

Content Validity and Telemedicine Satisfaction Measures

Robert Garcia
DePaul University
rgarci11@cdm.depaul.edu

Peter Kallio
Rosalind Franklin University
peter.kallio@rosalindfranklin.edu

Olayele Adelokun
DePaul University
oadelokun@cdm.depaul.edu

Abstract

Confirming the relevance of measures through content validity can be among the most important but often over-looked aspects of measurement design. With the growing need to evaluate telemedicine satisfaction it is important that researchers pay more consideration to the relevance of measures used to represent studied constructs. This research discusses a content-validity effort using a formative approach for designing measures. By presenting insights gained during this process this research contributes to the knowledge by demonstrating both the importance and challenges of content validity and measure development in practice. Results identify several issues with differences in non-expert views, measurement modifications, participant matching strategies and form usability.

1. Introduction

A key part of successfully evaluating Information Systems (IS) is the ability to interpret and derive meaning from reported results. Yet, this can present challenges for those evaluating satisfaction with complex IS such as telemedicine. These challenges are often due to uncertainty around the meaning of satisfaction, its antecedents and measurements used.

Telemedicine is a term used to describe an IS that provides remote medical care across geographic distances using telecommunications technology [1]. Telemedicine is considered a subset of telehealth that focuses directly on clinical practices, as opposed to overall healthcare needs. Telemedicine usage has continued to grow over the years [2]. As usage has increased, so has the number of different types of telemedicine services and technologies. For example, telemedicine technologies can range from video conferencing on mobile devices, to remote monitoring using Internet of Things devices, to telepresence in virtual reality. Telemedicine is being used in a variety of medical services. These can include assessments and consultations across medical domains. Telemedicine can be provided for primary, secondary, and tertiary

care. As a result, there is a growing need for researchers, practitioners, and decision makers to evaluate and compare different telemedicine systems. To evaluate telemedicine systems, researchers often examine factors such as patient satisfaction [3, 4]. However, because of different views of the meaning of satisfaction and its complexities, findings can be difficult to generalize and compare [5].

Some of these challenges are due to deficiencies in the methodologies used for conducting evaluations [6]. Methodologies typically used in telemedicine often fail to account for measurement adequacy and validity. For example, a lack of standardized measures make comparisons difficult and researchers often end up creating their own measures [7]. This presents a problem as many of these measures are often not thoroughly validated or checked for reliability. While researchers have discussed the need for more rigorous approaches towards measurement development, many evaluation still often fail to emphasize the importance of content validity [6, 8]. Content validity can be defined as “the degree to which elements of an assessment instrument are relevant to, and representative of, the targeted construct for a particular assessment purpose” [9]. Whereas construct validity is concerned with evaluating the theoretical aspects of underlying constructs [10], content validity is used to evaluate a measurement instruments design in representing the targeted constructs [9]. Yet, while studies may present findings and clear theoretical underpinnings for the importance of content validity, there are many practical aspects that are often not discussed.

This research discusses experiences and findings during testing for content validity of a questionnaire to evaluate dimensions of patient satisfaction with telemedicine. While studies on satisfaction are typical in the IS literature [5], telemedicine differs in its service dependencies [11]. For a telemedicine video service for example satisfaction can depend on both the technology’s ability to deliver content as well as a provider’s ability to establish rapport with a patient. Each aspects quality can influence the perception of

either aspect as well as a patient's overall satisfaction. The goal of this research is not to describe the instrument development process in its entirety or justify the selected measures in relation to satisfaction. Rather this study seeks to contribute to the knowledge by providing insights gained and challenges faced during the content validity process. This is being done to help demonstrate both the importance and challenges of content validity for measures in telemedicine satisfaction research.

2. Literature review

Patient satisfaction plays a unique role in the success of telemedicine [11]. Studies suggest that satisfaction can directly impact medical outcomes [12]. Others show that satisfaction's influence on patient attitudes towards their care can affect their compliance with treatment plans and ongoing needs [13]. Yet while satisfaction is an important consideration for telemedicine evaluations, it remains a complex and often misunderstood construct [5]. Satisfaction can be based on both cognitive evaluations and emotional reactions [14]. It is considered a multi-dimensional construct that can be viewed from both outcome oriented and process-oriented perspectives [5]. In the IS literature satisfaction is often viewed as an antecedent for evaluation of system success [15]. Satisfaction is often viewed as consisting of aspects related to service, systems and information quality while informing usage [16, 17]. However existing IS models may not adequately capture psychological and organizational aspects of satisfaction [18].

Differences in the way satisfaction's complexity is modeled can impact the results of evaluations [5]. Research has discussed the importance of fully defining similar complex constructs and their dimensions [8]. Still concerns are often raised about the lack of attention to the multidimensionality of satisfaction in the telemedicine literature [19]. Yet even recent telemedicine studies still often rely on single measures of overall satisfaction [20]. Although, overall satisfaction may be a good enough indicator for these studies, it is important to consider what is being evaluated and whether measures fit their suggested meaning. For example, when patients say they are satisfied with telemedicine are they responding to the technologies being used, the care being administered or something else entirely? Agreement on the meaning of measures, metrics and instruments that should be used for evaluating telemedicine in different contexts can aid in research [21]. However, as there is a lack of standards for

evaluating telemedicine satisfaction, it is often up to individual researchers to determine which measures to use [14]. In many cases researchers create their own measures for examining telemedicine satisfaction [22]. In some cases, the measures that are used are entirely new or specific to the study [23]. In others they are combinations or modifications of items from previous questionnaires [24, 25]. Different approaches can provide new insight and test novel evaluation methods. However, there can also be confusion in how new measures relate to previous ones. This can present challenges when trying to compare results of studies. It is often also unclear the extent to which modifications to existing measures may or may not change how a measure is interpreted. This has led researchers to raise concerns over the number of satisfaction measures which are never tested for validity or reliability [19, 26]. The importance of measure validity, reliability and methods for examining them are well-documented [8]. However, this is often overlooked in telemedicine studies. When considered in relationship to the lack of agreement on the meaning of satisfaction, this creates an even greater uncertainty on what studies are actually measuring [19, 21]. Therefore, it is important that studies examine both the measures and their intended meaning. This is typically done during instrument development during content validity tests. Content validity describes the extent to which measures reflect the constructs they are meant to measure [27]. Without ensuring content validity, it is difficult to determine if a researcher's intended meaning for a measure fits the views of those completing an evaluation. Researchers have presented a number of different methods for assessing content validity [28]. More recently researchers have begun discussing the need for more research into examining and developing best practices for content validity in IS [29]. This research aims to contribute to this area, by discussing experiences and insights gained during the content validity process.

3. Methods

This research focuses on the design and testing of measures for examining dimensions of patient satisfaction with telemedicine. Dimensions of satisfaction are defined using an outcome-oriented approach where satisfaction is considered the outcome of a patient's telemedicine usage [5]. This study follows procedures described by Hoehle and Venkatesh [30] for developing measurements and guidelines from the MacKenzie, Podsakoff and Podsakoff [8] framework. This study expands on this work by using a formative approach towards measure

and content validity evaluations. Figure 1 below provides an overview of the content development and testing process. This will be discussed in further detail throughout the following subsections.

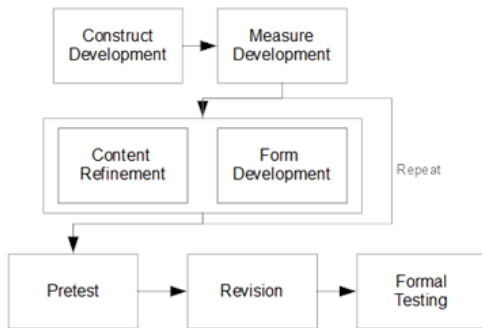


Figure 1. Content development and testing process

3.1. Construct and measure development

The preliminary steps in the process involved developing the constructs and measures to be evaluated. Constructs and their definitions were obtained from a previously published study conducted by the authors. In the previous study 18 constructs were identified using a grounded theory approach and expert feedback that was adapted from [30]. This approach was used to extract constructs from 23 previously validated telemedicine satisfaction questionnaires. The constructs were identified through expert feedback. They were then defined based on descriptions identified in the IS, healthcare, and telemedicine literature. This study began by developing the measures used to evaluate these constructs [8, 30]. To develop the measures, two researchers with expertise in telemedicine selected questions from coded questionnaires derived from the open coding process used in the previous study for construct creation. Items were only selected if they were labeled during the open coding process with labels that matched the specific construct. Two questions were selected for each construct in this group based on consensus of the measures face validity with a third researcher acting as a mediator. Although 3-5 measures are recommended to fully represent a construct, this research only sought to obtain preliminary findings. Future research will examine the constructs in more detail.

3.2. Content refinement

After the measures were selected, the next step was to refine and test the content to evaluate the measure validity. This was done simultaneously with form development that will be discussed in section 3.3.

A questionnaire was developed for this purpose. The questionnaire consisted of two forms that each contained a set of instructions, a matrix to be used for matching measures and a separate sheet containing the numbered definitions. This was a modification of the matrix described by [8, 31]. The modified matrix was used due to the length of each measure and the large number of measures. Each form contained measures corresponding to each construct for a total of 18 measures per form. Different measures were included in each form. Figure 2 below shows an example of the layout used for the form. Construct definitions were included on a separate sheet. The form was put through several rounds of testing to refine the wording of the items before a larger test was conducted. Testing was conducted with between 3-4 student participants. Ten rounds of testing were completed. Participants would complete the survey, provide verbal feedback and occasional notes. A narrative analysis of the qualitative feedback was conducted. During this time, the team would determine whether changes were needed to the form's design, or definitions and items required refinement.

Form 1: Measurement Items	Matching Definition – Write the # of the construct definition that best fits each item on the left
How well the telemedicine staff respected your freedom to protect your information from unauthorized disclosure	15
I trust my health care provider with concerns I have even when he or she does not ask	4

Figure 2. Design of initial form

During refinement, an issue was identified with the approach and the strategies participants used to complete the matching. After completing the survey participants were asked why they matched certain items together. Most participants discussed using a keyword matching strategy to relate items. Participants stated they would look for key words in the measures and match them to the same terms in the definitions. Participants discussed using this strategy as a time saving technique and to ease the mental workload required by having to learn a measure's meaning. The team felt keyword matching would make it difficult to ensure the measures had a meaning that matched the construct definition. Further, this strategy could result in false positives in cases where measures used terms that were in non-matching definitions. To address keyword matching the team analyzed the measures for key terms and attempted to reword them in the measures and definitions using synonyms or descriptions. Table 1 on the following page shows an example of this.

Table 1. Keyword matching example

Original definition	Adjusted definition
The degree to which patients perceive their privacy will remain protected and safe.	Patients willingness to share personal information and the control they have over that information is adequate
Original measure	Adjusted measure
How well the telehealth staff respected your privacy	How well the telemedicine staff respected your privacy

3.3. Form development and pretesting

Based on feedback and revisions during the content refinement process a new form was developed that integrated changes in definitions and measurements. To evaluate the effectiveness of the form a pretest was conducted using a within-subject design. The pretest was conducted at a University using paper and online versions of the form. Data was collected by recruiting on campus volunteers and online volunteers using Reddit. This was done in compliance with an approved process by the University Institutional Review Board (IRB). For content validity checks researchers suggest that using students can be appropriate in cases that rely on analytical thinking skills for sorting procedures [30, 32]. As the test involved analytically examining questions to determine corresponding definitions and not the validity of the constructs themselves, it was determined that there would be no realizable benefit by examining domain expertise. Further, students are often patients and any patient can potentially use telemedicine. Participants were provided an information sheet about the study, definition sheet and two matching forms. The definition sheet contained a list of numbered definitions. The matching forms contained a list of questions/statements and a box with instructions to write the number of the closest matching item. In total 129 paper surveys were collected and six online surveys. Data cleaning procedures were conducted to avoid inconsistencies in data. Surveys that were incomplete and/or had dimensions used for multiple or repeated constructs were removed. As these did not follow the instructions, they were deemed unusable. 79 surveys in total were evaluated. Two techniques were used. First an agreement analysis was conducted to determine whether items were successfully matched to their definitions. Second an analysis was performed to determine which items were most related. The agreement analysis was conducted using two

calculations: the proportion of substantive agreement (PSA) and the substantive validity coefficient (CSV) [30, 31].

$$PSA = \frac{nc}{N} \quad CSV = \frac{nc - n0}{N}$$

PSA measures the proportion of responses that assigned the correct dimension to the correct construct. nc is the number of responses that assigned the correct dimension to the correct construct and N is the total number of responses. Values for P range between 0-1. Higher values show greater agreement that the measure matches the correct construct. CSV measures the proportion of responses that were able to assign a dimension to a hypothesized construct instead of any other construct. n0 is the largest number of responses assigned to any other single construct. Values range between -1 and 1. A positive value indicates the dimension was assigned to the proper construct more than any other construct. A negative value would mean the opposite.

3.4. Revisions and Usability challenges

Following the pretest revisions were made to the form. During the preliminary testing phase several observations suggested usability challenges. The first observation was the number of incomplete forms. Only 58% of respondents that attempted the survey completed the forms. Participants were observed often complaining about the number of items per form. Many participants expressed frustration and confusion with the number of items and difficulty of the exercise. Some participants stated that after completing easy to match items they would randomly mark the rest. To address the usability concerns additional changes were proposed. Figure 3 below shows an example of the layout for the revised form edited to fit space requirements.

Definitions	
1. The environment in which the telemedicine session takes place.	
2. Patients feel they can access and receive all the information they deem important about their healthcare adequately.	
3. Patients' perceived cost or monetary expense of using telemedicine.	
Question	Number
How satisfied were you with the rooms and facilities?	1
Do you feel you were provided all the information you wanted during your visit to understand your medical condition?	2

Figure 3. Design of refined form

First it was decided that in order to reduce mental load and allow participants to focus more on the meaning of measures, the number of items per form

should be reduced to 7+/-2 [33]. Second was to include the definition list on each form to easily allow users to examine definitions. Finally, additional edits would be made to refine measures that participants had difficulty matching. A test was conducted using the re-designed form to evaluate any further usability challenges faced by participants. Participants were recruited in person on a University campus. Data was only collected for one day. In total 34 participants were recruited. Forms were distributed to participants on a rotating basis for the duration of the collection period. 17 participants completed the first form in its entirety and 16 completed the second form. The participant that did not complete the second form completed only the first grouping but not the second. Table 2 lists the constructs, definitions, and measures.

Table 2. Bolded Constructs followed by definitions and numbered questions evaluated in study

Constructs, definitions, and measures
<p>Comparison of care: Comparison between telemedicine and face-to-face visits</p> <ol style="list-style-type: none"> 1. I cannot be examined over telemedicine as well as I can by seeing a physician in person 2. I am as satisfied receiving care with telemedicine as I am receiving care in person
<p>Cost: Patients' perceived cost or monetary expense of using telemedicine</p> <ol style="list-style-type: none"> 1. Telemedicine cannot save me any money 2. Telemedicine reduces the price of health care
<p>Duration: The adequacy in the length of time patients spend in the actual visit with a medical provider and receiving care.</p> <ol style="list-style-type: none"> 1. The amount of time I was allowed to spend with the provider was not long enough to deal with everything I wanted 2. The amount of time spent with the physician you saw
<p>Ease of use: The system's technical functions are user friendly and easy to use</p> <ol style="list-style-type: none"> 1. The necessary telemedicine equipment seems difficult for me to use 2. I would imagine that most people would learn to use this system very quickly
<p>End User Support: The technical assistance and training provided by personnel to aid patients in using the technology</p> <ol style="list-style-type: none"> 1. I am getting the technical help I need by staff to access, use and understand the telemedicine system 2. How satisfied are you with the help you received by staff to use the telemedicine system?

<p>Environment: The environment in which the telemedicine session takes place.</p> <ol style="list-style-type: none"> 1. How satisfied were you with the location? 2. How satisfied were you with the rooms and facilities?
<p>Information completeness: Patients feel they can access and receive all the information they deem important about their healthcare adequately.</p> <ol style="list-style-type: none"> 1. I was given all the information I wanted about any continuing medical services I might need 2. Do you feel you were provided all the information you wanted during your visit to understand your medical condition?
<p>Interaction: The attitude in which medical care providers communicate with patients.</p> <ol style="list-style-type: none"> 1. When speaking with my medical providers I have been shown kindness and respect 2. The courtesy, respect, and sensitivity shown by medical staff during discussions
<p>Outcome: The resulting change in health of a patient due to a medical intervention</p> <ol style="list-style-type: none"> 1. As an outcome of your visit do you feel you are now better able cope with your illness? 2. As an outcome of your visit are you now able to deal more effectively with your illness?
<p>Provider benefits: Patient feels the system technology assists their medical providers in their work</p> <ol style="list-style-type: none"> 1. Telemedicine can help my nurse / physician perform their jobs 2. Telemedicine can help my physician care for patients more efficiently
<p>Privacy: Patients willingness to share personal information and the control they have over that information is adequate</p> <ol style="list-style-type: none"> 1. Sensitive data is protected from those who should not have access to it 2. How well the telemedicine staff respected your privacy
<p>Quality of care: The competency of the physician who cared for the patient</p> <ol style="list-style-type: none"> 1. Skill of provider: Ability to diagnose problems, thoroughness of examinations, skill in treating your condition, and scientific knowledge 2. The thoroughness, carefulness, and skillfulness of the provider who cared for you
<p>Relationship: The strength of the personal relationship developed between the patient and medical provider</p> <ol style="list-style-type: none"> 1. Because of our close ties this physician knows all about me 2. I trust my care provider and feel I can share health concerns I have with them even if they don't ask

<p>Reliability: The reliability, accuracy and consistency of the technology used.</p> <ol style="list-style-type: none"> 1. I cannot always trust the equipment to work 2. The system is stable in its performance
<p>Reuse: Patient thoughts on re-using the services and recommending it to others</p> <ol style="list-style-type: none"> 1. Would you suggest this form of treatment to someone else with your condition? 2. I will continue using the telemedicine system
<p>Scheduling: The time required for scheduling a session with a medical provider.</p> <ol style="list-style-type: none"> 1. The telemedicine appointments you make are set up quickly 2. The length of time to get a telemedicine appointment
<p>Treatment: The medicine, drugs and medical procedure given to a patient to manage their health condition.</p> <ol style="list-style-type: none"> 1. Satisfaction with medications, therapy and advice you were given for your illness 2. How satisfied are you with the therapy and/or medications you received for your illness?
<p>Usefulness: Patient believes using the system's technical functions enhance their task performance</p> <ol style="list-style-type: none"> 1. The technical features of the system were useful 2. Did you find the technical features of the system useful?

3.5. Dimensionality of usefulness

During the redesign testing a concern was raised on whether the Usefulness construct was being appropriately evaluated. The concern was based on the idea that usefulness like Satisfaction is multi-dimensional [34]. While constructs such as usefulness and usability are considered distinct from satisfaction they can potentially form or be reflective of satisfaction. Seminal work in the IS literature describes the influence of constructs such as these in contributing to the variances in user satisfaction [16]. Results from testing showed that usefulness was often matched with other measures by participants and many commented about it. Participants frequently matched measures for outcome and end user support to the usefulness construct in the pretest. For patient evaluations of telemedicine, it is possible that patients may decide that a technology that enhances their medical outcomes or supports it, makes it useful. A decision was made not to further evaluate usefulness and use already established measures in the final questionnaire. This would allow the study to examine constructs that were not typically evaluated and understand the degree to which their meaning may overlap without the influence of Usefulness.

3.6. Formal testing

Once the revised instrument was completed a formal test was conducted. The test was carried out using both paper and online versions. Participants were recruited in person for the paper version and via Reddit for the online version. The process was approved by the University IRB. A total of 632 participants were recruited for the formal testing. 10 participants took the online survey with the rest completing the paper copy. Of these 294 completed the first form and 285 the second. In the collected data, results in which a user either repeated responses or did not answer a response in a grouping were removed. For example, if a participant repeated a response in the first group but not the second, the second group's data would be evaluated but not the first. Removing repeated or missing responses was done to avoid skewing the results of CSV calculations.

4. Results

The results are described in three subsections. The first subsection presents the results of the pretest. The next subsection presents results of groupings following the pretest. A subsection then presents the results of formal testing. Finally, a section presents comparisons between the pretest and formal test.

4.1. Pretest results

The results of the pretest show a mixture of variations in performance on measures with participants having difficulties with some items yet performing better in others. The results of the pretest showed several difficulties with the matching of items. A threshold value of .60 was assigned to both parameters for the analysis based on suggestions by [30]. Cost, environment, information completeness and privacy were successfully matched on both forms. Many of the items had different results based on the form. Comparison of care and Ease of use were successfully matched on one form but not the other. Duration, relationship, reliability, reuse, and scheduling each had one item matching the .60 threshold for on one form for PSA but one item having a CSV below the threshold. Neither end user support, outcome, provider benefits, treatment nor usefulness had items meeting the .60 threshold. Table 3 on the following page shows the results of the pretest for both forms. Although the items were not separated in the forms, they are grouped in the table for comparison with formal testing.

Table 3: Results of pretest

Constructs	Pretest			
	Form1		Form2	
	PSA	CSV	PSA	CSV
Group 1				
Cost	0.949	0.949	0.962	0.962
Duration	0.734	0.646	0.646	0.43
Environment	0.823	0.785	0.861	0.823
Information completeness	0.671	0.608	0.696	0.633
Privacy	0.873	0.848	0.924	0.899
Reuse	0.658	0.582	0.772	0.709
Scheduling	0.797	0.747	0.709	0.481
Group 2				
Comparison of care	0.506	0.43	0.671	0.62
End user support	0.519	0.342	0.456	0.367
Interaction	0.608	0.481	0.57	0.443
Outcome	0.329	0.025	0.443	0.316
Provider benefits	0.557	0.481	0.582	0.494
Quality of care	0.633	0.532	0.481	0.342
Relationship	0.772	0.696	0.671	0.582
Treatment	0.282	0.013	0.418	0.253
Variable				
Ease of use	0.582	0.494	0.81	0.734
Reliability	0.744	0.692	0.633	0.544
Usefulness	0.532	0.418	0.418	0.304

4.2. Usability refinements

Following the results of the pretest items were separated between two forms. The items were separated based on the agreement with the hypothesized matches and their closeness to other measures that did not agree with the hypothesized matches. This was done to ensure that items compared on a form were tested against measures that were most similar in meaning. To accomplish this, measures were first separated into two forms based on the CSV and PSA agreement. Similar to Hoehle and Venkatesh [30] a .60 cut-off value was used. Items above a .60 threshold were grouped into one form and those below into a second. From additional analyses, it was shown that on form 1 interaction measures were often defined as relationship and quality of care by participants. This suggested that these items should be grouped together. On form 2 both CSV and PSA were above the .60 threshold while CSV was below .60 for both

duration and scheduling. Duration and scheduling were also frequently matched to each other and therefore grouped together. Ease of use had a PSA and CSV below .60 and was frequently matched with End user support on form 1. On form 2 however Reliability had a CSV below .60 and was more frequently matched with End user support and Treatment.

It was therefore decided to group Ease of use with End user support on form 1 and Reliability with End user support and Treatment on form 2. Two separate questionnaires were created. The first contained form 1 measures and the second form 2 measures. Each form was divided into separate pages. The first grouping was added on one page together with the related definitions list and the second grouping on the second page. The results of the groupings are shown in table 4.

Table 4: Revised item groupings

Form 1 items		
<i>First grouping</i>		
• Cost	• Information completeness	• Reuse
• Duration	• Privacy	• Reliability
• Environment		• Scheduling
<i>Second grouping</i>		
• Comparison of care	• Provider benefits	
• Ease of use	• Quality of care	
• End user support	• Relationship	
• Interaction	• Treatment	
• Outcome	• Usefulness	
Form 2 items		
<i>First grouping</i>		
• Cost	• Environment	• Privacy
• Duration	• Information completeness	• Reuse
• Ease of use		• Scheduling
<i>Second grouping</i>		
• Comparison of care	• Quality of care	
• End user support	• Relationship	
• Interaction	• Reliability	
• Outcome	• Treatment	
• Provider benefits	• Usefulness	

As discussed previously the groupings were tested before being used. No issues were observed with participants using the form. Researchers conducting the study observed most participants completing the survey within ten minutes. Although the data obtained was small, the results suggest minor improvements could be made to alleviate further difficulties. These were then done before the formal test.

4.3. Formal test

The results of the formal test show similar variations but improvements for some items over the pretest. The results shown in table 5 below are listed based on the groupings used for the form.

Table 5: Results of formal test

Constructs	Formal test			
	Form1		Form2	
	PSA	CSV	PSA	CSV
Group 1				
Cost	0.908	0.881	0.908	0.880
Duration	0.793	0.673	0.813	0.721
Environment	0.837	0.779	0.770	0.668
Information completeness	0.779	0.701	0.784	0.717
Privacy	0.820	0.745	0.845	0.802
Reuse	0.816	0.779	0.749	0.689
Scheduling	0.799	0.707	0.837	0.767
Group 2				
Comparison of care	0.788	0.727	0.678	0.615
End user support	0.635	0.485	0.470	0.276
Interaction	0.799	0.751	0.633	0.537
Outcome	0.626	0.486	0.608	0.392
Provider benefits	0.612	0.507	0.470	0.314
Quality of care	0.704	0.646	0.574	0.440
Relationship	0.830	0.789	0.650	0.523
Treatment	0.636	0.442	0.557	0.394
Variable				
Ease of use	0.602	0.442	0.809	0.760
Reliability	0.867	0.833	0.682	0.569
Usefulness	Na	Na	Na	Na

The results show agreement for measures of cost, duration, environment, information completeness, privacy, reuse, and scheduling on both forms. Several measures performed better on the formal test than on the pretest. Comparison of care on form 1 of the pretest did not pass the threshold for either measure. In the formal test it passed the threshold for both PSA and CSV. Several measures saw improvement in PSA values above the .60 threshold. These include end user support, provider benefits and ease of use on form 1, interaction and treatment on form 2 and outcome on both forms. Several results showed improvement in CSV values that put them above the threshold for both

indicators. These included reuse, interaction, and quality of care on form 1 and duration and scheduling on form 2.

4.4. Result comparison

Based on the results it was unclear whether form changes or changes to measures had a significant effect on PSA and CSV values. It is particularly important to consider whether the groupings were responsible for differences in results. To examine the relationship between the two forms a two-tailed independent t-test was performed using the Python Scipy library. The test was performed on items between forms and studies without considering usefulness as it was not evaluated in both tests. The results suggest a significant difference for between PSA values on form 1. The results $t(32) = 2.152, p = .039$ suggest a significant difference at $p \leq .05$ for form 1 pretest ($M = .6492, SD = .1792$) and formal test ($M = .7559, .0984$). No significant difference was observed for PSA results on form 2 or between CSV results on either form. To examine this further t-tests were performed on items based on groupings. Significant differences were only observable between pretest and formal test on form 1 group 2 results. For PSA, a significant difference was observed at $t(16) = 2.743, p = .014$ between pretest ($M = .532, SD = .1505$) and formal test ($M = .6924, SD = .0902$) at $p \leq .05$. A significant difference was also observed for CSV values at $t(16) = 2.205, p = .042$ between pretest ($M = .388, SD = .229$) and formal test ($M = .5862, SD = .1413$).

5. Discussion

Among the critiques of studies in the telemedicine literature is the lack of rigor in measurement development [19]. Content-validity is an important but often missing element reported in studies. By demonstrating the content-validity process using non-experts this research identified several challenges. Both formal and pretest results show variability between non-expert views of measures. Questions can be raised as to whether our experts selected the appropriate measures to represent constructs or if the items, are well-defined. Even if our experts erred on their selection of measures and definitions, the use of experts is common practice. As experts often validate measures and interpret results, this could suggest uncertainty in their use and the extent to which non-experts agree with their views [35, 36]. Even among non-experts there was no consensus between participants on several measures. This raises questions on whether patients may have unique views of what

survey measurements actually mean [29]. This is particularly an issue with self-designed measures in telemedicine with no content validity checks [6, 8]. Examining differences in CSV we observed how views on measurements overlap between constructs. Overlapping views of measures can lead to agreement in responses that are based more on similar interpretations of a measures wording than views of its relationship to an underlying construct. This can have impacts when conducting statistical analysis such as correlation or regression analyses as it can be uncertain as to where the agreement stems from. While our modifications may have altered the meaning of previously validated measures, question modification is a common practice in telemedicine [24, 25]. This suggests more caution is needed when interpreting results of modified questions and greater care should be taken in defining measures. Yet even when defining measures care must be taken to account for user evaluation strategies. Strategies such as keyword matching during tests may impact the extent to which checks truly reflect a participant's view. [30] discussed issues related to the usability with using matching forms. Our efforts show these challenges, their impact on participants and a grouping strategy to address this. By ensuring that items closely confused for each other were not separated an attempt was made to limit the impact of the separation Our analysis showed little variation between results of single-form and grouped items except for items on form 1 in group 2. It is unclear why exactly these discrepancies occur. Possible differences in form performance could be attributed to the number of participants, question wording, item grouping, enhanced usability, or removal of usefulness. While separating measures may reduce difficulties for participants, it can also reduce the broadness of comparisons. This can potentially result in measures being related to imprecise meanings because a lack of valid options.

6. Conclusions and implications

The goals of this research were to describe challenges and insights gained during the content validity process. The results of this study and adjustments made during the content validity process suggest several implications around differences in non-expert views, measurement modifications, participant matching strategies and form usability that should be considered for measurement design. First the variations in views demonstrate the value of non-expert evaluations of content validity and suggest their views may differ from those of experts. Second it is important to consider the extent to which modifications to or mixing of measurements from

previously validated instruments may affect participant views. Third user while it is unclear the extent to which user strategies for completing tests such as keyword matching may impact results it is important to consider how participants are evaluating tests. Fourth, it is important to consider the design of forms and how content is grouped, especially if using content across pages.

Finally, the results also demonstrated non-expert support for at least one measure for the satisfaction dimensions examined with the exception end user support, provider benefits, outcome, and treatment. Previous studies have used measures with CSV values less than .60 as long as PSA values were greater than .60 with minor edits [30]. Therefore, it is likely adjustments can be made to the wording of the remaining questions to address the differences and be used for testing. However, more research is needed to refine questions and construct meaning.

7. References

- [1] S. Sood., V. Mbarika, S. Jugoo, R. Dookhy, C.R. Doarn, N. Prakash, and R.C. Merrell, "What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings", *Telemedicine and e-Health*, 2007, 13, (5), pp. 573-590
- [2] Mordor Intelligence. e-Health Market- Growth, Trends, and Forecasts (2020 - 2025). <https://www.mordorintelligence.com/industry-reports/e-health-market>
- [3] R.E. Hanson, M. Truesdell, G.T. Stebbins, A.L. Weathers, and C.G. Goetz, "Telemedicine vs office visits in a movement disorders clinic: comparative satisfaction of physicians and patients", *Movement disorders clinical practice*, 2019, 6, (1), pp. 65-69
- [4] L.R. Moo, M.E. Gately, Z. Jafri, and S.D. Shirk, "Home-based video telemedicine for dementia management", *Clinical gerontologist*, 2020, 43, (2), pp. 193-203
- [5] R. Vaezi, A. Mills, W. Chin, and H. Zafar, "User Satisfaction Research in Information Systems: Historical Roots and Approaches", *Communications of the Association for Information Systems*, 2016, 38, (27), pp. 501-532
- [6] S. AlDossary, M.G. Martin-Khan, N.K Bradford, and A.C. Smith, "A systematic review of the methodologies used to evaluate telemedicine service initiatives in hospital facilities", *International journal of medical informatics*, 2017, 97, pp. 171-194
- [7] E. Batbaatar, J. Dorjdagva, A. Luvsannyam, M.M. Savino, and P. Amenta, "Determinants of patient satisfaction: a systematic review", *Perspectives in public health*, 2017, 137, (2), pp. 89-101
- [8] S.B. MacKenzie, P.M. Podsakoff, and N.P. Podsakoff, "Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques", *MIS quarterly*, 2011, 35, (2), pp. 293-334

- [9] S.N. Haynes, D. Richard, and E.S. Kubany, "Content validity in psychological assessment: A functional approach to concepts and methods", *Psychological assessment*, 1995, 7, (3), pp. 238
- [10] R.J. Cohen, M.E. Swerdlik, and S.M. Phillips, "Psychological testing and assessment: An introduction to tests and measurement". Mayfield Publishing Co, California city, CA, 1996..
- [11] P.H. Hu, "Evaluating telemedicine systems success: a revised model", In 36th Annual Hawaii International Conference on System Sciences, 2003, pp. 8.
- [12] G.D. Kennedy, S.E. Tevis, and K.C. Kent, "Is there a relationship between patient satisfaction and favorable outcomes? ", *Annals of surgery*, 2014, 260, (4), pp. 592
- [13] C.Y. Chou, and D.J. Brauer, "Temperament and satisfaction with health status among persons with rheumatoid arthritis", *Clinical Nurse Specialist*, 2005, 19, (2), pp. 94-100
- [14] E. Shirley, G. Josephson, and J. Sanders, "Fundamentals of Patient Satisfaction Measurement", *Physician leadership journal*, 2016, 3, (1), pp. 12
- [15] W.H. DeLone, and E.R. McLean, "The DeLone and McLean model of information systems success: a ten-year update", *Journal of management information systems*, 2003, 19, (4), pp. 9-30
- [16] S. Petter, W. DeLone, and E. McLean, "Measuring information systems success: models, dimensions, measures, and interrelationships", *European journal of information systems*, 2008, 17, (3), pp. 236-263
- [17] B.H. Wixom, and P.A. Todd, "A theoretical integration of user satisfaction and technology acceptance", *Information systems research*, 2005, 16, (1), pp. 85-102
- [18] N. Au, E.W. Ngai, and T.E. Cheng, "Extending the understanding of end user information systems satisfaction formation: An equitable needs fulfillment model approach", *MIS quarterly*, 2008, pp. 43-66
- [19] T.L. Williams, C.R., May, and A. Esmail, "Limitations of patient satisfaction studies in telehealthcare: a systematic review of the literature", *Telemedicine Journal and e-Health*, 2001, 7, (4), pp. 293-316
- [20] M.T. Nawas, H.J. Landau, C.S. Sauter, C.A. Featherstone, S.A. Kenny, E.S. Rodriguez, L.G. Johnson, S.A. Giralt, and M. Scordo, "Pilot Study of Telehealth Evaluations in Patients Undergoing Hematopoietic Cell Transplantation", *Biology of Blood and Marrow Transplantation*, 2020, 26(6) pp 135-137
- [21] S. Zhang, S.I. McClean, D.E. Jackson, C. Nugent, and I. Cleland, "Patient Satisfaction Evaluation of Telemedicine Applications Is Not Satisfactory", *Mediterranean Conference on Medical and Biological Engineering and Computing*, Springer, 2014, pp. 1140-1143
- [22] J.F. Robb, M.H. Hyland, and A.D. Goodman, "Comparison of telemedicine versus in-person visits for persons with multiple sclerosis: A randomized crossover study of feasibility, cost, and satisfaction", *Multiple sclerosis and related disorders*, 2019, 36, pp. 101258
- [23] J.H. DeAntonio, H.S. Kang, H.C. Cockrell, W. Rothstein, C. Oiticica, and D.A. Lanning, "Utilization of a handheld telemedicine device in postoperative pediatric surgical care", *Journal of pediatric surgery*, 2019, 54, (5), pp. 1005-1008
- [24] G. Rogers, "Using Telemedicine for Pediatric Preanesthesia Evaluation: A Pilot Project", *Journal of PeriAnesthesia Nursing*, 2020, 35, (1), pp. 3-6
- [25] D. Rickwood, A. Wallace, V. Kennedy, S. O'Sullivan, N. Telford, and S. Leicester, "Young People's Satisfaction With the Online Mental Health Service eheadspace: Development and Implementation of a Service Satisfaction Measure", *JMIR mental health*, 2019, 6, (4), pp. e12169
- [26] M.A. Iseli, R. Kunz, and E. Blozik, "Instruments to assess patient satisfaction after teleconsultation and triage: a systematic review", *Patient preference and adherence*, 2014, 8, pp. 893
- [27] R.F. DeVellis, "Scale development: Theory and applications", Sage publications, Newbury Park, CA, 2016.
- [28] E. Almasreh, R. Moles, and T.F. Chen, "Evaluation of methods used for estimating content validity", *Research in Social and Administrative Pharmacy*, 2019, 15, (2), pp. 214-221
- [29] J. London, K. Matthews, and V. Grover, "On Meaning and Measurement: A Review of Content Validity in IS", *Twenty-third Americas Conference on Information System. Proceedings of the*, 2017
- [30] H. Hoehle, and V. Venkatesh, "Mobile application usability: conceptualization and instrument development", *Mis Quarterly*, 2015, 39, (2), pp. 435-472
- [31] J.C. Anderson, and D.W. Gerbing, "Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities", *Journal of Applied Psychology*, 1991, 76, (5), pp. 732
- [32] D. Compeau, B. Marcolin, H. Kelley, and C. Higgins, "Generalizability of information systems research using student subjects—a reflection on our practices and recommendations for future research", *Information Systems Research*, 2012, 23, (4), pp. 1093-1109
- [33] G.A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information", *Psychological review*, 1956, 63, (2), pp. 81
- [34] R.G. Saadé, "Dimensions of perceived usefulness: Toward enhanced assessment", *Decision Sciences Journal of Innovative Education*, 2007, 5, (2), pp. 289-310
- [35] M. Yip, A.M. Chang, J. Chan, and A.E. MacKenzie, "Development of the Telemedicine Satisfaction Questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study", *Journal of Telemedicine and Telecare*, 2003, 9, (1), pp. 46-50
- [36] S. Bakken, L. Grullon-Figueroa, R. Izquierdo, N.J. Lee, P. Morin, W. Palmas, J. Teresi, R.S. Weinstock, S. Shea, and J. Starren, "Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire", *Journal of the American Medical Informatics Association*, 2006, 13, (6), pp. 660-667