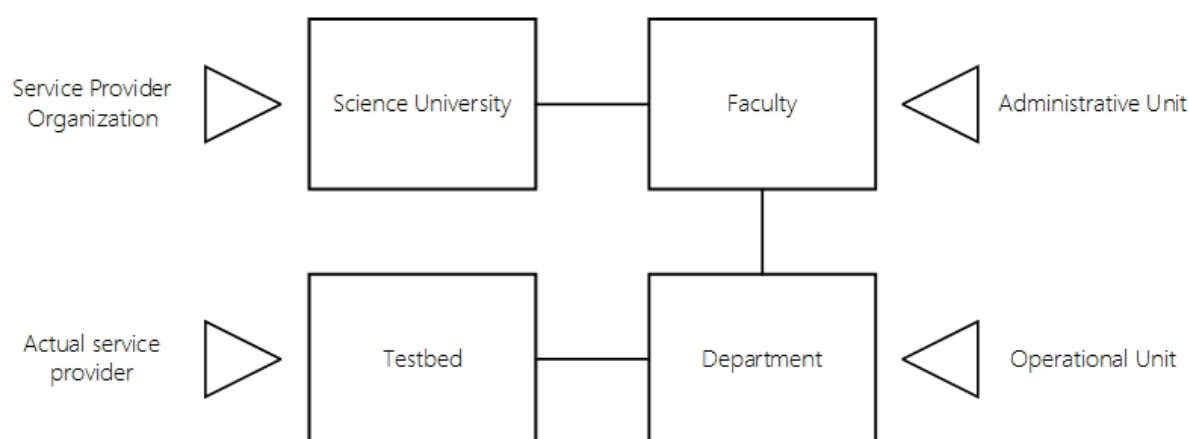


Figure 1. Service alignment, example.

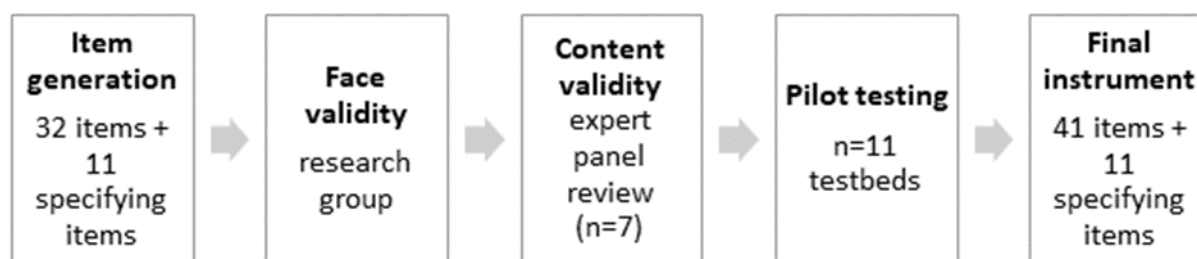


Figure 2. Instrument development process.

The structure and content of the instrument was first assessed by the research group and then by an external expert panel (n=7) in December 2021. Expert panel members' experience in testbed and innovation operations varied from 2 to 10 years (mean 5.6 years). The expert panel review included the clarity, importance and relevance evaluation for each instrument item using a four-point scale: e.g., 1=not at all relevant, 2=not very relevant, 3=quite relevant, 4=highly relevant. The experts could also comment on the content and suggest revision or other content or items that could be added to the instrument. The experts gave their reviews individually using an online platform. The content validity index (CVI) was measured for each item (I-CVI) by calculating the sum of ratings 3 and 4 divided by the number of experts. For each section (Average I-CVI), the CVI ranged from 0.71 to 0.93. The CVI for the whole instrument (S-CVI) was 0.85, meaning that the CVI can be considered acceptable [15,16]. Clarifications and some modifications were made to the items on the basis of the expert panel review, especially for those items with a lower I-CVI, e.g. for the language or response options

Data collection

The pilot testing of the instrument consisted of 11 testbeds. The testbeds were recruited through contacts and networks of researchers, such as the

national health care testbed network to which most Finnish health care testbed organisations belong. The testbed provider organisations could represent any sector (e.g. education or health and social services). The pilot testing was conducted as a form of friendly auditing, meaning that the testbed representatives self-assessed their testbed's status. The audits were facilitated remotely by the research team, and one or two representatives from each testbed took part in the auditing.

The instrument went through some improvements during the pilot testing process. Some clarifications were made to the wording, some new answer options were added to the items, one new item was added to the quality factor, and a totally new factor, sustainable development, with eight new items, was added. At the end of the pilot testing, the instrument had eight factors: operations (7 items), ecosystems (3 items + 3 specifying items), facilities (3 items), personnel (4 items + 2 specifying items), funding (3 items), quality (8 items + 5 specifying items), marketing and dissemination (5 items + 1 specifying item), and sustainable development (8 items) (Table 1). The maximum score for the instrument was 120 (operations 22, ecosystems 20, facilities 7, personnel 8, funding 4, quality 41, marketing and dissemination 8, sustainable development 10). The instrument was named Health Care Testbed Maturity Assessment Instrument (HCTM).

Table 1. Examples of the instrument items.

Factor	Item
Operations	Who is the primary customer of the testbed? I. Companies (2p) II. Public sector (excluding research organisations and universities) (2p) III. Third sector (incl. associations) (1p) IV. Research organisations, science agencies and universities (1p) V. Other, specify (1p)
Ecosystems	Are the testbed services part of an innovation ecosystem? I. No (0p) II. Yes. You may choose more than one ecosystem type. (1p) a. The services are part of a regional ecosystem [example] (1p) b. The services are part of a national ecosystem [example] (2p) c. The services are part of an international ecosystem [example] (3p)
Facilities	Are the testbed services provided in a real-world clinical environment, such as in a hospital ward or a nursing home? I. No (0p) II. Yes (3p)
Personnel	Can the testbed use actual clinicians or patients in its services (e.g. in usability testing or concept validation)? I. No (0p) II. Yes (1p)
Funding	How are the services priced as a rule for the primary customer? I. Market-based pricing (2p) II. Original cost pricing (prime cost) (2p) III. Subsidized from other sources, such as from projects (1p) IV. The services are offered freely (0p)
Quality	Is the operation of the testbed formally evaluated by an independent (certification) agency? I. No (0p) II. Yes (3p)
Marketing and Dissemination	Which marketing and communication channels are used frequently by the testbed? I. Web pages (testbed or testbed service provider organisation) (1p) II. Social media, blogs postings, videos or similar (1p) III. Traditional media (TV, radio, newspapers) (1p) IV. None of the above (0p)
Sustainable Development	Are the longevity and life cycle (recycling, ecological footprint, etc.) of the materials factors in the testbed's or testbed service provider organisation's procurement practices? I. No (0p) II. Yes (1p)

p=points

Data analysis

The data were analysed using descriptive statistics, including means, medians and minimum and maximum scores for each testbed and each factor of the instrument. As the sustainable development factor was added to the instrument at the end of the pilot testing, a maximum score of 110 was used in the analysis. The maturity levels were described as scores, and the testbed service provider organisations' sectors (education/health care/other) were compared. As the sample for this pilot testing was only 11 testbeds, no inferential statistics were used to analyse the data. Cut-off scores for low (<50%), good (<67%) and excellent (>67%) maturity levels were set for each instrument factor based on the data. Maturity levels were presented as a percentage because the scoring in each item and factor varied significantly.

Ethical considerations

This study followed the guidelines for the responsible conduct of research according to the Finnish National Board on Research Integrity [17]. Participation in this study was voluntary, and participants could withdraw from the study at any point without giving a reason. Personal data was collected and handled according to the European Union General Data Protection Regulation [18]. Written informed consent was requested from every study participant prior to data collection.

Results

Testbed organisations

In total, 11 testbeds from Finland participated in the pilot testing of the instrument. All the testbeds were health care testbeds, but the sector in which the testbed service provider organisation operated varied. Of these 11 organisations, six represented

the education sector, and all of them were higher education organisations. Three organisations represented health or social services, two of which represented specialized care and one primary care. The two remaining organisations represented other sectors, one was a research institution and the other was a regional development organisation. In terms of size (personnel), the testbed service provider organisations varied. One organisation had 10–49 employees, six organisations had 100–999, two organisations had 1000–4999, and two organisations had more than 5000 employees. The organisations represented different geographical areas in Finland, but the areas are not specified here because the testbeds could be identified.

Maturity of the testbeds

The average overall maturity level of the participating testbeds was 51% (56/110), meaning that the average maturity level of all testbeds was good (>50%). The highest average overall maturity was in those testbeds where the service provider organisation represented sectors other than health or social services or education (55%). The second highest overall maturity levels were in those testbeds where the organisation represented the education sector (51%), and the lowest were in those testbeds where the organisation represented health or social services (46%).

The average maturity level in operations was 73% (16/22). The highest average maturity levels were in testbeds where the service provider organisations represented the education sector (77%), and the lowest average maturity levels were in testbeds where the organisations represented other sectors (59%). The average maturity levels were excellent (>67%) in all testbeds except where the organisation represented sectors other than health or social services, or education. (Table 2)

In ecosystems, the average maturity level was 60% (12/20). The highest average maturity levels were in testbeds in health or social services organisations (65%), and the lowest were in service provider organisations representing other sectors (51%). The average maturity levels were good (>50%) in all testbed service provider organisation sectors. One testbed received full scores in this section (see Table 2).

In the facilities factor, the average maturity level within testbeds was 57% (4/7). The highest average maturity level was in testbeds where the service provider organisations represented sectors other than health or social services or education (71%). There was no difference in average maturity levels between the health or social services and education sectors. Three testbeds received full scores in this factor (see Table 2).

In the personnel factor, the average maturity level of testbeds was 75% (6/8), which was also the average maturity level of all testbed service provider sectors other than the health or social services sector, where it was 63%. In this factor, one testbed scored 100% maturity level (see Table 2).

The average maturity level in the funding factor was 50% (2/4). The highest average maturity levels were in testbeds where the testbed service provider organisations represented the education sector

(75%). Three testbeds received the full scores in this factor, and one received zero points (see Table 2).

In the largest instrument factor, quality, the average maturity level of the participating testbeds was 27% (11/41). In this factor, the testbeds where the service provider organisations represented sectors other than health or social services and education, had the highest average maturity levels (46%), and their maturity levels were relatively higher compared to the other sectors. In this factor, there was a lot of variation in the maturity levels between the testbeds, which can be partly explained by the modifications of this factor during the pilot testing process, and that one new item with one specifying item was added during the process (see Table 2).

In the marketing and dissemination factor, the average maturity level was 50% (4/8). The highest average maturity levels were in testbeds where the service provider organisation represented the education sector or other sectors (63%) (see Table 2).

Only one testbed answered the items related to the sustainable development of the testbed, as this factor was added to the instrument at the end of the pilot testing based on the feedback during the data collection. This testbed's service provider organisation represented the education sector, and the average maturity level in this factor of the instrument was 80% (8/10) (see Table 2).

Table 2. Average maturity levels of testbeds.

Maturity factor	Mean %/score	Median score	Min %/score	Max %/score
Operations (7 items: min 3 – max 22)				
All sectors	73/16	15	59/13	86/19
Health care sector	73/16	15	64/14	82/18
Educational sector	77/17	17	68/15	86/19
Other sectors	59/13	13	59/13	53/13
Ecosystems (3+3 items: min 0 – max 20)				
All sectors	60/12	11	40/8	100/20
Health care sector	65/13	10,5	40/8	100/20
Educational sector	58/11,5	11,5	48/9,5	68/13,5
Other sectors	51/10,25	10,25	40/8	63/12,5
Facilities (3 items, min 0 – max 7)				
All sectors	57/4	4	14/1	100/7
Health care sector	57/4	4	43/3	57/4
Educational sector	57/4	4	14/1	100/7
Other sectors	71/5	5	43/3	100/7
Personnel (6+2 items, min 0 – max 8)				
All sectors	75/6	6	38/3	100/8
Health care sector	63/5	6	38/3	88/7
Educational sector	75/6	6	38/3	100/8
Other sectors	75/6	6	75/6	75/6
Funding (3 items, min 0 – max 4)				
All sectors	50/2	3	0/0	100/4
Health care sector	50/2	3	0/0	100/4
Educational sector	75/3	2,5	25/1	100/4
Other sectors	50/2	2	25/1	75/3
Quality (8+5 items, min 0 – max 41)				
All sectors	27/11	7,5	7/3	51/21
Health care sector	22/9	7,5	12/5	33/13,5
Educational sector	24/10	5,75	7/3	45/18,5
Other sectors	46/19	18,75	41/17	50/41
Marketing and dissemination (5+1 items, min 0 – max 8)				
All sectors	50/4	4	13/1	88/7
Health care sector	38/3	2	13/1	63/5
Educational sector	63/5	4,5	50/4	75/6
Other sectors	63/5	5	38/3	88/7
Sustainable development (8 items, min 0 – max 10).				
Note: only one case included.				
All sectors	NA	NA	NA	NA
Health care sector	NA	NA	NA	NA
Educational sector	80/8	8	80/8	80/8
Other sectors	NA	NA	NA	NA

Discussion

Discussion of the results

The objectives of this study were twofold. First, to develop and test the maturity assessment instrument. Second, to assess the maturity of health care testbeds in Finland, using the developed HCTM instrument. The development process of the instrument consisted of five phases: 1) instrument generation, 2) face validity assessment, 3) content validity assessment, 4) pilot testing, and 5) instrument finalization (c.f. Figure 2).

In general, the highest average maturity levels of the participating testbeds were related to personnel. This indicates that the health care testbeds in Finland are mostly well-staffed, even though the personnel often worked in temporary positions (e.g., in projects). However, the size (personnel) of the testbed provider organization was no guarantee of more a mature testbed if the personnel was not allocated to the testbed.

In relation to the personnel factor, individual questions were about the roles, responsibilities and level of resourcing (part-time or full-time). Most of the testbeds (10 of 11) had a person in charge, but in most cases, supervising a testbed was a part-time role. In all 11 testbeds, actual clinicians or patients could be recruited for testing purposes, and all testbeds reported that use of complementary personnel (researchers, students, etc.) was an option as well.

In our opinion, high maturity in this factor can be explained with a) simple and ‘flat’ organisational structures (as is common in Scandinavia [19]) high commitment by the service provider organisations, and b) active collaboration with local ecosystems, that supports co-operation, and gives testbeds

some operational flexibility (e.g. in terms of resourcing, marketing, etc.).

The second highest average maturity levels in the pilot testing were related to operations (73%). In this factor, both education and health or social service provider organisations scored excellent maturity levels. In this factor, there were questions, amongst others, about the intended customer sector (who the primary customer of the testbed is), nature of service provisioning (are the testbed services related to the primary operation of the service provider organisation) and time in the market.

The primary customers in most cases were companies, and those testbeds typically scored highest in this factor. However, there was a significant amount of variation in the testbeds’ expertise, in other words, the number of served customers and the time in market varied significantly. Due to this variation, the scale in the question related to the number of served customers was edited during the pilot testing, and it is probable that the scale will be adjusted later, as the question did not segregate the testbeds very well. In the later adjustments, it is possible that the question needs to be co-factored with the type of service provider organisation, as a health or service provider’s testbed services seem to have a longer time to execution (i.e., from contract to the start of services), due to a more regulated environment.

Other factors that showed good average maturity levels (> 50%) were a) ecosystems, b) facilities, c) funding and d) marketing and dissemination. In ecosystems, testbed partnerships were evaluated; whether they belonged to a wider business ecosystem or to a testbed ecosystem that supported solicitation of customers. The ‘reach’ of the ecosystems was also analysed, and those that had international coverage were ranked higher. In this factor, regional and national ecosystems were

widely presented. Most of the testbeds were in a formal partnership with a region's health-related ecosystem and were linked to a service offering of a regional development organisation that offered different support services to the companies.

In facilities, testbeds that were related to the health or social services sectors, received the highest maturity scores. These testbeds offered real-world testing environments instead of simulated ones, and the differences in testing fidelity (i.e., match to real-world care) contributed towards higher scores. In the education sector, the simulated environments were understandably more prevalent, as the testbed facilities were used primarily for (undergraduate) teaching, and the testbed services themselves were built to serve the same purpose.

In Funding, the education sector received higher maturity scores compared to other sectors. This difference can be explained by the fact that industry collaboration in Finland is a core operation for higher education institutes (especially universities of applied sciences). In health care and social services, the core operations are focused on patient care, education and research [20], and public service providers have limited capabilities to offer testbed services on a commercial basis.

Similarly, in marketing and dissemination, the health or social services sectors scored lower than the education sector. The reasons for this are similar: there is no budget or resources for activities that fall outside the core operations.

Finally, in the largest instrument factor, quality, the average level of all testbeds was only 27%. The results demonstrate that sectors other than health or social services and education received higher maturity scores but still below 50%. In this factor, questions were about the evaluation of the testbed's operations, documentation (processes,

agreements, etc.) and customer interaction. In the testbeds where operations were evaluated, it was done by the testbed's personnel, not by a formal evaluator – or an external actor of any capacity. In one testbed, the operations were reportedly based on existing standards (ISO 9001); however, as the evaluation was conducted as a friendly audit, proof of this claim was not inspected.

The low maturity levels in this factor can be explained by the fact that most of the testbeds were new, and they have not yet had the possibility to create formal processes, or any kind of quality framework that could be used for steering their operations. On the other hand, low performance in this factor could also be traced back to the nature of the core operations of the service provider organisation. In Finland, the core operations of the provider are typically teaching, research, or patient care and co-creation with companies is of a lower priority [21].

Discussion of validity

The developed HCTM instrument went through face and content validity testing [22] before the pilot testing increasing the internal validity of the study. During the pilot testing process, the instrument was modified and further developed. The pilot testing was conducted with 11 testbeds, which can be considered as an appropriate sample for pilot testing of a new instrument. However, further testing, including psychometric testing [23], is needed to assess the instrument's construct validity, internal consistency and sensitivity. In addition, the pilot testing was based on a friendly audit method, meaning that the maturity assessment was mainly based on self-assessment, a subjective assessment. The disadvantage of subjective assessment is that testbeds may try to convince people, in this case researchers, of how mature their testbed is, instead of giving an honest assessment

of the maturity of their testbeds, which could also help them to develop their testbed. The testbeds may also think that they have existing processes and quality controls, for example, because it is obvious to them but not to people outside the organisation. In future, the instrument should be tested using formal audits, which could be considered a more objective way to assess the maturity of testbeds.

The variety of organisations helped develop the instrument during the pilot testing process to better answer the needs and features of different sectors. However, due to a small sample size, differences between the sectors could not be reliably addressed, and the data analysis was descriptive. Still, the service alignment structure (Figure 1) of most Finnish health care testbeds may increase the validity and usability of the instrument, as health care testbed activities are rarely organizations' core function. The testbeds for this study were recruited via the researchers' contacts and networks, but they represented different geographical areas in Finland, giving an overview of the testbeds' maturity on a national level. Thus, the study results cannot be generalised internationally, but further development and testing on a larger scale is

warranted, because the health care testbed maturity factors are common across countries.

The developed instrument had some similarities to the previously developed testbed maturity models and frameworks. For example, the factors of ecosystems, evaluation (quality), resources (personnel/funding) and environment (facilities) were also identified in two previous studies [9,10]. This indicates that the structure of the instrument is somewhat akin to other testbeds or living lab maturity assessment instruments and models. However, no instruments assessing the maturity of testbeds solely in the health care context were found, meaning that the developed instrument could not be compared to any other gold standard [24] instruments in its entirety.

Conflict of interest declaration

None declared.

Acknowledgements

This work was supported by the project Health Campus Turku 2.0 (337640), which is funded by the Academy of Finland.

References

- [1] Chernova I, Konina A. The report on health startups' experiences with testbeds, public procurement and regulations. Publications of the Ministry of Economic Affairs and Employment 2022:6. Ministry of Economic Affairs and Employment of Finland; 2022. p. 63. <http://urn.fi/URN:ISBN:978-952-327-598-0>
- [2] Haukipuro L, Väinämö S, Torvinen H. End-user involvement enhancing innovativeness in public procurement. Evidence from a healthcare procurement. Journal of Innovation Management. 2016;4(4):98–121. https://doi.org/10.24840/2183-0606_004.004_0007
- [3] Rosemberg C, Potau X, Leistner S, Dijkstal F, Vinnik A, Tiriduzzi C, et al. Regulatory Sandboxes and Innovation Testbeds: A Look at International Experience in Latin America and the Caribbean. Final report [Internet]. Inter-American Development Bank; 2020. <https://doi.org/10.18235/0002526>
- [4] Haukipuro L, Väinämö S, Hyrkäs P. Innovation Instruments to Co-Create Needs-Based Solutions in a Living Lab. Technology Innovation Management

- Review. 2018 May 29;8(5):22–35. <https://doi.org/10.22215/timreview/1156>
- [5] EUR-Lex. Regulation (EU) 2017/745 Of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. European Parliament, Council of the European Union; 2017. <http://data.europa.eu/eli/reg/2017/745/oj>
- [6] Kielo-Viljamaa E, Collanus E, Lahtiranta J, Tuomisto A. Maturity of Health Care Testbeds – A Qualitative Mapping at the Nordic Context. *FinJeHeW*. 2022 14(1):92–103. <https://doi.org/10.23996/fjhw.111734>
- [7] Jabin MSR, Nilsson E, Nilsson AL, Bergman P, Joke-la P. Digital Health Testbeds in Sweden: An exploratory study. *Digit Health*. 2022 Feb 14;8:20552076221075194. <https://doi.org/10.1177/20552076221075194>
- [8] Engels F, Wentland A, Pfotenhauer SM. Testing future societies? Developing a framework for test beds and living labs as instruments of innovation governance. *Research Policy*. 2019 Nov;48(9):103826. <https://doi.org/10.1016/j.respol.2019.103826>
- [9] Santonen T, Kjellson F, Andersson K, Hirvikoski T. Developing maturity model for transnational living lab collaboration. In: Bitran I, Conn S, Gernreich C, Heber M, Huizingh KRE, Kokshagina O, Torkkel M (Eds.). *Proceedings of the 2020 ISPIM Innovation Conference (Virtual) Event "Innovating in Times of Crisis"* held on 7 to 10 June 2020. International Society for Professional Innovation Management; 2020.
- [10] Veeckman C, Schuurman D, Leminen S, Westerlund M. Linking Living Lab Characteristics and Their Outcomes: Towards a Conceptual Framework. *Technology Innovation Management Review*. 2013 Dec 17;3(12):6–15. <https://doi.org/10.22215/timreview/748>
- [11] Jukić T, Pluchinotta I, Hržica R, Vrbek S. Organizational maturity for co-creation: Towards a multi-attribute decision support model for public organizations. *Government Information Quarterly*. 2022;39(1):101623. <https://doi.org/10.1016/j.giq.2021.101623>
- [12] Osorio F, Dupont L, Camargo M, Palominos P, Peña JI, Alfaro M. Design and management of innovation laboratories: Toward a performance assessment tool. *Creat Innov Manag*. 2019 Mar;28(1):82–100. <https://doi.org/10.1111/caim.12301>
- [13] DeVellis RF. *Scale Development: Theory and Applications*. 4th ed. Thousand Oaks, CA: Sage; 2017.
- [14] Rosemann M, de Bruin T. Towards a Business Process Management Maturity Model. In: *Proceedings of the 13th European Conference on Information Systems (ECIS) 2005*, 37. p. 521–32. <https://aisel.aisnet.org/ecis2005/37>
- [15] Grant JS, Davis LL. Selection and use of content experts for instrument development. *Res Nurs Health*. 1997 Jun;20(3):269–74. [https://doi.org/10.1002/\(SICI\)1098-240X\(199706\)20:3<269::AID-NUR9>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1098-240X(199706)20:3<269::AID-NUR9>3.0.CO;2-G)
- [16] Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*. 2007 Aug;30(4):459–67. <https://doi.org/10.1002/nur.20199>
- [17] TENK. Responsible Conduct of Research (RCR) [Internet]. TENK (Finnish National Board on Research Integrity); 2021 [cited 5 December 2022]. Available from: <https://tenk.fi/en/research-misconduct/responsible-conduct-research-rcr>

- [18] EUR-Lex. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). European Parliament, Council of the European Union; 2016. <http://data.europa.eu/eli/reg/2016/679/oj>
- [19] Andreasson U, Lundqvist M. Nordic Leadership. Analysis No. 02/208, Nordic Council of Ministers, 15. Copenhagen: Nordisk Ministerråd; 2018. <https://doi.org/10.6027/ANP2018-835>
- [20] Reunanen T, Kontio E, Lahtiranta J. Concurrent Research and Decentralized Decision Making as an Accelerator from Idea to Business – Case Turku Finland. In: Kantola JI, Nazir S, Salminen V, editors. Advances in Human Factors, Business Management and Leadership. AHFE 2020. Advances in Intelligent Systems and Computing, vol 1209. Cham: Springer International Publishing; 2020. p. 209–216. https://doi.org/10.1007/978-3-030-50791-6_26
- [21] Lahtiranta J, Reunanen T, Kontio E. Co-creation services boosting health technology potential – the case of Turku Finland. In: Markopoulos E, Goonetilleke RS, Luximon Y (eds). Creativity, Innovation and Entrepreneurship. AHFE (2022) International Conference. AHFE Open Access, vol 31. AHFE International, USA; 2022. <https://doi.org/10.54941/ahfe1001515>
- [22] DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A Psychometric Toolbox for Testing Validity and Reliability. J Nurs Scholarsh. 2007;39(2):155-64. <https://doi.org/10.1111/j.1547-5069.2007.00161.x>
- [23] Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A practical guide to their development and use. 5 edn. Oxford University Press; 2015. <https://doi.org/10.1093/med/9780199685219.001.0001>
- [24] Choudhary PK, Nagaraja HN. Selecting the instrument closest to a gold standard. Journal of Statistical Planning and Inference. 2005 Feb;129(1–2):229–37. <https://doi.org/10.1016/j.jspi.2004.06.049>