


RESEARCH ARTICLE

A methodological study on the combination of qualitative and quantitative methods in cognitive interviewing for cross-cultural adaptation

Miao Jia^{1,2} | Yao Gu³ | Yanhua Chen⁴ | Jinyi Tu^{1,2} | Yue Liu^{1,2} | Huiqi Lu^{1,2} |
Simin Huang^{1,2} | Jinjun Li^{1,2} | Hongzhen Zhou^{1,2} 

¹Department of Nursing, Nanfang Hospital, Southern Medical University, Guangzhou, China

²School of Nursing, Southern Medical University, Guangzhou, China

³Rehabilitation department of traditional Chinese medicine, Baiyunshan Hospital, Guangzhou, China

⁴Chronic disease management outpatient, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China

Correspondence

Hongzhen Zhou, Department of Nursing, Nanfang Hospital, Southern Medical University, No.1838 Guangzhou Avenue North, Guangzhou 510515, China.
Email: zhouhz_smu@126.com

Funding information

This work was supported by grants from the Science and Technology Program of Guangzhou, Guangdong, China (201803010084) and the Evidence-Based Nursing program of Nanfang Hospital, Guangdong, China (2021EBNa002) in the decision to submit the article for publication.

Abstract

Aim: The aim of this study is to explore the use of the Questionnaire Appraisal System with a combination of qualitative and quantitative methods in cognitive interviewing for cross-cultural adaptation.

Design: This is a descriptive methodological study.

Methods: Using the Mandarin version of the Post-Stroke Checklist as an example, cognitive interviews were conducted with 27 stroke survivors in Guangzhou between November 2020 and February 2021. The Questionnaire Appraisal System was applied as a codebook in focus group discussions to perform quantitative data collections and quantitative content analysis.

Results: Thirty-eight problems were proposed in focus group discussions and identified all but four of the 30 questions that emerged in the cognitive interviews. A new item was added to the Questionnaire Appraisal System for better categorization. Four categories and six subcategories of problems in the checklist were revealed.

KEYWORDS

cognitive interviewing, cross-cultural adaptation, focus groups, methodological studies, nursing

1 | INTRODUCTION

Guidelines for the translation and cross-cultural adaptation of survey questionnaires have been published, and cognitive interviewing is recommended to evaluate survey questionnaires and improve their design (Willis, 2005). However, although cognitive interviews might detect the presence of problems, they cannot identify the source of the problems and quantify them. Without hard evidence concerning the best way to conduct cognitive interviews, the strategies applied have varied from person to person, and different analysis methods

have been used based on qualitative methodologies. Studies on the combination of qualitative and quantitative methods in cognitive interviewing for cross-cultural adaptation are needed (Andersen et al., 2014).

2 | BACKGROUND

Cognitive interviewing represents an interdisciplinary fusion of the survey methodology and cognitive psychology. This method may

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Nursing Open* published by John Wiley & Sons Ltd.

help users identify the manner in which targeted subjects understand, mentally process and respond to material, with a special emphasis on the potential bias; in this way, cognitive interviewing can uncover problems that previously may have been unnoticed to minimize the bias and other classic questionnaire errors before field testing (Drennan, 2003; McColl et al., 2003; Willis, 2005). Thinking aloud and verbal probing are the most commonly used techniques in cognitive interviews. When the thinking aloud method is used for pretesting, subjects are encouraged not only to answer but also to vocalize all of their thoughts as they respond to the questionnaire, which always involves some practice at the start of an interview. Then, interviewers read each question aloud and record the subjects' verbal stream of thought or otherwise note the processes subjects use in arriving at an answer. During verbal probing, interviewers ask the targeted question and subjects answer it, but interviewers then follow up by probing for other specific information relevant to understanding how information is recalled, how subjects reached their answers and what the response categories are (Willis, 2005, 2006). Most importantly, scripted probes must be developed to ensure the successful process of the study. Card sorting is another technique to determine what subjects think; this method is used to determine how individuals organize concepts and, in particular, what they believe a concept includes or excludes (Willis, 2005). For example, interviewers could use a card sort task to investigate subjects' perceptions of medical conditions by instructing subjects to sort cards containing names of chronic diseases into piles that "go together."

Although researchers generally agree with the unique advantages of cognitive interviews (Gawad et al., 2021; Hewlett et al., 2016), a research methodology to guide best practice has not been fully established. Major problems, such as sample size, iteration cycles, participant characteristics, probes employed and data analysis, remain unresolved (Lee, 2014). The processes related to the problems, such as appraising items to find problems, developing probes based on problems and analysing items reported with problems, are inconsistent. The strategies applied have varied from person to person, and different analysis methods have been used based on qualitative methodologies (Eeg-Olofsson et al., 2020; Gerlach et al., 2020; Simwanga et al., 2019). However, there are key differences between cognitive interviews and qualitative analysis approaches (Cassidy et al., 2020). The analysis of cognitive interviews is conducted item by item, with the aim of revealing problems with each question in the questionnaire; meanwhile, qualitative analysis approaches involve in-depth and holistic data excavation across the whole interviews to uncover the sources of problems in the questionnaire, which is not necessary in some ways because the sources of problems in cognitive interviews are well defined by the four-stage cognitive model (Tourangeau, 1984). When qualitative methods detect the presence of problems, they cannot distinguish the source of the problems and the quantity of problems (Andersen et al., 2014). Due to the lack of standardized methods, cognitive interviewing has been criticized for being too subjective (Conrad & Blair, 2009). For these reasons, research is needed on how to combine qualitative and

quantitative methods in cognitive interviews (Andersen et al., 2014; Malterud, 2001). Historically, focus group discussions (FGDs) have been used as part of a methodical approach in which both qualitative and quantitative data are collected with the aim of further refining and interpreting previously gathered data (Stalmeijer et al., 2014), which is consistent with our purpose. Thus, we valued the role of FGDs in this study.

The Post-Stroke Checklist (PSC) is an 11-item checklist administered by healthcare professionals to standardize the process of identifying long-term care and facilitating referrals (Philp et al., 2013). It has been proven to be useful and feasible and has gained international recognition, being endorsed by the World Stroke Organization to improve the follow-up care of poststroke individuals. After obtaining permission to translate this measure, we developed a Mandarin version of the PSC (M-PSC) in keeping with guidelines (Wild et al., 2005). The Questionnaire Appraisal System (QAS-99) is designed to assist questionnaire designers in evaluating survey questionnaires and in finding and fixing problems before the questions "go into the field," and it is recommended for use in cognitive interviews (Willis & Lessler, 1999). This study explores the use of the QAS-99 with a combination of qualitative and quantitative methods in cognitive interviewing for cross-cultural adaptation, using the M-PSC as an example, and demonstrates a consistent approach to the processes related to problems in cognitive interviewing.

3 | THE STUDY

3.1 | Design

We present a descriptive methodological study that uses semistructured interviews and FGDs. The QAS-99 is applied as a codebook to provide quantitative information on categories and frequencies of problems identified in cognitive interviews and FGDs. Consolidated criteria for reporting qualitative research (COREQ) (Tong et al., 2007) were followed to ensure standardized reporting (see Supporting Information S1).

3.2 | Method

3.2.1 | Participants

Participants were recruited in Guangzhou from (1) a stroke rehabilitation outpatient department of a tertiary hospital in Yuexiu District, (2) a rehabilitation department using traditional Chinese medicine of a secondary hospital in Baiyun District and (3) a community in Baiyun District. Individuals were eligible to participate if they had experienced a cerebral infarction or intracerebral haemorrhage and had been more than 6 months since the most recent stroke. Those with cognitive impairment (Montreal Cognitive Assessment score <26), native Cantonese speakers who were not fluent in Mandarin and

those with dysphasia were excluded. Purposive sampling was used to achieve the maximum heterogeneity regarding education level, gender, age, occupation and marital status. We sampled from three different areas for this purpose. One set of nine practice rounds, where the number of rounds of iteration was determined based on the saturation of information, was conducted (Willis, 2005). Two researchers (Y. G. and Y. H. C.) were in charge of recruitment in the two rehabilitation departments, and a resident in the community in Baiyun District helped us to invite eligible individuals to participate in this study. Twenty-seven eligible participants showed their willingness to participate when they were contacted by the two researchers. Another two invited individuals rejected taking part in; one did not agree to recordings or notes during the interview, whereas the other did not admit to having a stroke.

3.2.2 | Data collection

All of the cognitive interviews were conducted by the first author (M. J.) from November 2020–February 2021. The first author was trained and had prior interview experience and performed each interview based on a cognitive testing protocol that was created from the FGD. The focus group consisted of one facilitator, three researchers, two clinical nurses and one observer. The QAS-99 was used as a codebook to provide quantitative information on problems of the M-PSC. Two researchers (J. Y. T. and Y. L.) were trained and given a copy of the M-PSC and some QAS-99 coding forms. The facilitator asked them to indicate whether the problem was present in the M-PSC by circling Yes or No on the form and, for each circled, note the reason a Yes code was assigned. The results were discussed to achieve uniformity, and then appropriate probe questions and the cognitive testing protocol were generated based on the results. The first author used the protocol to simulate a cognitive interview with another researcher (H. Q. L.) who played the role of the stroke survivor. The interviewer (M. J.) was not involved in the discussion. Meanwhile, two clinical nurses joined the discussion, and the facilitator asked the group to evaluate deficiencies of the protocol and the interviewer's behaviour. An observer was present during the whole discussion to take notes and ask questions to clarify issues raised if necessary. After receiving the protocol, the first author introduced the study to the participants, obtained written informed consent and practised helping them think aloud. In this case, they were asked "How many windows are there in this room?" and "Please tell me the details about how you came up with this answer." Then, a semistructured interview was performed using concurrent probing and immediate retrospective probing, and card sorting was also used in Q2b. No one else was present besides the participant and the interviewer, and no prior relationship existed between them. The first author took descriptive field notes during the interview, looked back over the protocol to fill in missing information at the end of each interview, organized these materials and sent them to the focus group. All interviews were audio recorded and conducted either at the hospital in an undisturbed office or by WeChat.

3.3 | Analysis

Two researchers (S. M. H. and J. J. L.) transcribed the audio recording of each interview verbatim, whereas another researcher (H. Q. L.) rechecked the differences between the two transcripts and corrected the major problems against the recording. The QAS-99 was applied as the codebook in quantitative content analysis in FGDs. Two trained researchers (J. Y. T. and Y. L.) independently coded all the corrected transcripts and reached an agreement with the guidance of the facilitator. The item of the M-PSC was modified if five participants found the question to be difficult to understand. The item appraisal, probe development, simulation testing and evaluation were subsequently redone, and the results guided further data collection. When five participants consecutively reported no problems or five of nine individuals did not report any issues, the item was modified appropriately.

3.4 | Trustworthiness

Triangulation was adopted to meet the criteria for credibility, and a thick description was provided to ensure transferability (Korstjens & Moser, 2018). Two researchers (Y. G. and Y. H. C.) were selected as quality control supervisors due to their years of experience in long-term follow-up stroke care. They were responsible for rechecking all the materials before the research entered the next stage. Several strategies were used to ensure that the enrolled participants met all eligibility criteria (Whitney et al., 1998). The information on the participants was checked by one of the researchers (Y. G.) because the recruiter in the community lacked medical knowledge. In addition, the interviewer (M. J.) conducted a maximum of three interviews a day to maintain energy (Willis, 2005).

3.5 | Ethics

This study was approved by the Ethics Committee of Nanfang Hospital (No. NFEC-2018-054) and proceeded after clearances have been obtained from all the settings involved. Participants were fully informed about the aim and procedure before the start of this study. Anonymity and confidentiality were ensured, and written informed consent was obtained from all participants.

4 | RESULTS

4.1 | Participant characteristics

Twenty-seven participants, including 24 individuals with ischaemic stroke, 1 individual with haemorrhagic stroke and 2 mixed stroke survivors, took part in the cognitive interviews. The participants consisted of 20 men and 7 women, and they had an average age of 66 years (range: 46–92 years). The median time since the last stroke

was 36 months (range: 6–276 months). Three of the participants were widowed, one was divorced and the others were married. Their professions varied from office worker, nurse, teacher, labour, civil servant, accountant, businessman, railwayman, pharmacist to unemployed. The median time of schooling was 9 years (range: 6–16 years), and two participants were never educated. The average time of the recorded interviews was 27 min (range: 16–47 min).

4.2 | Cognitive interviewing

Thirty-eight problems were proposed based on the QAS-99 in FGDs, which identified all but four of the 30 questions that emerged in the cognitive interviews (Table 1). Each item of the M-PSC was modified at most five times and at least once. One item was added, and one item was changed only in terms of the pronoun. With 6 of the 38 problems not up to standard, 87% of the problems identified in the cognitive interviews were identical to the problems raised in FGDs. We also added a new item, “Phonetic words,” to the QAS-99 for the better categorization of interviewer-administered questionnaires. Readers can find detailed descriptions of all original items of the M-PSC and identified problems as well as suggested revisions in Supporting Information S2.

4.2.1 | Category 1: Clarity

Wording

Some participants described difficulty with lengthy items.

TABLE 1 The coding results of the QAS-99

Categories	Subcategories	Frequencies of problems (n)	
		FGDs	CI
Clarity	Wording	3	2
	Technical terms	14	6 ^a
	Phonetic words		2
Assumptions	Inappropriate assumptions	11	11
	Double-barrelled items	7	6
Sensitivity/bias	Sensitive content	3	1
Other problems		0	4
Total		38	30
Reading	Not identified in FGDs or CI		
Instructions			
Knowledge/memory			
Response categories			

Note: Abbreviations: CI, cognitive interviews; FGDs, focus group discussions; QAS-99, question appraisal system.

^aTwo technical terms were eventually categorized as phonetic words after cognitive interviews, and the frequency was not counted in total.

Concerning the item “Advice on health related lifestyle changes, such as diet and exercise” (modified version of Item 1), one participant reported he was confused in grasping the meaning, whereas the other four had challenges interpreting it. Specifically, two participants could not remember the entire question, and the other two indicated that the question was easy to respond to but hard to repeat. “It’s easy to answer, but I can’t repeat it. It’s a little difficult to repeat this question” (P13). Shortening the sentence to “Lifestyle changes, such as diet and exercise” solved this problem.

The M-PSC provides examples to make it easier to define “things important to you” (Item 10). However, the participants found that their understanding of this item was limited by excessive and inappropriate examples. “I think it would be better to ask directly with no examples” (P4). “I don’t know how to answer this question because my husband passed away” (P5). We deleted parts of the examples to address this problem, and only the examples “work” and “hobbies” were retained.

Technical terms

Participants frequently had difficulty interpreting technical terms.

In the first round, five of the eight participants found that the translated term for “lifestyle changes” was undefined. Some reported difficulties in comprehending the term, whereas the others interpreted it to mean only either “food” or “exercise.” “That means a better diet, for example, eating light foods” (P1). The participants also failed to include “rehabilitation therapy” in their answers. For example, some participants considered walking and medications to be “rehabilitation therapy.” “In my understanding, the drug that can cure me is the rehabilitation therapy. The medicine I take now can just control disease, but not cure it. I understand it in this way” (P14). Some participants from the community did not even know what rehabilitation therapy was. “I don’t know what it means. I have never seen that” (P12). We solved the two problems above with an explanation or, more precisely, examples to make them clear.

Wrong references to body parts were also reported. The participants described the arm to mean only the upper arm. “(Pointing at the upper arm) This is the arm ... but that is the forearm, not the arm” (P7). Several participants also had problems regarding the parts of the leg. “Usually, I think the leg just means the thigh” (P4). In addition, a phonetic problem occurred with this item, so we handled it comprehensively. A concise illustration was made to replace the words in Chinese.

The participants indicated that the item “Do you feel anxious or depressed?” was easy to understand and respond to, but they had difficulty explaining the technical terms, especially the translated word “depression.” “Depression means ... uh ... I don’t know how to explain it, but I make sure I never suffered from it. I know what that means, but I have no idea how to explain it” (P8). An explanation was added to address this issue.

Additionally, the participants showed challenges understanding the word “participation.” “I am not clear what it means” (P13). In addition, words in Chinese, such as “medication suggestions,” “activities of daily living,” “daily living,” “activity” and “anxiety,” were challenging to some participants, even though they were unmodified. “That depends on how you interpret it. I don’t know what you are implying.

Is it the aspect of the labour or the family?" (P16) In particular, we considered the translated term "new pain" to be an understanding issue at first; however, it was identified as a phonetic problem by cognitive interviews.

Phonetic words

There were problems concerning phonetic words in Chinese when the participants were thinking aloud.

For example, the translated term "new pain," which is also a technical term in Chinese, was reported to sound like "heartache." "Of course, from the family, and the small society ... which means the emotional baggage" (P2). "(Pointing at his heart) I think you mean angina, do you?" (P4) We expanded the phrase into a sentence to make it easy to understand and clear.

Meanwhile, the word "arm" in Chinese was misheard as "the back of the hand," "the coldness of the hand" and "the paralysis of the hand." "That means your hand feels cold, which is called the 'the coldness of the hand'" (P10). A concise illustration was made as stated above.

4.2.2 | Category 2: Assumptions

Inappropriate assumptions

An inappropriate assumption was made about the "last assessment."

Five consecutive participants reported this issue when they were exposed to the M-PSC for the first time. "After the stroke, I ... as the 'last assessment', I have no idea about it ... what do you mean by that?" (P2) In addition, although used in all items and probes, we viewed this as one question and just asked the probe in Item 1. We addressed this problem by removing the "last assessment" from all items.

Double-barrelled items

Participants found it difficult to understand these items because they contained more than one implicit question.

Seven items of the M-PSC had this problem, and to our great surprise, we easily identified all seven items using the QAS-99. A typical example is as follows: "Since your stroke or last assessment, are you finding it more difficult to think, concentrate or remember things?" Five successive participants indicated that they were puzzled about how to answer such an overlapping question. "It's strange ... uh, I don't know how to answer it. My memory has gradually blurred these years, but I never found it was difficult to think and concentrate. I am not diagnosed with Alzheimer's disease, ha-ha (laughing)" (P5). Furthermore, the double-barrelled questions made the items lengthy. "The sentence is too long, and I think it would be better to disjoin it" (P3). Item decomposition was applied to solve this problem.

4.2.3 | Category 3: Sensitivity/bias

Sensitive content

The participants identified the family relationship as a sensitive topic, and two of them even cried after we talked about this problem.

"I'm afraid the disease might bring pressure to my son. His work and life might be influenced (sobbing)" (P19). We suggest that this item be treated according to the role of the medical personnel administering the questionnaire. If the individual is close to the family and can do something to improve the relationship, we strongly recommend asking this item and taking action. Otherwise, we recommend skipping this item as we did in the third iteration.

4.2.4 | Category 4: Other problems

Four problems not previously identified in FGDs were reported.

Contrary to our expectation, the participants failed to understand the meaning of "medications for preventing another stroke." "I don't know what it means ... I just take medicine as advised by the doctor" (P12). In addition, one participant considered medications to be the injections given in the hospital during the acute period. "I think it means the medication that can be used to treat the stroke. I didn't take the imported medicine in the hospital, although others were injected" (P10). We settled this problem by altering the term to the common phrase "medicine suggestions" in Chinese.

The participants reported that stiffness in their body parts was relieved gradually after rehabilitation training. "Nope, it becomes better instead, and that is the most improvement" (P3). We addressed this problem by adding the new item "Does the stiffness in [BODY PARTS] improve now?" and replacing the word "this" with "the stiffness in [BODY PARTS]" to the question "Is this interfering with activities of daily living?" to make the sentence read smoothly in the Chinese context.

The translation error for "preparing hot drinks" was very surprising. Three participants described this item as "hot tea," "boiled water" or "hot soft drinks" individually, whereas two participants reported they were completely puzzled about it. "Preparing hot drinks? Yes, it would be better to drink boiled water, not cold water" (P4). In fact, we had not considered this term to be a translation error before the quality control supervisors discussed it with us. They proposed that participants normally were not linguists, and they certainly could not describe what it meant in detail. A common and simple translation was needed. Thus, two phrases, "boiling water" and "pouring boiled water," were used to replace the original translation.

After all the examples had been deleted to make Q10 brief, the participants were confused with the definition of "things important to you." Thus, the examples "work" and "hobbies" were left.

5 | DISCUSSION

Using the M-PSC as an example, this study describes the active role of the QAS-99 in combination with qualitative and quantitative methods in cognitive interviews for cross-cultural adaptation. We used the QAS-99 as a codebook to provide quantitative information identified in FGDs and cognitive interviews. Meanwhile, the connecting link between the preceding and the following in processes related to problems was formed in this way.

Although many published studies consider expert reviews to be a pretesting method for item appraisal (Klarare et al., 2021; Wu et al., 2020; Zou et al., 2019), quantitative methods have rarely been used to appraise the readability and wording of questionnaires in cross-cultural adaptation (Egger-Rainer, 2019; Rapley et al., 2018). Although academically or professionally oriented experts often know the participants in the study well, they may not be very good questionnaire designers. The empirical opinions of experts must be accepted with discretion and considered to be modifiable based on other objective testing results (Willis, 2005). By employing the QAS-99, we conducted a quantitative item appraisal as an independent procedure in addition to an empirical expert review. The results of item appraisal guided the next process: probe development. Developing appropriate probe questions is not a small challenge for researchers. Some authors of published studies have shared the probes they applied, but there is very little documentation on how to formulate anticipated probes (Dabbagh et al., 2020; Flythe et al., 2019; Li et al., 2018). According to the results of item appraisals, probes can be easily developed using the QAS-99. Traditionally, researchers have devoted themselves to developing systematic methods of analysis based on qualitative strategies in dramatically different ways (Eeg-Olofsson et al., 2020; Gawad et al., 2020; Zarnowiecki et al., 2020), which has made it difficult to further compare and analyse. Some authors have suggested a thematic approach in the testing process. Simwinga et al. (2019) identified four categories after analysing transcripts from cognitive interviews, Wiegers et al. (2019) classified data into eight themes and Furtado et al. (2020) identified six categories in the same way. However, the categories varied from one another. In other studies, data were analysed deductively based on the four-stage cognitive model: comprehension, information retrieval, judgement and response process (Gerlach et al., 2020; Halls et al., 2018; Murphy et al., 2018). Although predefined themes were established, researchers also had to engage in classifying the problems into the right theme. Case analysis and content analysis have also been employed in some studies (Andersen et al., 2014; Eeg-Olofsson et al., 2020). Although qualitative methods might detect the problems, they cannot distinguish the source of the problems and quantify them (Andersen et al., 2014). At the same time, qualitative analysis methods may cost researchers considerable time and energy. Quantification of phenomena or categories could be performed to gain a better understanding of qualitative data (Malterud, 2001). In this study, the QAS-99 was applied as a codebook in quantitative content analysis to deal with data from cognitive interviews; in this way, effortless data analysis was performed. In addition to inappropriate assumptions that were present in each item, double-barrelled items and technical terms were the most frequent problems. Two technical terms were eventually categorized as phonetic similarity issues after cognitive interviews. Therefore, we added a new item, "phonetic words," to the QAS-99 for the sake of better categorization for interviewer-administered questionnaires. The problems related to reading and instructions were not identified; it might be attributed to expert reviews correcting these problems. However,

the implementation of expert reviews did not mean that these two problems would not be reported in the cognitive interviews. Similarly, the categories of the problems not identified in this study did not necessarily appear in other survey questionnaires. In addition to the newly added subitem, all items and subitems of the QAS-99 were retained to wait for further validation. Additionally, this study provides evidence that interviews with members of the target population to better characterize and assess participants' cognitive processes may result in contrary item changes, even in direction, which was not found in previous work (Kjörk et al., 2019; Turner et al., 2019; Ward et al., 2014).

The sample size in cognitive interviewing has been the subject of enduring discussion (Turner-Bowker et al., 2018). Current practices are based on the assumption that the most critical questionnaire issues can be revealed by a small sample (Beatty & Willis, 2007). One study challenged this assumption and claimed that significant issues continued to be uncovered even after 50 or more cognitive interviews (Blair et al., 2006). Another study also questioned this practice and proposed that cross-cultural studies often included a large number of interviews (Willis, 2015). It is better to run several iterations with a small number of participants than one or two iterations with a large number (Willis, 2005). Some practices have included employing approximately 5–15 subjects in a round to conduct cognitive interviews (Lee, 2014; Willis, 2005). The iterative approach was proven to be useful, but it was not clear how many rounds were usually needed. Some qualitative researchers have recommended that the sample size should be sufficient to reach theoretical saturation with a range between 20 and 30 interviews (Turner-Bowker et al., 2018; Vasileiou et al., 2018). We suggest, and applied in this study, one set of nine practice rounds, where the number of rounds of iteration is determined based on saturation of information. Participants recruited from outpatients and communities in the same district may have similar characteristics. To offer sufficient information power (Malterud et al., 2016), we chose an outpatient from a different district with the community. Again, we state the importance of the FGDs. Our experience in this study suggests that FGDs are a productive line of research that may benefit cognitive interviews. Indeed, we take the view that interviewers should be excluded from the discussion, seeing that their assumptions in the interviews may influence the results of the analysis.

Further research is also needed to explore how to define saturation in cognitive interviews, including items and alternative subitems. Few studies have provided empirical data on how saturation was achieved in cognitive interviews using qualitative or quantitative methods (Hennink et al., 2017). Willis (2005) proposed that a question should be modified when five of nine participants find the question to be difficult to understand. A related problem is as follows: Should we set the same criteria for items and alternative subitems? There is no doubt that more informants are required to reach saturation for alternative subitems because researchers cannot forecast which participant will be asked the alternative subitems. A prime example, which we also came across in this study, is when all the items reached saturation except for one alternative subitem, and

four consecutive participants reported no problems with it. Is it necessary to run another iteration or should the iterations be stopped to avoid waste? A further discussion on these matters is needed.

5.1 | Limitations

This study has several limitations. First, the interviews were conducted in normal settings, not a dedicated cognitive laboratory. Second, the participants in this study were untrained, and it was their first time taking part in cognitive interviews. It is suggested that in further studies, laboratory testing be conducted with experienced volunteers if possible, although this is not a critical requirement for cognitive interviewing. Third, this was a single study adapting a single interviewer-administered checklist. Further research is needed to explore the generalizability of the findings for both interviewer-administered questionnaires and self-administered questionnaires. Finally, the M-PSC should be validated with additional research.

6 | CONCLUSION

We demonstrated the active role of the QAS-99 in combination with qualitative and quantitative methods in cognitive interviews for cross-cultural adaptation and formed a connecting link between the preceding and the following in processes related to problems using the QAS-99. Further research may focus on the issues of saturation in cognitive interviewing and the effect of the combined approach for interviewer-administered questionnaires and self-administered questionnaires.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Michael Brainin for permitting to translate the PSC, to the recruiter Bai and 27 participants who took part in this study, to the Nanfang Hospital and Guangzhou Municipal Science and Technology Bureau for supporting this study.

CONFLICT OF INTEREST

No conflict of interest has been declared by the authors.

AUTHOR CONTRIBUTIONS

MJ conducted interviews. YG and YHC were involved in recruitment and quality control. SMH and JLL transcribed interviews, with HQL checking transcripts. Data analysis and interpretation were led by HZZ, JYT and YL, with HQL, SMH and JLL contributing. MJ wrote the manuscript and all other authors critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript and agree accountability for all aspects of the work in ensuring investigating and resolving questions related to the accuracy or integrity of any part of the work.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the Supplementary File.

ORCID

Hongzhen Zhou  <https://orcid.org/0000-0001-7740-6779>

REFERENCES

- Andersen, R. D., Jylli, L., & Ambuel, B. (2014). Cultural adaptation of patient and observational outcome measures: A methodological example using the COMFORT behavioral rating scale. *International Journal of Nursing Studies*, 51, 934–942. <https://doi.org/10.1016/j.ijnurstu.2013.10.006>
- Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71, 287–311. <https://doi.org/10.1093/poq/nfm006>
- Blair, J., Conrad, F., Ackermann, A. C., & Claxton, G. (2006). *The effect of sample size on cognitive interview findings*. American Association for Public Opinion Research Conference.
- Cassidy, S. A., Bradley, L., Cogger-Ward, H., Shaw, R., Bowen, E., Glod, M., Baron-Cohen, S., & Rodgers, J. (2020). Measurement properties of the suicidal behaviour questionnaire-revised in autistic adults. *Journal of Autism and Developmental Disorders*, 50, 3477–3488. <https://doi.org/10.1007/s10803-020-04431-5>
- Conrad, F. G., & Blair, J. (2009). Sources of error in cognitive interviews. *Public Opinion Quarterly*, 73(1), 32–55. <https://doi.org/10.1093/poq/nfp013>
- Dabbagh, A., MacDermid, J. C., Packham, T. L., & Macedo, L. G. (2020). Content validation of the Kamath and Stothard questionnaire for carpal tunnel syndrome diagnosis: A cognitive interviewing study. *Health and Quality of Life Outcomes*, 18, 359. <https://doi.org/10.1186/s12955-020-01614-7>
- Drennan, J. (2003). Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing*, 42(1), 57–63. <https://doi.org/10.1046/j.1365-2648.2003.02579.x>
- Eeg-Olofsson, K., Johansson, U. B., Linder, E., & Leksell, J. (2020). Patients' and health care professionals' perceptions of the potential of using the digital diabetes questionnaire to prepare for diabetes care meetings: Qualitative focus group interview study. *Journal of Medical Internet Research*, 22(8), e17504. <https://doi.org/10.2196/17504>
- Egger-Rainer, A. (2019). Enhancing validity through cognitive interviewing. A methodological example using the epilepsy monitoring unit comfort questionnaire. *Journal of Advanced Nursing*, 75, 224–233. <https://doi.org/10.1111/jan.13867>
- Flythe, J. E., Dorough, A., Narendra, J. H., Wingard, R. L., Dalrymple, L. S., & DeWalt, D. A. (2019). Development and content validity of a hemodialysis symptom patient-reported outcome measure. *Quality of Life Research*, 28, 253–265. <https://doi.org/10.1007/s11136-018-2000-7>
- Furtado, R., MacDermid, J. C., Bryant, D. M., Faber, K. J., & Athwal, G. S. (2020). Interpretation and content validity of the items of the numeric rating version short-WORC to evaluate outcomes in management of rotator cuff pathology: A cognitive interview approach. *Health and Quality of Life Outcomes*, 18(1), 88. <https://doi.org/10.1186/s12955-020-01339-7>
- Gawad, N., Wood, T. J., Cowley, L., & Raiche, I. (2020). The cognitive process of test takers when using the script concordance test rating scale. *Medical Education*, 54, 337–347. <https://doi.org/10.1111/medu.14056>
- Gawad, N., Wood, T. J., Cowley, L., & Raiche, I. (2021). How do cognitive processes influence script concordance test responses? *Medical Education*, 55, 354–364. <https://doi.org/10.1111/medu.14416>

- Gerlach, C., Taylor, K., Ferner, M., Munder, M., Weber, M., & Ramsenthaler, C. (2020). Challenges in the cultural adaptation of the German myeloma patient outcome scale (MyPOS): An outcome measure to support routine symptom assessment in myeloma care. *BMC Cancer*, 20, 245. <https://doi.org/10.1186/s12885-020-06730-7>
- Halls, S., Dures, E., Kirwan, J. R., Pollock, J., Baker, G., Edmunds, A., & Hewlett, S. (2018). Development and testing of candidate items for inclusion in a new rheumatoid arthritis stiffness patient-reported outcome measure. *Rheumatology (Oxford)*, 57, 263–272. <https://doi.org/10.1093/rheumatology/kex085>
- Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? *Qualitative Health Research*, 27, 591–608. <https://doi.org/10.1177/1049732316665344>
- Hewlett, S., Nicklin, J., Bode, C., Carmona, L., Dures, E., Engelbrecht, M., Hagel, S., Kirwan, J., Molto, A., Redondo, M., & Gossec, L. (2016). Translating patient reported outcome measures: Methodological issues explored using cognitive interviewing with three rheumatoid arthritis measures in six European languages. *Rheumatology (Oxford)*, 55, 1009–1016. <https://doi.org/10.1093/rheumatology/kew011>
- Kjörk, E. K., Carlsson, G., Sunnerhagen, K. S., & Lundgren-Nilsson, Å. (2019). Experiences using the poststroke checklist in Sweden with a focus on feasibility and relevance: A mixed-method design. *British Medical Journal Open*, 9(5), e028218. <https://doi.org/10.1136/bmjopen-2018-028218>
- Klarare, A., Wikman, A., Söderlund, M., McGreevy, J., Mattsson, E., & Rosenblad, A. (2021). Translation, cross-cultural adaptation, and psychometric analysis of the attitudes towards homelessness inventory for use in Sweden. *Worldviews on Evidence-Based Nursing*, 18(1), 42–49. <https://doi.org/10.1111/wvn.12477>
- Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *The European Journal of General Practice*, 24(1), 120–124. <https://doi.org/10.1080/13814788.2017.1375092>
- Lee, J. (2014). Conducting cognitive interviews in cross-national settings. *Assessment*, 21, 227–240. <https://doi.org/10.1177/1073191112436671>
- Li, C., Auld, G., D'Alonzo, K., & Palmer-Keenan, D. (2018). Communicating and assessing physical activity: Outcomes from cognitive interviews with low-income adults. *Journal of Nutrition Education and Behavior*, 50, 984–992. <https://doi.org/10.1016/j.jneb.2018.07.008>
- Malterud, K. (2001). Qualitative research: Standards, challenges, and guidelines. *The Lancet*, 358, 483–488. [https://doi.org/10.1016/S0140-6736\(01\)05627-6](https://doi.org/10.1016/S0140-6736(01)05627-6)
- Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26, 1753–1760. <https://doi.org/10.1177/1049732315617444>
- McCull, E., Meadows, K., & Barofsky, I. (2003). Cognitive aspects of survey methodology and quality of life assessment. *Quality of Life Research*, 12(3), 217–218. <https://doi.org/10.1023/A:1023233432721>
- Murphy, M., Hollinghurst, S., & Salisbury, C. (2018). Patient understanding of two commonly used patient reported outcome measures for primary care: A cognitive interview study. *BMC Family Practice*, 19(1), 162. <https://doi.org/10.1186/s12875-018-0850-2>
- Philp, I., Brainin, M., Walker, M. F., Ward, A. B., Gillard, P., Shields, A. L., & Norrving, B. (2013). Development of a poststroke checklist to standardize follow-up care for stroke survivors. *Journal of Stroke and Cerebrovascular Diseases*, 22(7), e173–e180. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2012.10.016>
- Rapley, T., Girling, M., Mair, F. S., Murray, E., Treweek, S., McColl, E., Steen, I. N., May, C. R., & Finch, T. L. (2018). Improving the normalization of complex interventions: Part 1 - development of the NoMAD instrument for assessing implementation work based on normalization process theory (NPT). *BMC Medical Research Methodology*, 18(1), 133. <https://doi.org/10.1186/s12874-018-0590-y>
- Simwina, M., Kumwenda, M. K., Dacombe, R. J., Kayira, L., Muzumara, A., Johnson, C. C., Indravudh, P., Sibanda, E. L., Nyirenda, L., Hatzold, K., Corbett, E. L., Ayles, H., & Taegtmeyer, M. (2019). Ability to understand and correctly follow HIV self-test kit instructions for use: Applying the cognitive interview technique in Malawi and Zambia. *Journal of the International AIDS Society*, 22(Suppl. 1), e25253. <https://doi.org/10.1002/jia2.25253>
- Stalmeijer, R. E., McNaughton, N., & Van Mook, W. N. (2014). Using focus groups in medical education research: AMEE Guide No. 91. *Medical Teacher*, 36, 923–939. <https://doi.org/10.3109/0142159x.2014.917165>
- Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19, 349–357. <https://doi.org/10.1093/intqhc/mzm042>
- Tourangeau, R. (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. National Academy Press.
- Turner, G. M., Mullis, R., Lim, L., Kreit, L., & Mant, J. (2019). Using a checklist to facilitate management of long-term care needs after stroke: Insights from focus groups and a feasibility study. *BMC Family Practice*, 20(1), 2. <https://doi.org/10.1186/s12875-018-0894-3>
- Turner-Bowker, D. M., Lamoureux, R. E., Stokes, J., Litcher-Kelly, L., Galipeau, N., Yaworsky, A., Solomon, J., & Shields, A. L. (2018). Informing a priori sample size estimation in qualitative concept elicitation interview studies for clinical outcome assessment instrument development. *Value Health*, 21, 839–842. <https://doi.org/10.1016/j.jval.2017.11.014>
- Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology*, 18(1), 148. <https://doi.org/10.1186/s12874-018-0594-7>
- Ward, A. B., Chen, C., Norrving, B., Gillard, P., Walker, M. F., Blackburn, S., Holloway, L., Brainin, M., & Philp, I. (2014). Evaluation of the post stroke checklist: A pilot study in the United Kingdom and Singapore. *International Journal of Stroke*, 9(Suppl. A100), 76–84. <https://doi.org/10.1111/ij.s.12291>
- Whitney, C. W., Lind, B. K., & Wahl, P. W. (1998). Quality assurance and quality control in longitudinal studies. *Epidemiologic Reviews*, 20(1), 71–80. <https://doi.org/10.1093/oxfordjournals.epirev.a017973>
- Wieggers, T. A., Hendriks, M., Malanda, U., & de Boer, D. (2019). Users' experiences with web-based health care information: Qualitative study about diabetes and dementia information presented on a governmental website. *Journal of Medical Internet Research*, 21(7), e11340. <https://doi.org/10.2196/11340>
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value Health*, 8(2), 94–104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x>
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Sage.
- Willis, G. (2006). Cognitive interviewing as a tool for improving the informed consent process. *Journal of Empirical Research on Human Research Ethics*, 1(1), 9–24. <https://doi.org/10.1525/jer.2006.1.1.9>
- Willis, G. B. (2015). The practice of cross-cultural cognitive interviewing. *Public Opinion Quarterly*, 79, 359–395. <https://doi.org/10.1093/poq/nfu092>
- Willis, G. B., & Lessler, J. (1999). *The BRFSS-QAS: A guide for systematically evaluating survey question wording*. Research Triangle Institute.

- Wu, J., Xie, S., He, X., Chen, G., & Brazier, J. E. (2020). The simplified Chinese version of SF-6Dv2: Translation, cross-cultural adaptation and preliminary psychometric testing. *Quality of Life Research*, 29, 1385–1391. <https://doi.org/10.1007/s11136-020-02419-3>
- Zarnowiecki, D., Byrne, R. A., Bodner, G. E., Bell, L. K., & Golley, R. K. (2020). Improving the reporting of young children's food intake: Insights from a cognitive interviewing study with mothers of 3–7-year old children. *Nutrients*, 12, 1645. <https://doi.org/10.3390/nu12061645>
- Zou, J., Liu, Y., Yang, Q., Liu, H., Luo, J., Ouyang, Y., Wang, J., & Lin, Q. (2019). Cross-cultural adaptation and validation of the Chinese version of the child food neophobia scale. *British Medical Journal Open*, 9(8), e026729. <https://doi.org/10.1136/bmjopen-2018-026729>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Jia, M., Gu, Y., Chen, Y., Tu, J., Liu, Y., Lu, H., Huang, S., Li, J., & Zhou, H. (2022). A methodological study on the combination of qualitative and quantitative methods in cognitive interviewing for cross-cultural adaptation. *Nursing Open*, 9, 705–713. <https://doi.org/10.1002/nop2.1121>