J Caring Sci. 2026;15(1):52-60
doi:10.34172/jcs.026.36566
https://jcs.tbzmed.ac.ir

Original Article

Enhancing Face Validity Using Cognitive Interviewing Techniques: Applying Tourangeau's Framework to the Nursing Talent Identification Scale

Farzaneh Bagheriyeh[1], Akram Ghahramanian[1,2*], Vahid Zamanzadeh[3], Leila Valizadeh[4], Behrouz Amini Kordkandi[2]

[1]Department of Medical Surgical Nursing, School of Nursing and Midwifery, Tabriz University of Medical Sciences, Tabriz, Iran
[2]Clinical Research Development Unit of Tabriz Valiasr Hospital, Tabriz University of Medical Sciences, Tabriz, Iran
[3]Department of Medical Surgical Nursing, School of Nursing and Midwifery, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[4]Department of Pediatric Nursing, School of Nursing and Midwifery, Shahid-Beheshti University of Medical Sciences, Tehran, Iran

Article Info

Article History:
Received: November 8, 2025
Revised: February 3, 2026
Accepted: February 24, 2026
ePublished: March 23, 2026

Keywords:
Cognitive interviewing, Face validity, Self-report instrument, Tourangeau's framework, Instrument development

***Corresponding Author:**
Akram Ghahramanian,
Email: ghahramaniana@gmail.com

Abstract

Introduction: Although cognitive interviewing is an efficient method for assessing and enhancing the face validity of self-report instruments, the lack of a structured framework can limit its effectiveness. This study aims to introduce and apply Tourangeau's Question-and-Answer Framework as a systematic guide for conducting cognitive interviews and enhancing face validity.

Methods: This methodological study was conducted in two phases. First, Tourangeau's theoretical framework comprising the four cognitive stages of comprehension, retrieval, judgment, and response was used to structure interview probes. Second, this framework was applied in a case study to improve the Nursing Talent Identification Scale. Cognitive interviews using think-aloud and verbal probing techniques were conducted with 20 first-year undergraduate nursing students selected via purposive sampling. Data were analyzed through qualitative content analysis explicitly guided by Tourangeau's stages.

Results: The application of Tourangeau's framework enabled the precise identification and categorization of item-level issues. Analysis of the initial 95-item pool revealed problems primarily related to ambiguous wording and misinterpreted intent. Most issues were traced to the comprehension (16 items) and judgment (11 items) stages, leading to the revision of 20 items. For example, items with vague phrases (e.g., "relationships between components") or negative wording were clarified.

Conclusion: Tourangeau's framework provides a systematic approach to conducting cognitive interviews and improving face validity in healthcare research. Its application enhanced the Nursing Talent Identification Scale; however, the small, region-specific sample may limit generalizability. Future studies should apply this framework in larger and more diverse populations to strengthen its external validity and broader applicability.

Introduction

Ensuring the accuracy and transparency of self-report instruments is a major concern in health and social science research.[1] These instruments are widely used to assess actions, attitudes, perceptions, and competencies.[2] However, their reliability and validity depend heavily on how respondents interpret the items.[3] Misunderstanding terms, variability in comprehension, or difficulties in recall can introduce error,[4] compromising data quality and threatening the validity of research findings.[1,5] Addressing these challenges is essential for generating credible and meaningful results.[3,6]

Traditional pre-testing methods often fail to identify the cognitive sources of such errors. Respondents' interpretations may vary due to language, cultural context, educational background, or personal experiences.[5,7] Even subtle wording differences can lead to ambiguity or inconsistent responses.[8] Moreover, participants may misremember events or inaccurately assess their own behavior.[8] These issues point to the need for careful pre-testing to ensure clarity, relevance, and cultural appropriateness.[7] To move beyond surface-level feedback and systematically uncover the cognitive origins of response error, researchers increasingly turn to cognitive interviewing.

Research Highlights

What is the current knowledge?

- Face validity is a key component ensuring clarity and interpretability of self-report instruments.
- Cognitive interviewing is an effective method to assess and enhance face and content validity.
- The lack of a structured framework may limit the effectiveness of cognitive interviews.
- Tourangeau's Question-and-Answer Model conceptualizes four cognitive stages in the response process.

What is new here?

- Tourangeau's framework was systematically applied to guide the cognitive interview process.
- The QAS checklist was integrated with Tourangeau's model to design targeted verbal probes.
- A structured analytical matrix was developed for organizing and interpreting interview data.
- Comprehension and judgment stages were identified as the main sources of response error.
- Applying a theoretical framework appears to have provided a systematic basis for clarifying items, which potentially strengthened face validity.
- The methodological framework provides a step-by-step guide that can be adapted for instrument development and refinement in other healthcare domains or professional fields.

Cognitive interviewing is widely recognized for its ability to detect such problems.[6,9] This method explores the cognitive processes respondents use when answering items, helping reveal ambiguities, misunderstandings, or other sources of measurement error.[10] Common techniques include think-aloud, in which respondents verbalize their thought processes, and verbal probing, in which researchers ask follow-up questions to clarify interpretations.[11,12] However, without a structured theoretical guide, the design of probes and analysis of findings can be unsystematic. To address this, cognitive interviewing often draws on theoretical models such as Tourangeau's Question-and-Answer Framework.[13]

Tourangeau's Framework describes the four stages of item response: comprehension, retrieval, judgment, and response.[14] Comprehension involves interpreting the intent of the question; retrieval requires recalling relevant information; judgment entails evaluating and integrating that information; and the response stage involves selecting or constructing an answer. This framework provides a structured lens to systematically guide probing, ensuring each cognitive stage is explored, and to analyze responses, thereby directly addressing the risk of unsystematicity in cognitive interview.[14,15] Understanding how respondents move through these stages enables researchers to pinpoint specific sources of error and improve item clarity and consistency.[15,16]

This method is particularly valuable when developing instruments designed to measure complex, multilevel, or subjective constructs.[11,16] "Nursing talent" is a prime example of such a complex construct. It is inherently multi-level, encompassing cognitive abilities (e.g., clinical reasoning), affective traits (e.g., empathy, resilience), and motivational values (e.g., commitment to care). It is also subjective, as its manifestation and perception can vary across cultural and personal contexts. This complexity creates significant potential for varied item interpretation among respondents. Identifying individuals who possess traits suited to nursing can enhance satisfaction, performance, and patient care quality.[17]

To develop a tool for this construct, an initial 95-item pool was created for the Nursing Talent Identification Scale (developed in Persian by Bagheriyeh in 2021 at Tabriz University of Medical Sciences).[17] As part of its validation, this study employed cognitive interviewing to examine this full item pool. Following this and other validation steps, the scale was subsequently refined to a 54-item instrument. Given the potential for diverse interpretations, cognitive interviewing plays a critical role in identifying ambiguous items and ensuring consistent understanding.[18]

This study aims to introduce and apply Tourangeau's Question-and-Answer Framework as a systematic guide for conducting cognitive interviews and enhancing face validity. By offering a systematic approach, this study enhances the clarity, usability, and reliability of self-report instruments, ultimately strengthening the interpretability of research findings in health and social science research.

Materials and Methods

This study employs a methodological design aimed at introducing and applying a structured framework for cognitive interviewing. The framework is grounded in Tourangeau's four-stage question-and-answer model, which examines the respondent's cognitive process through comprehension, information retrieval, judgment, and response selection.[15,19] To demonstrate the practical application of this framework, a descriptive qualitative case study was conducted. This research is part of an approved doctoral dissertation at Tabriz University of Medical Sciences, Iran. The study was carried out between March 2019 to June 2021.

The study participants consisted of 20 first-year nursing students purposively selected from the Schools of Nursing and Midwifery at three major universities in Northwest Iran: Tabriz (n = 7), Urmia (n = 7), and Ardabil (n = 6). A maximum variation sampling approach was employed based on three key criteria to ensure a representative and diverse sample: (1) Geographic and linguistic diversity (reflecting the regional dialects and cultural backgrounds of the three provinces), (2) Academic performance (including students with diverse GPAs), and (3) Gender balance. This structured selection ensured that the identified cognitive issues were representative of the broader first-year nursing student population in the region.

In this study, data were collected through cognitive interviews utilizing two primary techniques: the think-aloud method and verbal probing, after obtaining informed consent from the participants. Participants were presented with the initial 95-item pool of the Nursing Talent Identification Scale. These techniques allowed participants to verbalize their thoughts processes in real-time while enabling the researcher to adapt the interview strategy when participants encountered difficulties with thinking aloud. Verbal probing was implemented using both passive and active questioning techniques to explore participants' cognitive processes comprehensively. Passive probing was employed reactively, involving interventions only when participants exhibited signs of difficulty, such as hesitation or long pauses (e.g., 'I noticed you hesitated there; what were you thinking?'). In contrast, active probing involved proactive, pre-planned questions designed to evaluate specific cognitive stages. For instance, to assess comprehension, participants were asked: 'In your own words, what does the term «alleviation» mean to you?' or for judgment: 'How did you decide between the response options provided?' This dual approach ensured that both spontaneous difficulties and structured cognitive evaluations were captured.

A key methodological contribution of this study was the systematic integration of the Question Appraisal System (QAS) checklist with Tourangeau's four-stage cognitive model to develop the interview guide. The QAS is a standardized taxonomy designed to identify potential cognitive difficulties in survey questions by evaluating them across several categories, such as reading, instructions, clarity, and response categories. This integration was achieved through a structured mapping process, where specific QAS categories were aligned with corresponding stages of Tourangeau's model (comprehension, retrieval, judgment, and response) to generate targeted verbal probes:[20]

- *Comprehension stage:* Mapped to QAS categories concerning clarity (e.g., "Vague/Ambiguous Terms"), technical jargon, and question intent.
- *Retrieval stage:* Mapped to QAS elements concerning memory demands and recall requirements
- *Judgment stage:* Mapped to QAS categories assessing logical consistency, assumptions, and estimation problems. Probes here targeted the internal evaluation process, such as difficulty applying a personal standard or integrating multiple pieces of recalled information, distinguishing it from mere uncertainty caused by prior stages.
- *Response stage:* Mapped to QAS components examining response category adequacy (e.g., missing or overlapping options) and formatting issues.

The resulting semi-structured interview guide incorporated this integrated framework (see Table 1), Prior to implementation, the interview guide and procedural protocol underwent validation through a pilot test with two methodological experts, which resulted in refinements to probe phrasing and interview sequencing. For instance, based on the methodological experts' feedback, the phrasing of specific cognitive probes was modified to be more open-ended. For example, a directive question such as 'Was this question easy to understand?' was refined to 'In your own words, what is this question asking?' to avoid social desirability bias and better capture the participant's actual comprehension process. Additionally, the sequencing was adjusted by placing more sensitive items toward the end of the session to establish better rapport with the students first.

Cognitive interviews were conducted with participants at the participating Nursing and Midwifery Schools. Data collection employed think-aloud protocols alongside concurrent and retrospective verbal probing techniques.[18] During administration, the researcher documented behavioral observations including item skipping, response alterations, scale navigation difficulties, and hesitation patterns. Participants were instructed to verbalize their thoughts continuously while completing the 95-item scale. Semi-structured interviews further explored item clarity, comprehensibility, and face validity concerns. Each interview session lasted approximately 40-50 minutes, ensuring comprehensive coverage of all items in the pool while maintaining participant engagement.

Data Analysis

Data analysis followed a structured, multi-step process guided by the cognitive interviewing framework and utilizing a matrix analysis approach. The analysis was conducted systematically as follows:

Table 1. Cognitive Interview Probe Guide Integrating Tourangeau's Four-Stage Model and Supplementary Objectives

Probe Type / Cognitive Stage	Objective	Sample Probing Question
General Probes	To gather overall feedback on scale presentation and usability.	"What are your general thoughts on the scale's layout or length?"
Observation-Based Probes	To follow up on observed behaviors (hesitation, skipping).	"Why did you not answer this item? Why did you change your answer? I noticed you paused here. What was going through your mind?"
Comprehension Probes	To assess understanding of item intent, wording, and terminology.	"What does this question mean to you?" "Can you explain this term in your own words?"
Retrieval Probes	To explore the process of recalling relevant information or experiences.	"How did you remember or recall the information needed to answer this?"
Comfort Probes	To identify emotional or sensitivity-related barriers.	"Did any question make you feel uncomfortable? Why?"
Content Probes	To assess face validity and relevance.	"Is this item relevant for assessing nursing aptitude? Is there any missing item to identify nursing aptitude? "

Data Organization and Matrix Population

All cognitive interviews were audio-recorded, transcribed verbatim, and imported into MAXQDA software (Version 2020) for initial coding. Following transcription, a comprehensive analysis matrix was created based on the framework proposed by d'Ardenne and Collins[18,21] (Table 2). A separate matrix row was created for each of the 95 items in the initial pool, with eight columns: respondent details, survey answers, observational findings, think-aloud data, general probes, comprehension probes, retrieval probes, judgment/response probes, and other relevant findings.

Problem Identification and Categorization Process

Two researchers independently populated the matrix by extracting relevant data from transcripts and observational notes. The initial problem identification involved flagging any instances where participants demonstrated:

- Verbalized confusion or requests for clarification
- Inconsistent responses between think-aloud and final answers
- Observable difficulties (hesitation, answer changes, prolonged response time)
- Direct feedback during verbal probing

To ensure coding consistency, both researchers initially analyzed the same three transcripts, achieving an inter-coder reliability of 85% through Cohen's kappa calculation. Remaining discrepancies were resolved through consensus discussions with a third methodological expert.

Theoretical Framework Integration

The critical analytical phase involved mapping identified problems to Tourangeau's cognitive stages through a systematic classification process:

- *Comprehension issues* were classified when participants misunderstood item meaning, questioned terminology, or interpreted questions differently than intended.
- *Retrieval problems* were identified when participants struggled to recall relevant experiences or information.
- *Judgment difficulties* were categorized only when participants, after demonstrating clear comprehension and successful retrieval, expressed difficulty in evaluating, comparing, or integrating information to form a stable answer. This included overt statements of indecision between options due to internal conflict or lack of a clear personal benchmark.
- *Response issues* were noted when participants understood the item but faced challenges with response options or formatting.

This classification was achieved by examining the matrix data against predefined criteria for each cognitive stage, ensuring that every identified problem was explicitly linked to its corresponding stage in Tourangeau's model. The final analysis yielded a comprehensive overview of cognitive challenges across all scale items, directly informing the subsequent item revision process.

Results

The mean age of participants was 20 years (Standard Deviation [SD] = 2.5) and the majority (11 out of 20, 55%) were female. Based on the cognitive interviewing process, a total of 20 out of the 95 scale items were revised. The identified problems and subsequent modifications are reported below, structured according to Tourangeau's four-stage question-and-answer model. An overview of the findings is presented in Table 3. As shown, problems were not evenly distributed across cognitive stages; most issues were traced to the comprehension (16 items) and judgment (11 items) stages, while retrieval and response problems were less frequent.

Comprehension

Modifications in this phase were primarily aimed at improving participants' understanding of the items by simplifying wording and incorporating additional explanatory details. Problems were categorized into three main areas:

Intent and Clarity

Participants often misunderstood the rationale behind certain items or were unclear about what was being asked, leading to a mismatch between their interpretation and the researcher's intent.

For instance, in Item 31 (*"I am in contact with many people"*), the respondents failed to interpret this item as intended; they associated it with common interaction and not social skills. A respondent queried, "Why should one interact with everyone?" (P.12), while another noted, "*I don't see how having many contacts relates to my professional skills*" (P.4). Consequently, the item was revised to: "*I have good social relationships with others*".

In Item 85 (*"I evaluate the situations with a broad view and through considering the relationships between the components"*), the phrase "*relationships between the components*" was felt to be imprecise and ambiguous the participants reported: "*I did not understand what you mean by 'the relationships between the components'; it is vague*" (P.8), and another mentioned, "*This sounds like a physics question, not about my behavior*" (P.15). The item was modified to: *"I try to consider all aspects of a problem"*.

In Item 94 ("I respond quickly to visual and auditory stimuli such as sound and light"), Participants' interpretation of this item was other than intended;

Table 2. Example of an Empty Matrix for Item 7

Item 7: I promptly notify an error when it needs to be corrected.							
Respondent details (ID, age, and gender)	Survey answer	Thinking aloud/ observations	General probes	Comprehension probes	Retrieval probes	Judgment/ response probes	Other

Table 3. Item Issues and Revisions Informed by Participant Feedback

Item no.	Original Item Wording	Interview findings	Revised Item Wording
2	I try to perform the assigned tasks correctly and completely.	What does 'assigned' mean? I did not understand the meaning of this word.	I try to do my tasks correctly and completely.
7	I promptly notify an error when it needs to be corrected.	It is vague; do you mean a mistake by me or others?	If I make a mistake that needs to be corrected, I report it immediately.
11	I easily adapt to an environment with strict rules and regulations.	Putting the words 'easy' and 'strict' together makes it difficult to understand the meaning. The word 'strict' gives a negative impression and the image of a military camp comes to my mind. What is meant by 'strict regulations'?	I can comply with workplace rules and regulations.
12	In my view, it is important to observe the rules and regulations.	Why don't you measure pragmatism? You are measuring the view, but action is more important.	I respect the rules and regulations.
18	I enjoy caring for others.	It reminded me of taking care of something. Rewrite it as: "I enjoy taking care of others". Delete the word 'for'.	I enjoy taking care of others.
19	While dealing with people in pain, I try to alleviate their suffering.	I did not understand the meaning of alleviation. Can you express it with a clearer meaning?	While dealing with people in pain, I try to reduce their suffering.
20	I feel obligated to care for and support people in need and the helpless.	The word 'obligated' implied compulsion and it did not make me feel good. Maybe it would be better to use the word 'willing' instead of it. The item is long and difficult to understand; I read it three times to find out the meaning.	I consider it my moral responsibility to support the needy and disabled.
23	I listen to others with interest.	Which conversations do you mean? Who do you mean by others? Do you mean friendly chat?	I listen carefully to someone speaking.
31	I am in contact with many people.	It is associated with any kind of communication. It gives different meanings. Interact with everyone? Why interact with everyone? Have a lot of friends?	I have good social relationships with others.
33	I can change people's beliefs and attitudes by providing logical explanations.	It implied religious beliefs to me. Why do I have to change people's religious beliefs?	I can change people's views and attitudes by providing logical explanations.
34	I am an eloquent and audible speaker.	Two meanings came to my mind: 'to speak bluntly' and 'to speak frankly'.	I speak to others in a simple and understandable way.
41	I listen to the views of the opposition.	It is vague and distracting. If you mean teamwork, then add the word 'in teamwork' or 'collective decision'.	I listen to the views of the opposition in teamwork.
47	Without judging, I see things from the perspective of others.	This sentence was difficult to understand and participants had a long pause to respond to the item.	I can understand the problems of others without judging.
55	When others are aggressive, I calmly listen to them.	The sentence is unusual. You cannot be calm in aggressive situations, but the person tries to behave calmly.	When others are aggressive, I try to listen to them calmly.
72	I can do a lot of things without feeling tired.	It is somehow unusual or in other words exaggerated. It is an abnormal expectation. Certainly, a lot of work tires a person.	I get tired quickly with the least amount of work.
83	I do not insist on my wrong ideas and opinions.	The sentence is negative and confusing. I had a problem answering this question.	I am open to changing my wrong ideas and opinions.
85	I evaluate the situations with a broad view and through considering the relationships between the components.	It seemed a little vague; I did not understand the relationship between the components.	I try to consider all aspects of a problem.
87	Compared to other people, I pay attention to important details when doing things.	'Compared to other people' can be omitted from the beginning of the sentence. A person may be much more precise (being very meticulous is not really needed in nursing profession) compared to the respondent.	I pay attention to important details when doing things.
93	I am slow in performing my tasks accurately.	My impression of this item is that in order to be careful in doing things, the person should act slowly. The words 'accurate' and 'slow' imply two opposite meanings and make the sentence difficult to understand.	While I am fast, I am careful enough when doing things.
94	I respond quickly to visual and auditory stimuli such as sound and light.	It sounds to me as being bothered by sound and light.	I use my senses to be aware of my surroundings

they interpreted it as being irritated by sound and light. Their perception was that *"This item means that they are bothered by sound and light"* (P.5), and another participant added, "*It feels like you are asking if I have a headache or sensitivity to light*" (P.2). The final version became: *"I use my senses to be aware of my surroundings"*.

Item Word Choice

Some items contained difficult or ambiguous words that hindered comprehension.

For example, Item 34 ("*I am an eloquent and audible speaker*"), the words "eloquent" and "audible" were found to be difficult to understand (P.3). Meanwhile, some participants expressed that: "*This item implies speaking frankly*" (P.11). The item was simplified to: *"I speak to others in a simple and understandable way"*.

In Item 93 (*"I am slow in performing my tasks accurately")*, the participants' perception was that *"To do things accurately, you must be slow"* (P.18), while this was completely different from the intended meaning of

the research team. The participants also stated that "*The words 'accurate' and 'slow' imply opposite meanings and make the sentence difficult to understand*" (P.17). This was rephrased as: "*While I am fast, I am careful enough when doing things*".

Item Relevance and Context

Some items were perceived as less relevant or unrealistic within the participants' context.

Item 72 ("*I can do a lot of things without feeling tired*"), the participants stated that this item looks a bit unusual. They stated that *"Surely a lot of work tires a person. It is better to say that I get tired quickly with the least amount of work"* (P.7). This item was changed accordingly to improve its contextual appropriateness.

Information Retrieval

This stage evaluates the strategies participants employed to recall information needed to answer each item. The interview protocol included targeted probes (e.g., "How did you remember or recall the information needed to answer this?") to explicitly explore this cognitive stage. Analysis of participant responses indicated that no significant difficulties with information retrieval were reported for the scale items. Participants typically recalled recent academic or clinical experiences to provide their answers. For instance, one student noted, "I just thought about my behavior during the last clinical placement in the hospital; it was easy to remember how I acted there" (P.14). Another participant explained, "I recalled my general habits during my daily studies and interactions with classmates over the past semester" (P.6). These responses confirm that the timeframe and context of the items were appropriate for the target population". All participants were able to effectively recall relevant past and current experiences to respond to the items, suggesting that their retrieval strategies were functioning as expected. Specifically, participants did not report instances of being unable to remember relevant events or experiences, nor did they express that the items required inaccessible or overly burdensome memory tasks. Through several key indicators during the cognitive interviews. First, response latency was monitored; participants who responded without prolonged hesitation or signs of cognitive strain were deemed to have efficient retrieval. Second, the specificity of recalled events served as a critical criterion. When probed with "How did you remember this?", participants who provided concrete, recent examples (e.g., specific clinical shifts or particular timeframes) demonstrated successful retrieval. Furthermore, the absence of "Don't know" responses or requests for clarification regarding the timeframe confirmed that the items did not impose inaccessible or overly burdensome memory tasks.

Judgment

This component examines the level of confidence or uncertainty participants experienced when selecting their answers. Ambiguity in wording led to uncertainty in responses for specific items.

For Item 7 (*"I promptly notify an error when it needs to be corrected")* in the intended manner and there were different interpretations. Participants reported that *"It is not clear whether the error was made by the individual or other people"* (P.9). To resolve this ambiguity, the item was clarified to: *"If I make a mistake that needs to be corrected, I report it immediately"*.

In Item 87)*"I pay attention to important details in doing things compared to others"*(, the participants stated that *"A person may be much more precise (being very meticulous is not really needed in nursing profession) compared to the respondent"* (P.16). The comparative phrase was removed during revision.

Response

This component examines how participants selected and used the response options. The feedback primarily concerned the response scale and the phrasing of items.

Participants were asked to comment on the response format. While they generally agreed that the five-point Likert scale (from strongly agree to strongly disagree) was reasonable (e.g. *"they are reasonable"* (P.13) or *"did not notice a problem"* (P.8). However, some participants suggested that the scale be reduced to only two or three answer options: *"it is sufficient to ask to agree or disagree"*. *"There is no need for strongly agree or strongly disagree"* (P.11). Many participants preferred that the five-point Likert scale was easier and more accurate than the two or three-point Likert scale.

A key observation was that participants had difficulty responding to negatively worded items. This was particularly evident for Item 83: (*"I do not insist on my wrong ideas and opinions").* To improve clarity and response consistency, this item was rewritten as an affirmative statement (*"I am open to changing my wrong ideas and opinions"*).

Discussion

Cognitive interviewing was a highly effective method for optimizing scale items, significantly enhancing their clarity, comprehensibility, and overall quality.[22] While psychometric evaluation focuses on item performance using quantitative methods, cognitive interviewing provides a detailed examination of the respondents' cognitive processes, ensuring that survey items function as intended from the perspective of the participants.[10,12] Techniques such as think-aloud and verbal probing allowed us to examine in real-time how respondents processed each item. By applying Tourangeau's four-stage model, this examination moved beyond a generic exploration. Instead, it systematically diagnosed error sources across comprehension, retrieval, judgment, and response.[23]

A key factor contributing to comprehension difficulties in this study is the gap between the language and cognitive framework of the tool developers (researchers

and professors) and the participants (students). Often, researchers use academic or technical language that may be familiar in scholarly contexts but is less accessible to participants, especially those who are still developing their professional or academic literacy. This was evident when items used phrases like "relationships between the components" or "I am in contact with many people," which participants interpreted in ways that diverged from the researchers' intentions. This highlights the importance of considering how terms are perceived by the target population. The cognitive model of Tourangeau (1984) emphasizes that comprehension involves both understanding the words used and interpreting them within the context of the respondent's own knowledge and experience. If these words or concepts are not clear, the entire process of comprehension is compromised.[15,18] The predominance of problems in the comprehension stage underscores its vulnerability as the foundational gatekeeper of the response process; errors introduced here propagate through all subsequent stages, distorting the final response.

Our findings underscore that the process of responding to a survey items are complex and multidimensional. Tourangeau's framework was instrumental in deconstructing this complexity. By separating the cognitive process into distinct stages, we could systematically pinpoint whether a problem originated in understanding the question (comprehension), accessing a memory (retrieval), evaluating information (judgment), or mapping an answer (response).[15,24] Cognitive interviewing informed by this model helped identify the root causes of potential item nonresponse, revealing not only how respondents interpreted the items but also flaws in the tool itself. A key insight was that even minor rewording or structural alterations were found to markedly enhance an item's clarity [16], demonstrating that initial poor comprehension does not necessarily warrant item removal.

This approach aligns with the COSMIN (Consensus-based Standards for the selection of health Measurement Instruments) guidelines, which advocate for refining items by incorporating the language and perceptions of the target population.[7,25] Ultimately, scales written in clear and unambiguous language facilitate more successful and accurate responses from participants.[6,26] Furthermore, the structured application of Tourangeau's model offered a significant advantage over an unstructured cognitive interview. It ensured comprehensive coverage of all potential cognitive pathways for error, guided the development of targeted probes for each stage, and provided a consistent taxonomy for analyzing and categorizing problems. This systematization enhances the reliability and replicability of the cognitive testing process.

This vulnerability in the comprehension stage emphasizes the importance of contextualizing language in research instruments. The fact that small changes in wording can significantly improve understanding reinforces the value of co-developing instruments with the target population in mind. Researchers must be mindful that even minor ambiguities in wording can lead to incorrect responses, which can skew the validity of the survey results. Therefore, attention to clarity and simplicity in language especially in terms of cultural and linguistic appropriateness is crucial for ensuring the accuracy and reliability of survey-based research.[22,27]

Furthermore, participants consistently struggled with negatively worded items, which created confusion and complicated response selection. This finding is consistent with prior research recommending the avoidance of negative phrasing to prevent cognitive strain and maintain response consistency.[10,28]

In summary, participant feedback was invaluable in refining the scale. Although cognitive interviewing is resource-intensive, its application during development ensures that instruments like the Nursing Talent Identification Scale are valid, reliable, and well-aligned with the nursing profession. This study demonstrates that employing a structured theoretical framework, such as Tourangeau's model, can significantly enhance the rigor and diagnostic power of cognitive interviews. Researchers must recognize the importance of language clarity,cultural relevance and systematic cognitive testing in designing survey tools to ensure the accuracy and effectiveness of their measurements.

Limitations

This study has several limitations. First, while the sample size of 20 participants and purposive sampling were appropriate for an in-depth qualitative exploration and provided sufficient data to identify major item flaws, it does impose limits. The participants were selected from only three western provinces of Iran. Consequently, the findings may not fully represent the diverse perspectives of all potential respondents, which limits the generalizability of the results. Although data saturation was reached for prominent item-level issues, it is possible that a different or larger sample might reveal additional, nuanced problems**.** Second, the participants' relatively homogeneous cultural and linguistic backgrounds may not capture the potential comprehension challenges that could arise in more culturally or linguistically diverse groups. Finally, an inherent limitation of cognitive interviewing is the potential for social desirability bias, where participants may provide answers they believe the researcher expects. While techniques like think-aloud can mitigate this, it remains a consideration when interpreting feedback on sensitive or value-laden items.

Conclusion

This study suggests that applying Tourangeau's Question-and-Answer Framework provides a systematic and effective approach to conducting cognitive interviews, potentially enhancing the face validity of self-report instruments. By integrating this framework with the QAS checklist and employing structured think-aloud and verbal probing techniques, researchers were able to

identify and revise 20 out of 95 problematic items in the Nursing Talent Identification Scale. Most issues originated in the comprehension and judgment stages, highlighting the importance of clear wording, contextual relevance, and item interpretability. The findings underscore that cognitive interviewing guided by a theoretical framework not only improves item clarity and response accuracy but also contributes to the overall validity and reliability of the instrument. This methodological approach can be adapted for the development and refinement of self-report tools in nursing and other healthcare domains, ensuring that instruments are interpretable, culturally appropriate, and aligned with the cognitive processes of respondents. Ultimately, the study reinforces the value of structured cognitive interviewing as a critical step in high-quality instrument development. Effective cognitive interviewing, underpinned by a solid theoretical framework, is essential not just for identifying and fixing issues in survey instruments, but for ensuring they are both scientifically robust and culturally relevant. Researchers should prioritize this method to create tools that truly capture the nuances of respondents' experiences and perspectives.

Acknowledgments

The authors sincerely thank all participants for generously sharing their time and valuable experiences.

Authors' Contribution

Conceptualization: Farzaneh Bagheriyeh, Akram Ghahramanian, Vahid Zamanzadeh.
Data curation: Farzaneh Bagheriyeh, Akram Ghahramanian.
Formal analysis: Farzaneh Bagheriyeh, Akram Ghahramanian, Vahid Zamanzadeh, Leila Valizadeh, Behrouz Amini Kordkandi.
Funding acquisition: Farzaneh Bagheriyeh.
Investigation: Farzaneh Bagheriyeh, Akram Ghahramanian, Vahid Zamanzadeh, Leila Valizadeh, Behrouz Amini Kordkandi.
Methodology: Farzaneh Bagheriyeh, Akram Ghahramanian, Vahid Zamanzadeh, Leila Valizadeh.
Project administration: Farzaneh Bagheriyeh, Akram Ghahramanian, Vahid Zamanzadeh.
Resources: Farzaneh Bagheriyeh, Akram Ghahramanian, Vahid Zamanzadeh, Leila Valizadeh.
Supervision: Farzaneh Bagheriyeh, Akram Ghahramanian, Vahid Zamanzadeh, Leila Valizadeh.
Validation: Farzaneh Bagheriyeh, Akram Ghahramanian.
Visualization: Farzaneh Bagheriyeh, Akram Ghahramanian.
Writing–original draft: Farzaneh Bagheriyeh, Akram Ghahramanian, Vahid Zamanzadeh, Leila Valizadeh, Behrouz Amini Kordkandi.
Writing–review & editing: Farzaneh Bagheriyeh, Akram Ghahramanian, Vahid Zamanzadeh, Leila Valizadeh.

Competing Interests

The authors declare they have no competing interests.

Data Availability Statement

The datasets used and/or analyses the current study are available from the corresponding author upon reasonable request.

Ethical Approval

This study was approved by the Medical Ethics Committee of Tabriz University of Medical Sciences (Approval Code: IR.TBMED.REC.1397.583) and was performed in accordance with the ethical standards outlined in the Declaration of Helsinki. Written informed consent was obtained from all participants after they were fully informed about the study's purpose and procedures. Participants were assured of the anonymity and confidentiality of their data. They were informed that the audio recordings would be used solely for transcription purposes and would be permanently deleted thereafter, and that their names would not be disclosed in any publications. Additionally, participants were made aware of their right to withdraw from the study at any stage without any penalty.

Funding

This study was financially supporting by Tabriz University of Medical Sciences. The funding part had no role in the design of study, the collection, analysis and interpretation of data, or in writing the manuscript.

References

1. Alcaraz-Ibáñez M, Paterna A, Sicilia Á, Griffiths MD. Examining the reliability of the scores of self-report instruments assessing problematic exercise: A systematic review and meta-analysis. J Behav Addict 2022;11(2):326–47. doi:10.1556/2006.2022.00014
2. Landsverk NG, Olsen NR, Brovold T. Instruments measuring evidence-based practice behavior, attitudes, and self-efficacy among healthcare professionals: a systematic review of measurement properties. Implement Sci 2023;18(1):42. doi:10.1186/s13012-023-01301-3
3. Ahmed I, Ishtiaq S. Reliability and validity: Importance in Medical Research. J Pak Med Assoc 2021;71(10):2401–6. doi:10.47391/jpma.06-861
4. Shaheen N, Shaheen A, Ramadan A, Hefnawy MT, Ramadan A, Ibrahim IA, et al. Appraising systematic reviews: a comprehensive guide to ensuring validity and reliability. Front Res Metr Anal 2023;8:1268045. doi:10.3389/frma.2023.1268045
5. Balza JS, Cusatis R, McDonnell SM, Basir MA, Flynn KE. Effective questionnaire design: How to use cognitive interviews to refine questionnaire items. J Neonatal Perinatal Med 2022;15(2):345–9. doi:10.3233/npm-210848
6. Meadows K. Cognitive Interviewing Methodologies. Clin Nurs Res 2021;30(4):375–9. doi:10.1177/10547738211014099
7. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res 2021;30(8):2197–218. doi:10.1007/s11136-021-02822-4
8. Ebadi A, Zarshenas L, Rakhshan M, Zareiyan A, Sharifnia S, Mojahedi M. Principles of scale development in health science. Tehran: Jame-e-negar. 2017;6(1):402.
9. Scott K, Ummer O, LeFevre AE. The devil is in the detail: reflections on the value and application of cognitive interviewing to strengthen quantitative surveys in global health. Health Policy Plan 2021;36(6):982–95. doi:10.1093/heapol/czab048
10. Patel-Syed Z, Becker S, Olson M, Rinella H, Scott K. What do you think it means? Using cognitive interviewing to improve measurement in implementation science: description and case example. Implement Sci Commun 2024;5(1):14. doi:10.1186/s43058-024-00549-0
11. Spoto A, Nucci M, Prunetti E, Vicovaro M. Improving content validity evaluation of assessment instruments through formal content validity analysis. Psychol Methods 2025;30(2):203–22. doi:10.1037/met0000545
12. Hall S, Beatty S. Assessing spiritual well-being in residents of nursing homes for older people using the FACIT-Sp-12: a cognitive interviewing study. Qual Life Res 2014;23(6):1701–11. doi:10.1007/s11136-014-0627-6
13. Wolcott MD, Lobczowski NG. Using cognitive interviews and think-aloud protocols to understand thought processes. Curr Pharm Teach Learn 2021;13(2):181–8. doi:10.1016/j.cptl.2020.09.005

14. Castillo-Díaz M, Padilla García JL. How Cognitive Interviewing can Provide Validity Evidence of the Response Processes to Scale Items. Social Indicators Research 2013;114(3):963-75. doi:10.1007/s11205-012-0184-8
15. Tourangeau R. Cognitive sciences and survey methods. Cognitive aspects of survey methodology: Building a bridge between disciplines. Washington, D.C.: National Academy Press;1984; 73-100.
16. DiBenedetti DB, Price MA, Andrews EB. Cognitive interviewing in risk minimization survey development: patient and healthcare professional surveys. Expert Rev Clin Pharmacol 2013;6(4):369–73. doi:10.1586/17512433.2013.811824
17. Bagheriyeh F. *Development and Psychometric Evaluation of "Nursing Talent Identification" Instrument*. Dissertation. Tabriz University of Medical Sciences; 2021.
18. D'Ardenne J. Developing interview protocols. Cognitive interviewing practice. London: SAGE; 2015:101-25. doi:10.4135/9781473910102
19. Tourangeau R, Rips LJ, Rasinski K. The psychology of survey response. Cambridge University Press;2000. doi:10.1017/CBO9780511819322
20. Willis GB. Questionnaire design, development, evaluation, and testing: Where are we, and where are we headed? Advances in questionnaire design, development, evaluation and testing. Hoboken, NJ: John Wiley & Sons, Inc. 2020:1-23. doi:10.1002/9781119263685.ch1
21. Collins D. Cognitive interviewing: origin, purpose and limitations. Cognitive interviewing practice. London: SAGE; 2015: 3-27. doi:10.4135/9781473910102
22. Wenemark M, Ngwili N, Ndoboli D, Wieland B, Roesel K. "How Are My Age and Cows Related?" Cognitive Interviewing as a Tool to Pretest Survey Questions in Two Limited Resource Settings. Front Vet Sci 2022;9:833748. doi:10.3389/fvets.2022.833748
23. Scott K, Gharai D, Sharma M, Choudhury N, Mishra B, Chamberlain S, et al. Yes, no, maybe so: the importance of cognitive interviewing to enhance structured surveys on respectful maternity care in northern India. Health Policy Plan 2020;35(1):67–77. doi:10.1093/heapol/czz141
24. Arthur EK, Menon U, Browning K, Overcash J, Wills CE. Challenges of Cognitive Interviewing in Sensitive Health Topic Research. Nurs Res 2021;70(5):376–82. doi:10.1097/nnr.0000000000000530
25. Jang MK, Kim S, Collins EG, Quinn LT, Park CG, Ferrans CE. Enriching the quality of cross-cultural instrument development through cognitive interviewing: Implications for nursing research. Jpn J Nurs Sci 2020;17(2):e12301. doi:10.1111/jjns.12301
26. Goodfellow LT. An Overview of Survey Research. Respir Care 2023;68(9):1309–13. doi:10.4187/respcare.11041
27. Poll GH, Petru J. Exploring the contributions of expert review and cognitive interviewing to evaluating the content validity of items for anew measure of adolescent social communication, the Transition Pragmatics Interview. Clin Linguist Phon 2023;37(12):1124–40. doi:10.1080/02699206.2022.2148131
28. Meadows KA, Ferrante N, Geršić T. Cognitive Interviews: Recommendations for Best Practices in Clinical Outcome Assessment (COA) Measure Development and Validation. Patient 2025;18(6):609–21. doi:10.1007/s40271-025-00752-8