Appendix xx: Key themes and quotes considering the key directional decisions and challenges as expressed by the expert panel participants in the one-to-one interviews (N=55)

Frequencies reflect the number of participants who favored a specific direction.

EE = eHealth Expert
HCP = Healthcare professional
IE = Insurance Expert
In = Investor
PA = Patient Advocate
Ph = Pharma
RE = Regulatory Expert
Rs = Researcher
TP = Tech Provider

Universal versus contextual criteria		
	Favors contextual (n=55, 100%)	"I think it's very fair and it has to be very clear if this is going to be useful, that there is a specific distinction between assessing the inherent functionality of the specific digital health tool and solution, which I would assume mostly captured through the core criteria versus assessing how the solution engage with its context with the ecosystem that's going to adopt it" P9-EE-Rs "So, it makes sense because healthcare is local. I mean, there is a high variability of the practice of healthcare from one side to the other, even in the same city" P11-HCP "You always kind of-- you have to be going back and forth between these criteria. But at the same time, you're constrained by some of those contextual criteria. So, it will work maybe well in Germany, but it's not going to work well in France" P38-Ph "The contextual factors for me as well really resonated… and the contextual, I absolutely support because … it could make or break successful implementation and eventually sustainability and scale up. So, I absolutely like the idea of having a minimal set of items or criteria that address the contextual characteristics of where and how the tool will be implemented" P54-Rs
Single score versus scorecard		
	Favors scorecard (n= 43, 78%)	"it's challenging to come up with one size fits all. So, I think you're giving it an opportunity to tease that out to take those differences into consideration" P2-EE "I think just having a single score would be very one dimensional… And with the scorecard, so you're taking different perspective into account, which I think is very important" P5-EE "The value of an assessment tool is not just a score… It's understanding the breakdown" P6-EE-In "A tool that helps you design for quality and focus on the right quality metrics, then by definition, you want to know where you stand on each individual dimension so that you can pick the right dimensions and assess for quality" P10-EE "I like the scorecard idea, and I totally appreciate how it's super contextual, what things assessors will value … and where they're willing to give some leeway" P14-HCP-EE "I think the scorecard is giving more bigger image about the evaluation and the assessment… it's nearer to the reality. It can give us better idea" P19-HCP-Rs

		"Because you've got so many moving parts... And one part kind of drags the other one in one direction, but the other one drags it in the other… A single score, may not do justice" P18-HCP-EE "I think that this kind of a tool is valuable, especially in that it gives the understanding of the strengths and on the weak points of each solution… I'm strongly in favor of the scorecard view because this gives you the ability to zoom in on those items, then you know that this is something that I really need to develop in order to be more successful in this assessment" P27-In "The scorecard is probably a bit more suitable and provides a fairer evaluation because it takes into consideration the various criteria" P28-In-TP "I think it's a good idea to have the scorecard… because I think that there's so many different criteria, that the only way to do it is to have some kind of scoring system that you can see them very quickly, where you need to improve on" P36-PA-TP "A global number would not mean anything, so that doesn't make sense. And depending who's using the tool, obviously, there are different priorities… the scorecard shows that this assessment is comprehensive and it takes into consideration the voices of different parties, the customer, the doctors, and so on and so forth" P34-PA-TP "So single-score doesn't work because there is no single objective for two different products. So … you as a project owner, you will be able to identify what's your priority. So, because it's your priority, you will be able to choose what is important to you" P41-Ph-TP "The scorecard just gives you a broader mapping, right, than a single score is going. It's not perfect, but it definitely offers a better understanding" P42-RE "To me, the scorecard is the solution. And yes, there are tradeoffs, but you at least have higher chances to really elaborate on a more holistic view than just the single score" P46-TP "So, kind of some of these areas might be more relevant depending on a bit on the history and the context, so that's why at least highlighting those in the kind of scorecard format instead of single score might make actually more sense" P52-TP
	Favors a mix of scorecard and composite score (n= 12, 22%)	"I'm a big fan of the scorecard … the risk there is if you have a lot of solutions that are being assessed (comparability becomes challenging)" P49-TP-EE "That's like a bit of both. So, I'm used to using scores when they'll have a sub scales, and so it'll probably give me, for example, overall symptom burden, but when I break it down… I might have a score for the physical symptoms and then the more psychological symptoms. So, I think, ideally, there will be a lot of tools work as you get an overall score. But you get sub scales are valid in their own right. And I like a bit of both" P53-HCP-Rs "Adding an individual score is probably-- but it might give a quick reference for somebody who's comparing multiple solutions to sort of say: Okay, here's the overall score now. Let me go into depth and see where that score is coming from… if they're comparing several apps at the same time" P54-Rs
Proactive versus mass appraisal		
	Favors proactive appraisal (n= 36, 65%)	"I may be biased here with an investor hat on, but it's due diligence in my world. And we would never-- even in the first instance, we're sourcing information from the provider. And it's the only way to do it. You have to kind of lift the stones" P6-EE-In

		"I think you'll never get down to the point of really understanding what the app can offer if you don't talk to the developer. I think we've seen this firsthand in different cases" P7-EE "I think the point that the mass appraisal is really dependent on the tool promotion, more than the tool efficacy or the tool usefulness, and I think the approach to choose the proactive appraisal, I think it's the better way to assess or to appraise the tool" P14-HCP-Rs "I think also the other pieces when I think about it, because I'm also a clinician, I have to put hands on it to verify" P15-HCP-EE "I would agree to the focused appraisal because the more specific the testing is, the more information you get from it and the higher the quality of the information also is" P16-HCP-EE "I think the proactive is very important because that enables you to go beyond just the way it's marketed" P17-HCP-EE "I don't think you can come up with ratings without actually having tried it because, again, there's bias in how things are going to be presented, and you need to see for yourself" P20-HCP-Rs "I think that proactive, more hands-on approach is so important. Because there's also a lot that you can find online, and when you see the reality of the tool can be very different" P47-TP "They need to look at this proactively, gather the information, look at the evidence supporting it, underpinning what the app does, the intervention, but also the evidence of the app in terms of the claims that it's making" P53-HCP-Rs "I definitely see the value of the proactive. I can't think of any other way to not go the proactive route" P54-Rs
	Disfavors proactive appraisal (n= 4, 7%)	"50 criteria are already a lot to do to an assessment in the real world. If you have an academic company, or if you have a regulatory mean, I think then 50 is easy because the more accurate, but in, as you said, with budget and stuff like that, I'm not sure if 50 or now it looks like 33. And 33 is if you do them in depth, that's a lot" P24-IE "I mean, if for example, the solution doesn't give any data and then you have to do this by yourself, I'm not sure that some of the assessors will do that hands-on. That's a bit challenging" P50-TP-Ph
	Favors a mix of proactive and mass appraisal (n= 10, 18%)	"For me, the first suggestion (mass appraisal), it could be a little bit superficial. It sounds to me like a bit desk research, but not more. But then, you would still need do this the other approach (proactive) which, of course, requires more time, but then you have a really matching feedback, which is really about platform or about the tool that was developed. I think the combination of the two would be ideal, I would say" P5-EE "Mix does sound good. I mean, it's a nice way to also leverage these new technologies, and it lets you kind of do the foundational, but kind of you can go a step above if you want" P14-HCP-EE "So having the screening process mass appraisal in a high-quality way is a good start. But then I think the proactive approach on top of it to validate is very valuable" P23-HCP-TP

		"I think that focused versus mass-- maybe a combination of the two, but then defining that criteria for what is focused and what's mass might take time, right? But I think it'll pay off eventually" P28-In-TP "For me the mass appraisal might be the first step if you really have to give me the huge volume. But in any case… then you need the proactive part afterwards" P31-PA-EE "This is a very critical part because, three months ago, things are completely different. So, I would be the proactive-- mass appraisal doesn't work because with whatever is screening all the internet and try to get the information is not good… But now, with what's happening with the new AI revolution that is taking place and the ability to start controlling what type of information you are able to collect, it's very debatable that when we are launching these guidelines, things will completely change. So, a mass appraisal could work" P41-Ph-TP "The focused trials, while they're necessary for, as you said, ease of use or some interoperability tests maybe, they're not required, so you can very well launch that product in the market without-- or with having done very little focused appraisal. So, I agree that, in an ideal world, it is better, it is necessary, and that in the field, if we're looking at the technologies that are being launched and especially in the B2C context, I think a lot of manufacturers get away with measuring outcomes without human subjects' intervention or with very little human subjects' intervention" P42-RE
Subjective criteria		
	Assessor diversity (n= 34, 62%)	"I think that you would need to test it through diverse populations… Even if it's just a small group, I think that will be okay, but just to be aware of this diversity" P5-EE "I think definitely diversity in the team assessing the tool… particularly for a patient-facing tool, a good demographic mix of patients. I think that's probably the most important" P12-HCP "Maybe depending on who the tools might be designed for it might be interesting to have different user groups or user personas" P14-HCP-EE "And also, thinking about diversity and inclusion. As you say, making sure that you think about this from a socioeconomic perspective, from an age perspective, from an education perspective, often from an ethnicity perspective, all of these. And you have to understand or have a reasonable level of understanding what level of variability you're going to get" P38-Ph "So, I do think having multiple people kind of from a diversity of perspectives rate that criteria would be helpful" P45-Rs "For instance, just taking a sample of end user and basically, they test the solution and provide the feedback" P55-Ph
	Research evidence and validation (n= 20, 36%)	"All sort of user experience-related, it should be the end user has evaluated that or scored it through any number of mechanisms, through a survey through, through a usability test, through a user research study" P10-EE "I guess, most of the vendors, they just published data about usability, for example. If they use the usability score, which is the universal measurement-- and you cannot use the usability score if you don't follow the steps, and that should be a public information. And then you can say, yeah, this solution, they measure the usability, they publish this, and that's the score. For me, that's reliable data" P50-TP-Ph
	Specify and explain to reach a common understanding (n= 16, 29%)	"I think making it clear is key: This is what you're supposed to be doing. This is what we're measuring, and these are your tasks" P23-HCP-TP

		"There are some broad, I guess, design considerations about software which you can't ignore. If you're going to be required to navigate up and down for menus or enter data repeatedly to get to the next level, these are basic software design considerations where you can have rules-- sort of you have a set of rules which are well-established rules which you follow" P38-Ph "I think you can actually create some parameters… you can actually sort of break out into sort of: (Here's three boxes), so you could sort of narrow that down" P48-TP "If you assess, for example, usability, and you don't follow certain procedure, you cannot say usability is low or high because you didn't follow the needed steps to assess this…. when you are an assessor of usability, you should follow certain steps" P50-TP-Ph "I would say that, I guess in general, if there are clear examples, like what is meant with some of those… that might help to steer it towards same kind of understanding and how different people approach that kind of question" P52-TP
	Tool's ratings as proxy criteria (if critical mass is achieved) (n= 9, 16%)	"Customer reviews can be so important because that's just the real-world utilizers as opposed to these very motivated people who agreed to participate in a study and even know how to participate in a research study and have access to advertisements for things like that" P20-HCP-Rs "From my experience, unless you've got something with thousands of reviews, it's pretty easy for these star ratings to be skewed, especially if somebody's annoyed with the product or the provider" P30-PA "If there are overall ratings of usability by a critical mass and it's high, I would say that's a good indicator that it likely has good usability" P54-Rs
	Tool's use metrics as proxy (n= 4, 7%)	"We had a framework on verification validation, and very recently, we were looking into-- there could be a great tech tool technical components of it that they can also deliver good health outcomes, but if people don't use it, you're not able to collect the data... So, for us, when we were discussing about usability and utility criteria—we were thinking about: Do they enjoy using it based on the score? Will they use it again if they had to? Did they finish the complete sessions?... Or how many times they were hitting that button on tech support? I think those were objective things that we were analyzing" P8-EE "If you have a solution prescribed, for example, what is the rate of declining to take that into use, or what is the rate of non-registering? What is the rate of going from registration to actually use? What is the range from going to taking that actually into use and being engaged within a relevant time frame? Let's say in a 12week engagement or something like that" P27-In
Optionality of some criteria		
	Favors optionality (n= 39, 71%)	"I think we can't get away from having the optionality. I feel, however, there needs to be some rationale if something is selected as not relevant or not applicable. That's, I think, the risk mitigation around having optionality" P6-EE-In "I think you can give the option to exclude criteria and say it's not applicable but with a mandatory request for justification" P9-EE-Rs "My impression is that we will not escape the not applicable. Because eHealth tools are so various and the use case are limitless. So, we will not escape it" P11-HCP "I think it's good to have flexibility… because not everything is going to be black and white. There's always going to be gray, and you don't want to exclude something because it's not black or white" P12-HCP

		"I would argue there is no way around actually being able to exclude stuff. And I know this makes the scientific evaluation very hard. But yeah, I have very clear opinion there" P22-HCP-TP "I think there's so much variety in these devices and this medical solution. So yeah, I think it would also make this a little bit of optional and this is not that it makes it weaker or something like this" P34-PA-Rs "I completely agree on the optionality thing. I would say in those where criteria might be because we talked a little bit about the different use cases where it's applicable or not. So, you cannot just throw one framework at every use case, as we discussed. But I like the idea of the optionality because, as you said, it gives the point for reflection. That means basically if you skip it, you skip it why" P37-Ph "I'm supporting optionality for the fact that the context or let's say the niche in a different disease area are fulling different" P52-TP
	Disfavors optionality (n= 4, 7%)	"You don't do this, you don't give it, because if you do optionality, you say that, yeah, you can misunderstand it. So, you're allowed to not understand it… So, you give way, you give leeway to the assessor" P24-IE "I would put everything mandatory, but then I would leave a kind of open space in case the one who are doing the assessment might have some observations to include" P32-PA-EE "This can bring a lot of variability, which I'm not sure it's good for the framework" P50-TP-Ph
Current versus progressive criteria		
	Favors current criteria (n= 50, 91%)	"I think that the criteria should reflect the reality of today" P11-HCP "For the clinical decision making, this is a legal challenge that can't be met by any digital applications" P16-HCP-EE "… in terms of where we are today and in terms of making this framework as useful and relevant today. I think it makes sense... that's the reality of today. Maybe that will be different in the next 5 to 10 years. But probably this is an evolving tool. So maybe that's something worth revisiting then" P27-In "I think it is a living list. I think it's going to have to be evaluated on a regular basis. Something that is patient safety today might not be patient safety tomorrow. Things change, technological advances, medical advances, especially with the world we're going into genomics and AI and precision medicine and all of these domains" P28-In-TP "I think we don't want to be dehumanizing the way that healthcare is delivered. I mean, I think that technology should always anyway augment, complement the interaction between-- broadly, the interaction between humans. But obviously, in this case, between healthcare professionals and patients" P38-Ph
	Favors progressive criteria (n=3, 5%)	"It's very shortsighted… I think we are already at a point where AI is more precise. People make mistakes all the time. And the more under pressure they are, more mistakes they make… so, when you take that into consideration, this notion that somehow human knows better seems outdated" P2-EE "I side with the visionary who's looking down the trail" P30-PA