

# Content validity of the credibility and expectancy questionnaire in a pain rehabilitation setting

Citation for published version (APA):

Mertens, V.-C., Moser, A., Verbunt, J., Smeets, R., & Goossens, M. (2017). Content validity of the credibility and expectancy questionnaire in a pain rehabilitation setting. *Pain Practice*, 17(7), 902–913. <https://doi.org/10.1111/papr.12543>

## Document status and date:

Published: 01/09/2017

## DOI:

[10.1111/papr.12543](https://doi.org/10.1111/papr.12543)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

---

## ORIGINAL ARTICLE

---

# Content Validity of the Credibility and Expectancy Questionnaire in a Pain Rehabilitation Setting

---

Vera-Christina Mertens, PhD<sup>\*,†</sup>; Albine Moser, PhD, MPH, RN<sup>‡,§</sup>; Jeanine Verbunt, MD, PhD<sup>†,¶,\*,\*\*</sup>; Rob Smeets, MD, PhD<sup>†,¶,††</sup>; Mariëlle Goossens, PhD<sup>†,‡‡</sup>

*\*Research Unit INSIDE, Institute for Health and Behaviour, University of Luxembourg, Esch-sur-Alzette, Luxembourg; †Department of Rehabilitation Medicine, School for Public Health and Primary Care, Maastricht University, Maastricht; ‡Department of Family Practice, School for Public Health and Primary Care, Maastricht University, Maastricht; §Faculty of Healthcare, Research Programme Autonomy and Participation of Chronically Ill People, Zuyd University of Applied Sciences, Heerlen; ¶Adelante Centre of Expertise in Rehabilitation and Audiology, Hoensbroek; \*\*Department of Rehabilitation Medicine, Maastricht University Medical Center, Maastricht; ††Libra Rehabilitation and Audiology, Eindhoven/Weert; ‡‡Department of Clinical Psychological Sciences, Faculty of Psychology and Neurosciences, Maastricht University, Maastricht, The Netherlands*

### ■ Abstract

**Background:** Content validity, the proper reflection of the concept to be measured, is yet unknown for the Credibility and Expectancy Questionnaire (CEQ). It is frequently used in pain rehabilitation because treatment expectancy is influential on the outcome.

**Objective:** To test and improve the content validity of the CEQ in a sample of patients with chronic pain in different phases of their treatment.

**Methods:** A qualitative observational study design using the Three-Step Test-Interview method was used. Therein, data collection, analyses, and adaptations occur iteratively.

Address correspondence and reprint requests to: Vera-Christina Mertens, PhD, Research Unit INSIDE, Institute for Health and Behaviour, University of Luxembourg, Campus Belval, Maison des Sciences Humaines, 11, Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg. E-mail: veracmertens@gmail.com.

Submitted: April 1, 2016; revised September 22, 2016;

Revision accepted: October 16, 2016

DOI: 10.1111/papr.12543

**Results:** Seventeen patients with chronic pain in different stages of treatment participated through convenience sampling from the mother sample of a randomized controlled trial. The main study parameter is content validity, which is defined as (1) interpretations and responses of the participants and (2) the identification of response problems operationalized, and resulting in changes in the CEQ. For patients waiting for treatment, the written instruction of the CEQ allowed different interpretations. After changing the instructions, the CEQ became an easy-to-understand and content-valid questionnaire. For patients who had already undergone treatment, changes regarding time frame and recall period were necessary to overcome interpretation and response problems to the CEQ. **Discussion:** After small changes, the CEQ appeared to be a content-valid measurement instrument for patients waiting for treatment. However, for patients who had already undergone treatment, the content validity of the CEQ was less, and considerable changes were necessary. ■

**Key Words:** validation studies (MeSH), self-administered questionnaire, cognitive interviewing, think aloud, treatment expectancy, pain rehabilitation treatment, questionnaire development, content validity

## INTRODUCTION

Expectancy, the affective base of “improvements that clients believe will be achieved on the basis of a particular treatment” (p. 829<sup>1</sup>), is an important concept. Its influential role in several medical conditions is known (eg, on surgery outcomes, drug efficiency, and psychiatric treatment<sup>2-4</sup>).

There are various instruments or methods for measuring the concept of expectancy in musculoskeletal disorders,<sup>5,6</sup> but the inconsistent use of valid measurement instruments is criticized.<sup>5</sup> Information regarding development and testing, as well as the reliability and validity of those instruments, is frequently omitted.<sup>6</sup> Due to this, there is an urgent need for a standardized and valid method or tool<sup>5,7</sup> that can be used to measure the level of expectancy, preferably even at multiple times.

Therefore, a measurement instrument is needed from which it can be taken for granted that it entails the concept one wants to measure. Content validity, one property of the domain validity,<sup>8</sup> refers to “the degree to which the content of an. [...] instrument is an adequate reflection of the construct to be measured” (p. 743<sup>8</sup>).

By doing so, the process and eventually the development of expectancy over time could be explored in more detail. As this is currently unknown, it is of critical importance to assess this by a content valid questionnaire. By doing so, the actual meaning of the outcomes derived from a questionnaire,<sup>9</sup> and to examine inconsistent interpretations, incomplete concept coverage, and misunderstandings<sup>10</sup> in relevant target groups at different stages of their treatment, can be interpreted properly. Several best practices exist to establish content validity in new or existing questionnaires, and an iterative approach is recommended.<sup>11-14</sup> Therein, the interpretation of and response to questionnaire items are explored. By doing so, response problems are identified. As this is an iterative approach, changes in the questionnaire can be made. And in the next step, the cycle of interpretation and response to adapted questionnaire items can be entered again.

A frequently used existing questionnaire is the adapted and Dutch-translated version<sup>15</sup> of the Credibility and Expectancy Questionnaire (CEQ) of Devilly and Borkovec.<sup>16</sup> The CEQ has good psychometric properties (internal consistency and test-retest reliability), and the 2-factor structure of construct credibility and expectancy has been confirmed.<sup>17</sup> It is often used in the pain rehabilitation setting.<sup>15,17</sup> However, its content validity is currently unknown.

The aim of this study was to test and improve the content validity of the CEQ by obtaining perspectives from a diverse sample of patients with chronic pain who are at different stages of treatment.

## Research Questions

In this study, the following research questions will be answered:

- 1a. How do patients with chronic pain before and during rehabilitation treatment interpret and respond to the items of the CEQ?
- 1b. If response problems are encountered during the completion of the CEQ, which adaptations in the items and/or instructions should be made to improve the content validity of the CEQ?

## METHODS

### Design

A qualitative observational study design<sup>18</sup> was used. A qualitative approach is indicated because we wanted to improve the validity of the CEQ.<sup>19</sup> The aim was to uncover how people with chronic pain interpret the meaning of questions and possible responses, what they think about the questions, how they decide on their answers, and what their answers mean.

For this, a specific form of cognitive interviewing (described in detail in the Interview Procedure and Observations section), the Three-Step Test-Interview (TSTI) method, was used. An advantage therein is that 2 subtypes of cognitive interview methods, think-aloud interviewing and probing techniques, were combined.<sup>19</sup> Think-aloud interviewing means that as a questionnaire is filled in, the cognitive processes are simultaneously verbalized by the participant. Probing strives to get more insight into the understanding and the retrieval or judgment process by having an interviewer ask for more specific information.<sup>19</sup>

### Participant Selection/Sampling

Seventeen participants with nonspecific chronic musculoskeletal pain or fibromyalgia who participated in a randomized controlled trial, the Pre-Pain Rehabilitation Study (PREPARE; total  $N = 163$ ),<sup>20</sup> and who agreed to be contacted for related research, were included. The inclusion and exclusion criteria for the current study were identical to those of the clinical trial.<sup>20</sup> First, we

started with a convenient sample frame,<sup>21</sup> meaning that all persons who were available were approached, without balancing specific entry requirements for this study. This choice was made to strive toward a broad dispersion of the participants' demographic characteristics, functioning levels, and treatment phases.

During the first interview round, we saw that participants who had finished rehabilitation treatment encountered specific problems with regard to the time frame of the questionnaires and recall. Therefore, we changed our sampling strategy to purposive sampling. We specifically approached participants who had finished their treatment. We asked them to imagine their situation before the start of treatment so that we could attempt to resolve the response problems in the questionnaire for this group.

This resulted in the final study population consisting of 12 participants who had already finished their pain rehabilitation treatment, 2 of whom were on the waiting list, 1 of whom was undergoing treatment, 1 of whom discontinued treatment prematurely, and 1 of whom had never started treatment (Table 1).

### Interview Procedure and Observations

Think-aloud interviewing using the TSTI technique combines 3 steps.<sup>22</sup> The first step is observing response behavior<sup>23</sup> (eg, how a participant chooses a certain response option) and concurrent verbalization (“think aloud”) while the participant answers self-report questionnaires. The second step involves follow-up probing into specific aspects of this response process (eg, the verbalization of formulations in the questionnaires that are difficult to understand). In the third step, the experiences and opinions of the participant are elicited.<sup>22</sup>

### Procedure

As thinking aloud is not a common activity in daily life, a grounded explanation and preliminary training were

necessary to elicit a sufficient amount of think-aloud behavior.<sup>19</sup> Therefore, the participants were asked to first practice this by thinking aloud and filling in a questionnaire unrelated to this study. Then, the actual think-aloud interview with the CEQ followed. The participants were given the following instructions: “Try to speak aloud everything that comes to your mind while you complete the questionnaire.”

Directly after the completion of the CEQ (Appendix S1) and while the patient was asked to think aloud, probing questions were asked to reveal more in-depth information regarding his or her thought processes. Examples of the probing questions can be found in Table 2, which also served as an interview guide. It contained questions related to the interpretations and responses of the participants regarding the CEQ and the identification of response problems in concordance with the TSTI method.

The interviews were conducted at the participants' homes to come as close as possible to the real-life situation in which the patient-related measurement instrument is completed.<sup>22,24</sup> One participant preferred to be interviewed at the university building. A pilot interview with one participant was conducted to test the interview procedure, interview guide, and interview probes. As this pilot did not lead to changes in the interview procedure, the data from this pilot interview were also used in the analysis. Figure 1 describes the iterative procedure.

Each participant was interviewed once, and all interviews were audio taped. Field notes<sup>25</sup> were made during and after every interview to document the interviewer's impressions of the interview and the interview process.

As an imperative,<sup>11</sup> data collection and analysis occurred in an iterative way (see Figure 1), and themes that emerged during the sequential analysis were added in the interview guidelines and were verified in the subsequent interviews by asking probing questions specific to these themes. For instance, we added

**Table 1. Demographic Characteristics of Participants Included for Cognitive Interviews**

| Interview number                                   | 1         | 2   | 3   | 4         | 5  | 6  | 7         | 8   | 9   | 10        | 11  | 12 | 13        | 14 | 15  | 16 | 17 |
|--|-----------|-----|-----|-----------|----|----|-----------|-----|-----|-----------|-----|----|-----------|----|-----|----|----|
| Participant identification                         | A         | B   | C   | D         | E  | F  | G         | H   | I   | J         | K   | L  | M         | N  | O   | P  | Q  |
| Sex  | F         | M   | F   | F         | F  | F  | F         | F   | M   | F         | M   | M  | M         | F  | M   | F  | F  |
| Age (years)  | 37        | 49  | 48  | 51        | 37 | 44 | 25        | 62  | 39  | 53        | 50  | 54 | 71        | 42 | 57  | 45 | 46 |
| Type of pain problem*                              | CPS       | CPS | CPS | CPS       | FM | FM | CPS       | CPS | CPS | FM        | CPS | FM | CPS       | FM | CPS | FM | FM |
| Position in rehabilitation trajectory <sup>†</sup> | a         | a   | a   | b         | c  | a  | d         | a   | a   | a         | a   | e  | a         | e  | a   | a  | a  |
| Round of analysis                                  | 1st round |     |     | 2nd round |    |    | 3rd round |     |     | 4th round |     |    | 5th round |    |     |    |    |

\*Fibromyalgia (FM) vs. centralized pain syndrome (CPS).

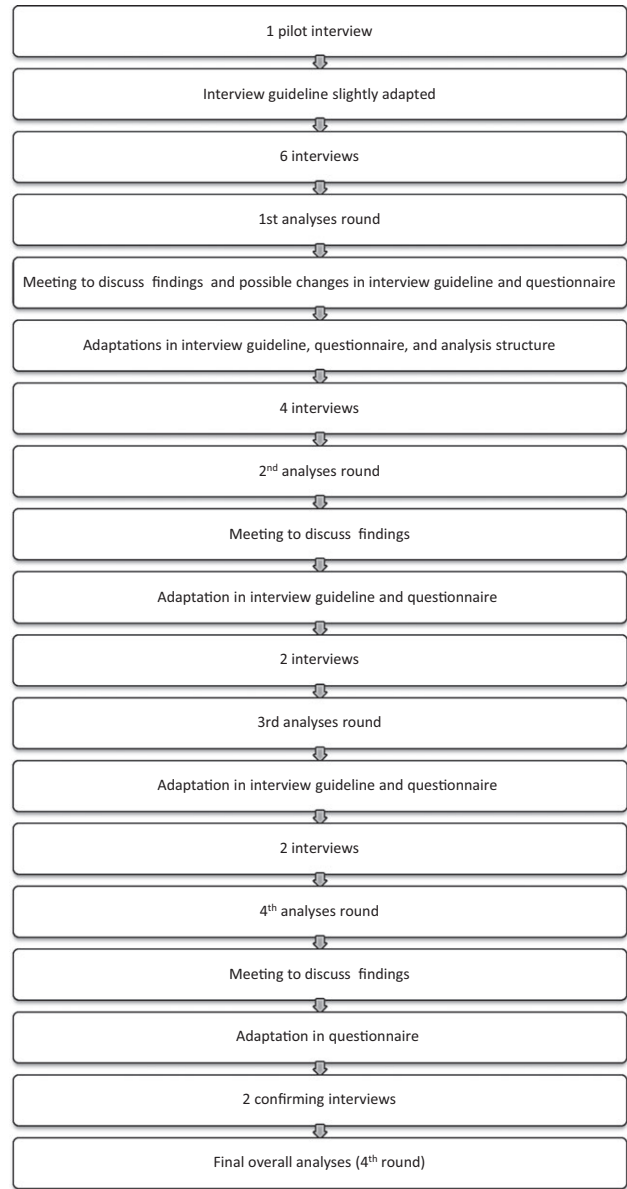
<sup>†</sup>a, rehabilitation treatment finished; b, no rehabilitation treatment undergone; c, treatment prematurely finished; d, at the beginning of rehabilitation treatment; e, waiting to start rehabilitation treatment. F, female; M, male.

**Table 2. Cognitive Components and Specific Interview Probes Based on Tourangeau,<sup>27</sup> Bloem et al.,<sup>29</sup> Patrick et al.,<sup>9</sup> and Liu et al.<sup>28</sup>**

| Main Parameters: Interpretations and Responses of the Participants, and Identification of Response Problems |  | Interview Probes Incorporated into Interview Guidelines   |
|---|--|---|
| Cognitive Component   | Definition   | Goal  |
| Comprehension   | Whether the respondent understands the question in the way the researcher intended them to | To understand the participant's interpretation of the task to be performed  |
|   |  | Can you describe any confusion/difficulty you had in understanding the instructions? What does the target construct in item (eg, participation) mean to you? Is the phrasing of the specific items in the questionnaire clear to you? What do the response scores and accompanying text mean to you? Can you explain this term/question in your own words?                                      |
| Retrieval   | Recalling relevant information   | To identify how patients retrieve information or remember situations or events  |
|   |  | Can you suggest any changes that would improve the questionnaire's wording? How did you remember that? What period of time did you think about when you were completing the questionnaire? Did you have a particular time frame in mind? Which times did you compare in order to choose an answer? What does the time frame mean to you? Did you compare yourself to someone or something else? |
| Judgment  | Making decisions regarding the response  | Can you explain how you decided to choose this response option?   |
| Reporting and response  | Matching answers to an available response option   | Were you able to find a response option for the first answer that came to your mind? What caused you to choose this response option? And why did you not choose a higher/lower score?   |
|   |  | Could you tell me how you decided on this response option and not another one, for example, XY (mention alternative response option)? Did you also consider another response option? If so, why did you consider this? Did the specified response options fit your answer?  |
| Preference  | Content coverage   | What other experiences do you have with the concepts of credibility and expectancy that are not covered in this questionnaire? Which part of the questionnaire were you more comfortable with answering?  |
|   | Format<br>Length   | What is your opinion about the format? Do you have any suggestions for improvement? What do you think about the amount of time it takes to complete the questionnaire?  |

**Table 2. (Continued)**

| Main Parameters: Interpretations and Responses of the Participants, and Identification of Response Problems |   |
|---|---|
| Cognitive Component   | Definition  |
| Editing the response  | Matching answer to an available response option   |
| Interview Probes Incorporated into Interview Guidelines   | <p>Goal</p> <p>To determine whether the length of time it takes to complete the questionnaire is reasonable</p> <p>How did you make sure that your answer fits with the available response options?<br/>                     In the case of response problems: How would you change the questionnaire so that your answer fits with the available response options?</p> |



**Figure 1.** Iterative process of interviews, analyses, and adaptations.

questions concerning the meaning of the word “credibility” and opinions regarding the scoring options (see the Results section for details).

The interviews were conducted by a researcher and a research assistant (details omitted for double-blind reviewing) trained in interview skills. The interview lasted approximately 1 hour. All interviews were audio taped.

### Main Parameters

Content validity is whether a questionnaire captures the concept to be measured adequately, and whether the

respondent understands and interprets the items of the questionnaire and decides on an answer as intended.<sup>8,26</sup>

If an assessment of content validity is dismissed, it could be because the concepts that are intended to be measured are not embodied in the actual questionnaire or are understood differently than intended by the respondents.<sup>25</sup> Therefore, the TSTI interviews to assess content validity focused on 2 main parameters<sup>23</sup>: the interpretations and responses of the participants, and the identification of response problems.

**Interpretations and Responses of the Participants.** The 4-stage model<sup>27</sup> is a common representation for the underlying cognitive processes a respondent uses to answer items in a questionnaire.<sup>10</sup> According to this model, 4 actions have to be fulfilled to answer a question: comprehending the question, retrieving necessary information, making a judgment, and responding to the question. An overview of the related cognitive components<sup>9,28,29</sup> and specific interview probes to explore these processes in depth are presented in Table 2.

**Identification of Response Problems.** While answering the questions in the CEQ, potential problems were observed and explored (eg, problematic formulations or the lack of a time frame). Those response problems link back to the previous main parameter. Response problems are, for example, problems regarding recall, time frame, and avoidance of extreme scores. Additionally, preferences of the respondent and how a respondent edited a response in the questionnaire were explored during the cognitive interview. Those explorations led to insights into a potential mismatch between the meaning behind the items of the questionnaire and the participant's behavior in answering these items. In the next step, these items were changed by the researchers and an adapted questionnaire was presented to the next respondent.

#### The Evaluated Questionnaire: The Credibility and Expectancy Questionnaire

The CEQ that was used in the PREPARE study and in the current study was adapted from the Dutch CEQ from de Leeuw et al.,<sup>15</sup> who had adapted the original version from Devilly and Borkovec<sup>16</sup> to be used in chronic pain rehabilitation.

The CEQ (see Appendix S1) consists of an explanatory preface, the instructions for part 1, part 1 (3 main items), and the instructions for part 2, part 2 (2 main items). Part 1 is related to current expectancy and

credibility regarding the treatment in general. Part 2 is related to the specific personal situation. The CEQ uses a Likert scale, ranging from 1 (not at all) to 9 (very much).

The CEQ's wording of the items is adapted to the actual pain rehabilitation treatment objective. Therefore, expectations about the treatment in general and the expected success in terms of the patient's individual increase in participation, decrease in disability, and decrease in pain intensity are rated. To clarify the meaning of the term *participation* for the respondents of the PREPARE study, an explanation was added to the CEQ in a footnote (see Appendix S1).

The CEQ has good psychometric properties (internal consistency and test-retest reliability),<sup>16</sup> and the 2-factor structure of construct credibility and expectancy has been confirmed.<sup>17</sup>

#### Data Analysis

Data were analyzed using a directed content analysis<sup>30</sup> and constant comparison<sup>31</sup> by searching for supportive and disconfirming data. For the directed content analysis, we used an a priori framework, which is a scheme of cognitive components based on earlier studies.<sup>10,23,29</sup> The components were comprehensiveness, retrieval, judgment, reporting and response, preferences, and editing the response (see Table 2).

We first read the data (interview transcripts and field notes), then marked text fragments that fit the research questions, coded them based on the above-mentioned components, and compared and contrasted text fragments and codes to deepen and broaden our analysis concerning cognitive components and response problems.

Furthermore, analysis took place in a cyclical process: Interviews and analysis alternated with meetings of the project team (authors V-C.M., A.M., and M.G.) to discuss the findings together (for a detailed overview, see Figure 1), resulting in sharpening and densifying the analysis.

Data collection and analysis were continued until analytical data saturation was reached.<sup>11</sup> Data saturation occurred after 15 interviews. To confirm the findings, another 2 interviews were held (Table 3). NVivo 9 software (QSR International, Cheshire, U.K.) was used.

Data triangulation, referring to the usage of different sources to increase validity,<sup>32</sup> was achieved by including information from different sources such as interview transcripts and field notes, as well as by combining think-aloud interviews with subsequent probing

**Table 3. Identification of Response Problems and Changes**

| Cognitive Component                            | Round in which Response Problem First Appeared |         |         |         |         |
|--|--|---------|---------|---------|---------|
|  | Round 1  | Round 2 | Round 3 | Round 4 | Round 5 |
| Comprehension                                  |  |         |         |         |         |
| Instruction                                    | x  |         |         |         |         |
| Retrieval                                      |  |         |         |         |         |
| Recall   |  |         | xx      |         |         |
| Time frame                                     | xx   |         |         |         |         |
| Standards of comparison                        |  | xx      |         |         |         |
| Judgment                                       |  |         |         |         |         |
| Avoidance of extreme values                    | x  |         |         |         |         |
| Reporting and response                         |  |         |         |         |         |
| Response options                               | x, xx  |         |         |         |         |
| Other  |  |         |         |         |         |
| Difference pain/disability resulting from pain | xx   |         |         |         |         |

x, participants who are waiting for treatment or have just started treatment; xx, participants who have already finished treatment.

questions. During the analysis, the project team (authors V-C.M., A.M., and M.G.) details omitted for double-blind reviewing) discussed the findings to strive for investigator triangulation.<sup>32</sup>

### Ethics

The Medical Ethics Committee of Medical Ethical Committee of University Hospital Maastricht and Maastricht University and Adelante zorggroep, Hoensbroek approved the study protocol. The study is registered in a public trial registry (Netherlands Trial Register NTR 3065). All participants provided written informed consent.

## RESULTS

The 2 main parameters, interpretations and responses of the participants (for specific content see Table 2) and identification of response problems (see Table 3), were used as structuring elements. This process was accompanied by making changes in the CEQ (see Appendix S1) for each round.

### Round 1

*Interpretations and Responses of the Participants and Identification of Response Problems.* Concerning the

instruction text for part 1, all participants did not immediately know what the statement “in general” was referring to when they were asked to describe problems in standard of comparison. Most participants solved this by filling in part 1 as if they were referring to their views “in general” regarding their own rehabilitation treatment rather than what was intended, which was to describe their views regarding rehabilitation treatment of everyone “in general.” All participants mentioned that they were not able to exclude their personal views while answering items regarding their expectancy “in general.” As the instruction text for part 2 referred to the specific personal situation, some participants perceived part 2 (and especially items 4a to c) as being redundant. Also, the participants did not know which reference of comparison to use.

All participants understood the word *participation* when asked about their expectations regarding increased participation. They could give a description, paraphrase it, and mention examples. However, none of the participants mentioned paid work as part of the concept of participation. Interestingly, all of their examples referred to free time, social contacts, and unpaid (voluntary) work. Also, the “quality of participation,” which refers to a secondary aspect next to frequency, was part of the participants’ definition of participation. By this, they meant that lowering the frequency of participation could improve the perceived quality of it.

Participants scored items regarding expectations of a “decrease in pain disability” (2b, 4b, 5b) and a “decrease in pain intensity” (item 2c, 4c, 5c) differently, depending on whether they were scheduled on the waiting list or had already finished pain rehabilitation treatment. Participants waiting for treatment gave rather congruent scores (meaning the same numbers) for those items. Participants who had already finished treatment emphasized during the interview that the decrease in the disability-related items (2b, 4b, 5b) was scored in a noncongruent way. They differentiated between a decrease in disability resulting from pain and a decrease in pain as 2 entirely separate concepts. Participants mentioned that this process of separation was a result of the treatment they successfully underwent:

This difference became clearer to me during the therapy. In the past, I linked these more to each other; now I’m separating them. (participant I, had finished pain rehabilitation treatment)

The rehabilitation treatment is very good. But again, the pain will remain. You have learned to



deal with it even though the pain will always be there. (participant P, had finished fibromyalgia treatment)

As the general Dutch rating scale ranges from 1 to 10 and the CEQ's response options use a 9-point Likert scale, one participant experienced the questionnaire's response rating scale as rather unusual. In subsequent interviews, it became apparent that 2 participants had filled in the questionnaire as if it used a 10-point rating scale (participants F and G).

**Changes after Round 1.** Due to the observed problems in the standard of comparison, the instruction texts were adapted. The instructions were reformulated to emphasize the participant's perspective on thoughts regarding rehabilitation treatment in general (part 1) and expectancies and expected success of the participant's own treatment specifically (part 2).

## Round 2

**Interpretations and Responses of the Participants.** Participants who were at the beginning of the treatment period or were waiting for treatment interpreted and responded to the CEQ as intended.

**Identification of Response Problems.** Another problem in the standard of comparison arose: In general, 2 different standards of comparison were used to answer the questionnaire's items. Those differed according to whether the participant was at the start of treatment or had finished it. Participants prior to the start of treatment were looking forward in terms of time to answer an item:

I do have certain expectations about how it will be afterwards compared to how it is now. This refers to how it usually is today and how I hope it will become [in the future]. Based on this, I have answered those questions. (participant N, waiting for start treatment)

Participants who had finished treatment encountered problems in choosing a suitable time frame for answering the items in part 1 and part 2 because the reference time frame had changed. The instructions for part 1 refer to the expected situation, and as a consequence, the items in part 1 refer to the future. However, these participants had already finished treatment:

The phrasing of the question is confusing "which is offered to you." I have already finished it, but it mentions this moment right now. (participant I, had finished rehabilitation treatment)

**Changes after Round 2.** As more problems were encountered by patients who had finished their treatment compared to patients at the beginning of treatment, we developed a different version of the CEQ for them. We named this version the CEQ-2. In this version, the wording of the items in terms of reference period (standard of comparison) was adapted by explicitly asking participants in the instructions to remember or imagine how they felt before or while just beginning treatment.

## Round 3 and Round 4

**Identification of Response Problems.** Participants who had already undergone treatment still encountered problems with understanding what time frame (recall period) was referred to in part 2 of the CEQ-2. Participants were still uncertain regarding the reference period to compare with because they had already finished rehabilitation treatment:

[Frame of reference, part 2:] For this, I refer to after the treatment. [Frame of reference, part 2, item 4:] This is a question about what you think about the treatment and what you expect, even though my treatment is finished. It is about the time after treatment only. [Suggestion frame of reference:] I would mention somewhere in the questionnaire to look back in time when the treatment started. (participant O, had finished rehabilitation treatment)

**Changes after Round 3 and Round 4.** By defining the recall period more specifically for the subsequent participants, we found that changes in wording as well as in tense were necessary to increase the clarity regarding the specific manner of comparison. Specifically, the CEQ-2 differs from the original version of the CEQ in the following points: The wording of the time frame was changed to present tense (items 1, 2a to c, 3); the words "attended treatment" were added (instructions for part 2, items 4a to c) to increase clarity; the recall period regarding expectations of improvement was adapted

(items 5a to c). Therein, the CEQ-2 asks patients to take themselves back in the time to before the time they started treatment.

### Round 5

*Interpretations and Responses of the Participants.* After implementing the changes in round 4, we found that the recall period was interpreted and responded to as intended by all of the participants, without causing any confusion.

### General Observations Throughout All Rounds

Two general observations were made throughout the entire interview process. The first is: When reasons for choosing a certain score were explored during probing questions, it became clear that several participants were inclined to avoid extreme values on the response option scale, indicating the probability of end-aversion bias. Participants wanted to avoid disappointments by having expectations that were too high, and participants who had not started treatment yet argued that they were uncertain about what to expect as a result of the treatment. As a consequence, they chose scores in the middle of the possible range.

During the entire interview procedure, equipoise between participants regarding the preference for or against a 10-point Likert scale occurred (see Round 2).

Furthermore, only 1 participant was able to express verbally that he could distinguish between the meanings of “to think” (in the participant’s words “to expect”) and “to hope”:

I hope that my complaints lessen. But I do not expect anything. Hope is not what you expect. I have my expectations and I have my hope. (participant L, waiting for rehabilitation treatment)

A summary of the main changes in the CEQ and the CEQ-2 is given in Box 1 and Box 2.

## DISCUSSION

The aim of this study was to test and improve the content validity of the CEQ. The main result is that patients with chronic pain waiting for or undergoing rehabilitation treatment interpreted and responded to the CEQ as intended.

**Box 1:** Recommendations for the Credibility and Expectancy Questionnaire (version for patients before or just beginning the treatment)

**Instructions**

Slight adaptations in the instructions text for part 1 and part 2

**Box 2:** Recommendations for revised Credibility and Expectancy Questionnaire (CEQ-2; for patients who are undergoing or have already finished the treatment)

Change wording regarding time frame  
Add words “attended treatment”

Items 1, 2a to c, 3  
Instructions for part 2,  
items 4a to c

Adapt recall period

Items 5a to c

Interviews with patients with chronic pain waiting for or just beginning rehabilitation treatment showed that only minimal changes in the wording of the instruction text for part 1 and part 2 of the CEQ were required. No other response problems could be identified. The only slightly necessary adaptations can be explained by the previous process of psychometric validation.<sup>16</sup> Furthermore, these slight changes indicate a direct link between some aspects of psychometric quality (test–retest reliability, factor structure, internal consistency, and the scale’s ability to predict outcome) of the CEQ resulting in the concept of credibility and expectancy and the actual meaning of the CEQ as it is used in clinical or research settings.

For patients with chronic pain who had already undergone rehabilitation treatment, we had to change the CEQ considerably, resulting in the development of a second version of the CEQ, the CEQ-2. We developed this version after round 2, and we tested it in the 3 subsequent rounds. The CEQ-2 differs from the original version of the CEQ in terms of the wording regarding the time frames and adapted recall periods (Appendix S2).

The application of the CEQ in the context of participants at different stages of treatment highlights the urgent need for increasing our insight into the process of change in expectancy.<sup>5</sup> This is important, as the predicting role of expectancy is acknowledged<sup>33</sup> and can even be modified, so it needs to be assessed as a potential process measure of treatment success.<sup>7,17</sup>

The iterative TSTI approach worked out well, like peeling an onion. We got through several layers before the underlying problems in interpretation, response, and understandability were encountered, unmasked, and ultimately solved. In each round, another problem was identified that had not been visible in the previous round. This provided depth to our study.

According to Nicklin et al.,<sup>13</sup> conducting cognitive interviews after the validation process of a questionnaire is completed would be too late to be able to make changes (eg, in the wording) that would increase its validity. Therefore, several authors recommend using an iterative approach when developing a new questionnaire.<sup>11–14</sup>

Saturation occurred after 15 interviews, and we feel that this indicates robust and complete findings and results in a content-valid measurement instrument.

We found that 1 participant was able to verbally express the difference in meaning between *think* (by mentioning “expect”) and *hope*, which was also differentiated in the concept approach of Devilly and Borkovec.<sup>16</sup> From our point of view, this indicates a successful translation and validation process and that there is no reason for concern that this specific aspect of the CEQ is not understood by patients.

We made an unexpected but confirmative determination of the discriminative ability of the CEQ: The primary goal of cognitive behavioral therapy as it relates to pain rehabilitation was increasing activity and participation, rather than decreasing pain. This is achieved by having the patient reconceptualize pain and disability resulting from pain. We found that the questionnaire items regarding the therapeutic goals differentiated between participants who had already finished treatment and those who were waiting for or had just started treatment.

We began recruiting participants by means of convenience sampling. This means that we interviewed the participants who came in first. This had the advantage that variability in the levels of functioning and the stage of treatment was present. By doing it this way, especially at the beginning, an oversampling of participants who had already finished rehabilitation treatment occurred. To explore this in more depth, we changed our sampling strategy to purposive sampling, which is commonly performed to create variation in the confirming or disconfirming cases.<sup>11</sup> This resulted in the observation that participants who had finished treatment encountered specific problems interpreting and responding to the questionnaire. Furthermore, this improved the

method as a heterogeneous range of respondents is recommended for the assessment of content validity.<sup>9</sup>

Evidence from previous studies shows that treatment expectancy is a strong predictor of treatment outcome,<sup>34</sup> and it has been postulated that expectancy is a modifiable variable of treatments.<sup>34</sup> Even in patients with chronic pain, expectancy can change over time<sup>33</sup> and seems to be modifiable.<sup>7,17</sup> Expectancy is also a predictor of long-term outcomes.<sup>33</sup> However, the whole process regarding changes in expectancy is not yet fully understood. And it is also uncertain whether a single measurement of expectancy can encapsulate the entire construct of expectancy for the whole treatment period and when expectancy should be measured.<sup>5</sup> Therefore, it is postulated that multiple measurements are needed throughout the duration of treatment to assess possible changes in expectancy.<sup>5</sup> There is a need to untangle the concept of expectancy and to estimate when and how often expectancy needs to be measured. But first, a proper measurement tool is needed that can be used in longitudinal designs, for example.<sup>5</sup> The current study makes a first step in this direction by developing a content-valid questionnaire. We developed 2 versions: (1) one that gives the opportunity to untangle the concept of expectancy at the start of and during treatment and (2) another that, for the participants who had already completed treatment, taps into what their expectancy was before or at the beginning of treatment (the CEQ-2).

A next step should be the examination of the practical applicability in a research setting, for example, by conducting studies on the correlation between the CEQ-2 and the original version of the CEQ.

Furthermore, in the future, particular attention should be paid to end-aversion bias (central tendency bias), which we observed during the cognitive interviews. If it is confirmed that patients filling in the CEQ tend to avoid the upper or lower end of the rating scale, the adaptation of answer categories toward less absolute statements at the endpoints or the addition of “throw-away” categories, which would serve as anchors at the ends of the rating scale, should be reconsidered.<sup>35</sup>

Along with Pool et al.,<sup>23</sup> we recommend that qualitative approaches should be an integral part of the development of a questionnaire. However, an iterative approach is time consuming.

The TSTI approach is a useful tool for exploring the process of interpretation and response, and for identifying response problems in the CEQ in a systematic way. More generally, qualitative methods

have an added value in the improvement of questionnaires.

All in all, this cyclical process of cognitive interviews, analysis, and adaptation of the questionnaire was an advantageous approach in the current study. By doing so, we gained a deeper understanding of how the questionnaire was interpreted by the participants and thereby could improve its validity.

As of now, the CEQ is a content-valid questionnaire with 2 versions: a minimally adapted version for patients waiting for or just beginning rehabilitation treatment and an adapted version for patients who are undergoing or have finished the treatment.

We conclude that after minimal changes the CEQ is a content-valid measurement instrument for patients waiting for treatment. However, for patients who had already undergone treatment, the content validity of the CEQ was less, and considerable improvements were needed.

### CONFLICT OF INTERESTS

There are no conflict of interests to declare.

### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** The Credibility and Expectancy Questionnaire used in the PREPARE trial.

**Appendix S2.** Adapted Credibility and Expectancy Questionnaire after content validity process (for patients before the start of the treatment).

### REFERENCES

1. Kazdin AE. Therapy outcome questions requiring control of credibility and treatment-generated expectancies. *Behav Ther.* 1979;10:81–93.
2. Auer CJ, Glombiewski JA, Doering BK, et al. Patients' expectations predict surgery outcomes: a meta-analysis. *Int J Behav Med.* 2016;23:49–62.
3. Bingel U, Wanigasekera V, Wiech K, et al. The effect of treatment expectation on drug efficacy: imaging the analgesic benefit of the opioid remifentanyl. *Sci Transl Med.* 2011;3:70ra14.
4. Lewin AB, Peris TS, Lindsey Bergman R, McCracken JT, Piacentini J. The role of treatment expectancy in youth receiving exposure-based CBT for obsessive compulsive disorder. *Behav Res Ther.* 2011;49:536–543.
5. Van Hartingsveld F, Ostelo RW, Cuijpers P, de Vos R, Riphagen I, de Vet HC. Treatment-related and patient-related expectations of patients with musculoskeletal disorders: a systematic review of published measurement tools. *Clin J Pain.* 2010;26:470–488.
6. Haanstra TM, van den Berg T, Ostelo RW, et al. Systematic review: do patient expectations influence treatment outcomes in total knee and total hip arthroplasty? *Health Qual Life Outcomes.* 2012;10:152.
7. Bialosky JE, Bishop MD, Cleland JA. Individual expectation: an overlooked, but pertinent, factor in the treatment of individuals experiencing musculoskeletal pain. *Phys Ther.* 2010;90:1345–1355.
8. Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63:737–745.
9. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. *Value Health.* 2011;14:978–988.
10. Collins D. Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res.* 2003;12:229–238.
11. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res.* 2009;18:1263–1278.
12. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Mayo FDAP-ROCMG. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health.* 2007;10(suppl 2):S94–S105.
13. Nicklin J, Cramp F, Kirwan J, Urban M, Hewlett S. Collaboration with patients in the design of patient-reported outcome measures: capturing the experience of fatigue in rheumatoid arthritis. *Arthritis Care Res.* 2010;62:1552–1558.
14. Turner RR, Quittner AL, Parasuraman BM, Kallich JD, Cleland CS, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcomes: instrument development and selection issues. *Value Health.* 2007;10(suppl 2):S86–S93.
15. Leeuw M, Goossens ME, van Breukelen GJ, et al. Exposure in vivo versus operant graded activity in chronic low back pain patients: results of a randomized controlled trial. *Pain.* 2008;138:192–207.
16. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry.* 2000;31:73–86.
17. Smeets RJ, Beelen S, Goossens ME, Schouten EG, Knottnerus JA, Vlaeyen JW. Treatment expectancy and credibility are associated with the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. *Clin J Pain.* 2008;24:305–315.

18. Wollny A, Marx G. Qualitative Sozialforschung-Ausgangspunkte und Ansätze für eine forschende Allgemeinmedizin. Teil 2: Qualitative Inhaltsanalyse vs. Grounded Theory [Qualitative social research—origins and approaches for research in family medicine]. *Z Allg Med*. 2009;85:467–476.
19. Willis GB. *Cognitive Interviewing. A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage Publications; 2005.
20. Mertens V-C, Goossens M, Verbunt J, Koke A, Smeets R. Effects of nurse-led motivational interviewing of patients with chronic musculoskeletal pain in preparation of rehabilitation treatment (PREPARE) on societal participation, attendance level, and cost-effectiveness: study protocol for a randomized controlled trial. *Trials*. 2013;14:90.
21. Battaglia M. Convenience sampling. In: Lavrakas P, ed. *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications; 2008:149.
22. Hak T, Van der Veer K, Jansen H. The Three-Step Test-Interview (TSTI): an observational instrument for pretesting self-completion questionnaires. *Surv Res Methods*. 2008;2:143–150.
23. Pool JJ, Hiralal SR, Ostelo RW, van der Veer K, de Vet HC. Added value of qualitative studies in the development of health related patient reported outcomes such as the Pain Coping and Cognition List in patients with sub-acute neck pain. *Man Ther*. 2010;15:43–47.
24. Hak T, van der Veer K, Jansen H. The Three-Step Test Interview (TSTI): An observational instrument for pretesting self completion questionnaires. ERIM reports 2004 ERS-2004-029-ORG). Available at SSRN: <https://ssrn.com/abstract=636782>.
25. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health*. 2009;12:1075–1083.
26. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. *Value Health*. 2011;14:967–977.
27. Tourangeau R. Cognitive sciences and survey methods. In: Jabine T, Straf M, Tanur J, Tourangeau R, eds. *Cognitive Aspects of Survey Methodology: Building a Bridge Between the Disciplines*. Washington, DC: National Academy Press; 1984:73–100.
28. Liu R, Buffart L, Kersten MJ, et al. Psychometric properties of two physical activity questionnaires, the AQUAA and the PASE, in cancer patients. *BMC Med Res Methodol*. 2011;11:30.
29. Bloem EF, van Zuuren FJ, Koeneman MA, et al. Clarifying quality of life assessment: do theoretical models capture the underlying cognitive processes? *Qual Life Res*. 2008;17:1093–1102.
30. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res*. 2005;15:1277–1288.
31. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *Br Med J*. 1995;311:42–45.
32. Merriam SB. *Qualitative Research: A Guide to Design and Implementation*. San Francisco: Jossey-Bass; 2009.
33. Goossens ME, Vlaeyen JW, Hidding A, Kole-Snijders A, Evers SM. Treatment expectancy affects the outcome of cognitive-behavioral interventions in chronic pain. *Clin J Pain*. 2005;21:18–26; discussion 69–72.
34. Crow R, Gage H, Hampson S, Hart J, Kimber A, Thomas H. The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review. *Health Technol Assess*. 1999;3:1–96.
35. Streiner D, Norman G. *Health Measurement Scales: A Practical Guide to their Development and Use*. New York: Oxford University Press; 1995.