## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Assessing patient experience with patient safety in primary care: development and validation of the ASK-ME-questionnaire |
| **AUTHORS** | Stahl, Katja; Reisinger, Anna; Groene, Oliver |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Pedrosa, V<br>Polytechnic of Leiria |
| **REVIEW RETURNED** | 15-Apr-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | Well done, a very important and nice topic. congratulationz |

| | |
|---|---|
| **REVIEWER** | Mavrodi, Afentoula<br>University of Macedonia, Business Administration |
| **REVIEW RETURNED** | 20-Apr-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | Exceptional work that adds to the patient safety existing literature. Adequate study design, sample size and methodology. |

| | |
|---|---|
| **REVIEWER** | Ye, Xu-Chun<br>Second Military Medical University, School of Nursing |
| **REVIEW RETURNED** | 06-May-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | This manuscript focused on patient experience with patient safety in primary care and developed a self-report ASK-ME questionnaire, which may contribute a lot to patient safety management. However, even lots ways were adopted for the questionnaire development and validation analysis, this manuscript still need some revision or clarify to make it more clear and easier for reading.<br><br>1. P.3, L.25-28, It seemed that the results in the abstract part were confusing, “The final ASK-ME-questionnaire consisted of 22 items in five domains……while the EFA yielded 3 factors”. The authors developed a 75 items questionnaire in 5 dimensions (generally as theoretical dimensions) and then deleted to 22 items (the methodology steps were not very clear). After EFA and CFA, a three-factor structure questionnaire was developed. Therefore, generally we would agree that the final questionnaire consisted 22 items in 3 factors/domains, but not 5 domains.<br>2. P.4, L.29-41, In the background part, the last two aims of the project didn't fit the aim of this manuscript. Thus, it seems no need to report these two aims in this manuscript.<br>3. P.6, L27-59: The development of items to be overly simplistic and in some cases, confusing. Should the authors describe the steps in more details how and why 75 items are deleted to 22 |

| | |
|---|---|
| | items, especially in a more normalized methodology way? What's the criteria for item delete?<br>4. P.5, L.42-46: 'Items were excluded from psychometric evaluation if there were more than 15% missing values (including those who choose a 'does not apply' -response option', it seems not a well-recognized method for item exclusion. Should the authors provide some methodology evidence?<br>5. P.8, L.38: What's the meaning "condition last › 3 months" in table 1?<br>6. P.14, L.48: 'All three scales were……', maybe 'three subscales' are more reasonable? |

**VERSION 1 – AUTHOR RESPONSE**

Reviewer: 1
Dr. V Pedrosa, Polytechnic of Leiria Comments to the Author:
Well done, a very important and nice topic. congratulationz

Reviewer: 2
Dr. Afentoula Mavrodi, University of Macedonia Comments to the Author:
Exceptional work that adds to the patient safety existing literature. Adequate study design, sample size and methodology.

Reviewer: 3
Dr. Xu-Chun Ye, Second Military Medical University Comments to the Author: This manuscript focused on patient experience with patient safety in primary care and developed a self-report ASK-ME questionnaire, which may contribute a lot to patient safety management. However, even lots ways were adopted for the questionnaire development and validation analysis, this manuscript still need some revision or clarify to make it more clear and easier for reading.

1. P.3, L.25-28, It seemed that the results in the abstract part were confusing, "The final ASK-ME-questionnaire consisted of 22 items in five domains……while the EFA yielded 3 factors". The authors developed a 75 items questionnaire in 5 dimensions (generally as theoretical dimensions) and then deleted to 22 items (the methodology steps were not very clear). After EFA and CFA, a three-factor structure questionnaire was developed. Therefore, generally we would agree that the final questionnaire consisted 22 items in 3 factors/domains, but not 5 domains.

Thank you for pointing this out. This is a wording lapse (in the text we are referring to these theoretical dimensions as dimensions). We adjusted the wording in the abstract to that used in the text.

2. P.4, L.29-41, In the background part, the last two aims of the project didn't fit the aim of this manuscript. Thus, it seems no need to report these two aims in this manuscript.

We included all three aims of the study because we thought it worthwhile make the context of the study transparent to the reader. It also goes a way to explain the sample size (see second paragraph in the method section 'data collection'). We would therefore like to keep this paragraph. However, if the article is deemed too long, we agree that this paragraph could be shortened by not reporting the other two study aims.

3. P.6, L27-59: The development of items to be overly simplistic and in some cases, confusing. Should the authors describe the steps in more details how and why 75 items are deleted to 22 items, especially in a more normalized methodology way? What's the criteria for item delete?

We included information on criteria for item selection during the Delphi process in the results section and specified the information in the method section. We also reworded the sentence on the items referring to actual experience with PSE's.

4. P.5, L.42-46: 'Items were excluded from psychometric evaluation if there were more than 15% missing values (including those who choose a 'does not apply' -response option', it seems not a well-recognized method for item exclusion. Should the authors provide some methodology evidence?

We were not sure whether this refers to "15%" which is an extent that is seen as impairing any interpretation and for which we provide a reference, or whether it refers to treating 'Does-not-apply answers' as missing data. We did the latter because we do not know how these participants would have answered if the situation would have applied. Keeping them for analysis would have required to interpret the answer in one way or another (good or bad experience). We changed the wording.

5. P.8, L.38: What's the meaning "condition last › 3 months" in table 1?

Thank you for this, we agree that this may be misleading. Meant is a condition that has been existing for 3 months or longer (indicating a chronic condition). We changed the wording.

6. P.14, L.48: 'All three scales were……', maybe 'three subscales' are more reasonable?

We think that this comment refers to page 12? We changed the wording (throughout the text where applicable).

Reviewer: 1
Competing interests of Reviewer: no

Reviewer: 2
Competing interests of Reviewer: No competing interests.

Reviewer: 3
Competing interests of Reviewer: no competing interests

**VERSION 2 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Ye, Xu-Chun<br>Second Military Medical University, School of Nursing |
| **REVIEW RETURNED** | 03-Sep-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | 1. This manuscript focused on an interesting work in the area of patient safety in primary care, which is really needed more well-designed study. The authors tried to develop a user-friendly tool measuring patient experience with patient safety in primary care for routine use. On the whole, the instrument development was based on a well-designed research process.<br>2. However, I'm afraid I have to suggest the authors make more clear about the methodological difference between instrument development and questionnaire for status quo analysis. They are different from each other, both in the purpose and the methodology. As I know, the main purpose for instrument development is to develop an effective instrument to measure a certain feature composed of several components with coordinating relation, which requires a strict research process including theoretical dimension development, item generation, large sample |

| | |
|---|---|
| | test following with EFA and CFA, as well as item selection using some statistical strategies. While developing a questionnaire for status quo analysis, there is no so many strict steps needed. In the background part, the authors described threefold aims, mainly measuring factors contributing to the occurrence of patient safety events (influencing factors) and the patient experience of patient safety events (the outcome indicators). It is logically ok if for status quo analysis with a self-developed questionnaire with 22 items. Actually, the authors presented a well-designed instrument development research steps in the methods part, from 75 items with 5 theoretical dimensions to the pretest questionnaire with 22 items, and at last 14 items with 3-factor structure after EFA and CFA. But what confused me was that they put both the influencing factors and the outcome indicators in the same instrument. Actually, the outcome indicators (items about patient experience of patient safety events) in this study was excluded after EFA, with only 14 items 3 factor structure left. But what made me confused again was that the authors presented the result as “The final ASK-ME-questionnaire consisted of 22 items covering five dimensions” (abstract). Should it not be “14 items 3-factor structure”? Though some items could be kept for some professional considerations, additional psychometric properties measurement analysis should also be required. |

**VERSION 2 – AUTHOR RESPONSE**

Reviewer: 3
Dr. Xu-Chun Ye, Second Military Medical University

Comments to the Author:

1. This manuscript focused on
an interesting work in the area of patient safety in primary care, which is really needed more well-designed study. The authors tried to develop a user-
friendly tool measuring patient experience with patient safety in primary care for routine use.
On the whole, the instrument development was based on a well-designed research process.

Thank you. We understand that there is no comment in this paragraph suggesting specific changes.

2. However, I’m afraid I have to suggest the authors make more clear about the methodological diff erence between instrument development and questionnaire for status quo analysis. They are different from each other, both in the purpose and the methodology. As
I know, the main purpose for instrument development is to develop an effective instrument to measure a certain feature composed of several components with coordinating relation, which requires a strict r esearch process including theoretical dimension development, item generation, large
sample test following with EFA and CFA, as well item selection using some statistical strategies.

We agree and for instrument development we started with a literature research on
relevant dimensions and existing tools for item generation. We discussed these in a modified 3-round-Delphi process with experts (researchers, patient safety experts, clinicians and patients)
and consented the items for the pretest version of the questionnaire. We conducted cognitive intervie ws with patients before the pilot test (see paragraphs ‚Item development‘ in the methods and results s ection). We piloted this version in 22 practices, generating a dataset a large sample

(over 3000 cases) for psychometric testing, using EFA and CFA,
internal consistency testing and checked for discriminating validity of the scales (see data analysis in t he methods section and the whole results section)

While developing a questionnaire for status quo analysis, there is no so many strict steps needed.

We agree that, in principle, a full psychometric validation of the instrument may not be required for a simple status quo analysis. However, there are various reasons why we employed the rigorous process as described above. First, we believe that the steps “literature research” to identify relevant instruments and dimensions, “Delphi panel” to reach consensus and “cognitive testing” are indispensable in any questionnaire development. Second, we believe the psychometric testing is indispensable to identify any overlapping dimensions and items in order to provide a concise and parsimonious instrument fit for use in the busy routine clinical context.

In the background part, the authors described threefold aims, mainly measuring factors contributing to the occurrence of patient safety events (influencing factors)
and the patient experience of patient safety events (the outcome indicators). It is logically ok if for stat us quo analysis with a self-developed questionnaire with 22 items.

There seems to be something missing in the first sentence. We understand it that our methods are de emed appropriate to measure status quo with a 22-item questionnaire.

Actually, the authors presented a well-
designed instrument development research steps in the methods part, from 75 items with 5 theoretical dimensions to the pretest questionnaire with 22 items, and at last 14 items with 3-
factor structure after EFA and CFA.
But what confused me was that they put both the influencing factors and the outcome indicators in the same instrument.

Our aim was to develop a questionnaire suitable for routine use (short), applicable in
all practices (generic), for internal quality management (actionable,
i.e. containing information on both processes (indicators)
and outcomes (frequencies of events)). Therefore, we included items on bothspan style="font-family:'Segoe UI'; color:#ff0000; background-color:#ffffff"> aspects in the questionnaire (i.e. we did not include the questions on frequencies of eve nts with the intent to use them for testing criterion validity) and we also decided to keep the outcome-items although they did not form
a factor (which was due to the fact that we had to exclude them from analysis because the precentag e of missings exceeded 15%, see second paragraph of „Exploratory factor analysis“
in the results section). We discuss this along with the fact that PSE-
frequencies collected with our questionnaire differ from existing evidence and that further research is needed in how to adequately assess frequency of PSE. See
last paragraph of the discussion section: „In contrast to existing measures (2,15), experience of PSEs and resulting harm was assessed with two generic questions. Both had to be excluded from further psychometric analyses because more than 15% of the patients had either not answered the question or were not sure whether what they had experienced represented a PSE. The proportion of patients who did experience a PSE was low (2,9%) compared to studies using more detailed measures. However, these studies did either not use a self-report measure (2,34) or had high proportions of patients skipping the detailed section on PSE experience and a low overall responserate (29). Further research is needed in how to adequately assess number and type of experienced events in routine measurements.”

We added a sentence at the end of the paragraph „Exploratory factor analysis“
in the results where we state that the questions excluded from the EFA-

model were still kept in the questionnaire for content-related reasons. We added a column in the table in the online supplement 2 where we explain the con tent-related reasons.

Actually, the outcome indicators (items about patient experience of patient safety events) in this study was excluded after EFA, with only 14 items 3 factor structure left.

Exclusion of these items was due to the percentage of missing values exceeding 15%. See method section „data analysis“: „Items were excluded from psychometric evaluation if there were more than 15% missing values (either due to omitted answers, ‘does not apply’-answers or due to filtering questions”)

and also results section „Exploratory factor analysis“: „Items with more than 15% missing data (either due to omitted or ‘does-not-apply’-answers (Q14, Q18, Q19, Q20) were not incorporated in the EFA.” (-> Q18, Q19 are the variables measuring the frequency of the events (outcome indicators))

and Online supplement 2, table „EFA excluded items“ where the reasons for exclusion are given.

But what made me confused again was that the authors presented the result as “The final ASK-ME-questionnaire consisted of 22 items covering five dimensions” (abstract). Should it not be “14 items 3-factor structure”?

The 22-item-*questionnaire* covers 5 dimensions (access, communication, medication safety, care coordination and experience of PSE), with access and care coordination being covered by single item scales, medication safety covered by a 4-item-scale, communication by a 4-tem-scale and a 6-item-scale, and experience of PSE by two single questions). We can see that this wording might be misleading, t herefore we changed it to make the difference between dimensions and factors clearer in the abstract . We also added the number of items comprised in the factors to clarify that not all items in the questionnaire make up the factors.

We also discuss the fact that there are two single-item-scales: „The dimensions access and care coordination were assessed by single items and did not result in a factor. Given that these dimensions are also considered highly relevant for patient safety and patients are well placed to report on these dimensions, they were kept in the questionnaire.”

Though some items could be kept for some professional considerations, additional psychometric prop erties measurement analysis should also be required.

Due to word constraints we do not report in the abstract on the reasons for keeping items in the questi onnaire that were not included in the EFA-model. However, we added a sentence in the text (Results section, EFA) why the 8 items, that were excluded from the final EFA, were still kept in the questionnaire as w ell as the reasons for keeping them in Online Supplement 2. We also included a sentence that considering adding questions to strengthen the single-item-scales could be worthwhile, however it would be a trade-off between questionnaire length/practicability/acceptability and detail.

Reviewer: 3
Competing interests of Reviewer: No competing interests.

## VERSION 3 – REVIEW

| | |
|---|---|
| **REVIEWER** | Ye, Xu-Chun<br>Second Military Medical University, School of Nursing |
| **REVIEW RETURNED** | 29-Jan-2022 |

| | |
|---|---|
| **GENERAL COMMENTS** | Compare to the first version, this paper has been better organized. The questionnaire development process was clearer than the first version. However, what confused me is still the logically consistent. The authors declared that the ASK-ME-questionnaire was developed consisting of 22 items covering five dimensions, while the EFA and CFA yielded 3 factors, 14 items. I could understand the eight items excluded from EFA were kept for content-related reasons, even too much items were kept. However, it's really confused me that the 22-item questionnaire showed 'very good acceptability, good construct validity……' but based on a 14-item and 3-factor analysis. For the methodologically and logically consistent, maybe a final 14 items questionnaire were more acceptable. Alternatively, if the authors want to keep the 22 items questionnaire, psychometric properties analysis based on the 22-item data may be more reasonable. |

## VERSION 3 – AUTHOR RESPONSE

Reviewer: 3
Dr. Xu-Chun Ye, Second Military Medical University Comments to the Author:
Compare to the first version, this paper has been better organized.
The questionnaire development process was clearer than the first version. However, what confused m e is still the logically consistent. The authors declared that the ASK-ME-
questionnaire was developed consisting of 22 items covering five dimensions, while the EFA and CFA yielded 3 factors, 14 items.
I could understand the eight items excluded from EFA were kept for content-
related reasons, even too much items were kept.
We agree that an argument can be made that keeping 8 items may be too high a number. However, we decided to keep them for content-related reasons. The two single-
item scales covering the theoretical dimensions access (Q1) and
care coordination (Q16) were kept because these dimensions are considered highly relevant for for p atient safety and patients are well placed to<="" span="" style="font-family: "Segoe UI"; color: rgb(192, 0, 0); background-color: rgb(255, 255, 255);">report on these dimensions.
A further three items were excluded from factor analysis according to accepted definitions for exclusio n of items in factor analysis (multiple factor loading, communality <0.3 and
>15% missings (in this case because the item contained a „does not apply"-
option that was as missing)). We decided to keep them because they were considered important by e xperts and patients in cognitive interviews and were answered by nearly all respondents. We present t he reasons for exclusion in Online
Supplement 2. Excluding these items would have meant to risk reducing the content validity of the qu

estionnaire. Content validity is considered a very important criterion in patient-reported outcome but also experience measures.
The remaining three items referred to outcomes and were excluded from factor analysis because the percentage of missings exceeded 15%
(we explained the reason for including both items on influencing factors and outcomes in our last resp onse) and we also decided to keep them. We report this in the second paragraph of „Exploratory facto r analysis“ in the results section and we discuss this along with the fact that PSE-
frequencies collected with our questionnaire differ from existing evidence and that further research is needed in how to adequately assess frequency of PSE. See
last paragraph of the discussion section: „In contrast to existing measures (2,15), experience of PSEs and resulting harm was assessed with two generic questions. Both had to be excluded from further psychometric analyses because more than 15% of the patients had either not answered the question or were not sure whether what they had experienced represented a PSE. The proportion of patients who did experience a PSE was low (2,9%) compared to studies using more detailed measures. However, these studies did either not use a self-report measure (2,34) or had high proportions of patients skipping the detailed section on PSE experience and a low overall response rate (29). Further research is needed in how to adequately assess number and type of experienced events in routine measurements.”
However, it’s really confused me that the 22-
item questionnaire showed ‘very good acceptability, good construct validity……’ but based on a 14-item and 3-factor analysis.
Assessment of acceptability is not based on the results of the factor analysis but
on the survey response rate and further acceptabiity criteria which we describe in the results section: All items had answers that included the full range of the response scales. Median item non-response rate was 4.2% (interquartile range 3.4% to 4.7%) which is considered low to moderate [25]. Missing data did not increase with question number.
We appreciate that the sentence in the discussion section may have contributed to confusion and therefore added a sentence in the discussion section to make this clearer.
The assessment of „good construct validity, very good to acceptable levels of internal consistency and good discriminant validity“ is based on accepted criteria for factor analysis results that we define and reference in the method section when we describe the data analysis.
For the methodologically and logically consistent, maybe a final
14 items questionnaire were more acceptable. Alternatively, if the authors want to keep the 22 items q uestionnaire, psychometric properties analysis based on the 22-item data may be more reasonable.
An argument could be made for reducing the questionnaire to only the 14 items that make up the thre e factors. However, we think it is important to consider that a patient experience measure needs to be distinguished from a psychometric test and that an important criterion for the former is its content vali dity which is the reason for keeping the items that had to be excluded from factor analysis. We discus s the keeping of the single-item-
scales in the last two sentences at the end of the second paragraph in the discussion section:
The dimensions access and care coordination were assessed by single items and did not result in a factor. Given that these dimensions are also considered highly relevant for patient safety and patients are well placed to report on these dimensions, they were kept in the questionnaire [30, 31]. An argument can be made for strengthening the single-item-scales by adding further items covering these dimensions. However, this would be a trade-off between questionnaire length and detail that also needs to be considered with regard to practicability and acceptability.
We discuss the keeping of the outcome items in the last paragraph of the discussion section:
In contrast to existing measures [2, 15], experience of PSEs and resulting harm was assessed with two generic questions. Both had to be excluded from further psychometric analyses because more than 15% of the patients had either not answered the question or were not sure whether what they had experienced represented a PSE. The proportion of patients who did experience a PSE was low (2,9%) compared to studies using more detailed measures. However, these studies did either not use

a self-report measure [2, 34] or had high proportions of patients skipping the detailed section on PSE experience and a low overall response rate [29]. Further research is needed in how to adequately assess number and type of experienced events in routine measurements.