



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Physiotherapy Theory and Practice*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Elvén, M., Hochwälder, J., Dean, E., Söderlund, A. (2018)

Development and initial evaluation of an instrument to assess physiotherapists' clinical reasoning focused on clients' behavior change

*Physiotherapy Theory and Practice*, 34(5): 367-383

<https://doi.org/10.1080/09593985.2017.1419521>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:mdh:diva-42887>

# Development and initial evaluation of an instrument to assess physiotherapists' clinical reasoning focused on clients' behaviour change

Maria Elvén, MSc, PT<sup>a</sup>, Jacek Hochwälder, PhD<sup>b</sup>, Elizabeth Dean, PhD, PT<sup>c</sup>, and Anne Söderlund, PhD, PT<sup>a</sup>

<sup>a</sup>*Division of Physiotherapy, School of Health, Care and Social Welfare, Mälardalen University, Västerås, Sweden,* <sup>b</sup>*Division of Psychology, School of Health, Care and Social Welfare, Mälardalen University, Eskilstuna, Sweden,* <sup>c</sup>*Department of Physical Therapy, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada.*

## Abstract

*Background and Aim:* A systematically developed and evaluated instrument is needed to support investigations of physiotherapists' clinical reasoning integrated with the process of clients' behaviour change. This study's aim was to develop an instrument to assess physiotherapy students' and physiotherapists' clinical reasoning focused on clients' activity-related behaviour and behaviour change, and initiate its evaluation, including feasibility and content validity. *Methods:* The study was conducted in three phases: 1) determination of instrument structure and item generation, based on a model, guidelines for assessing clinical reasoning, and existing measures; 2) cognitive interviews with five physiotherapy students to evaluate item understanding and feasibility; and 3) a Delphi process with 18 experts to evaluate content relevance. *Results:* Phase 1 resulted in an instrument with four domains: Physiotherapist; Input from client; Functional behavioural analysis; and Strategies for behaviour change. The instrument consists of case scenarios followed by items in which key features are identified, prioritized, or interpreted. Phase 2 resulted in revisions of problems and approval of feasibility. Phase 3 demonstrated high level of consensus regarding the instrument's content relevance. *Conclusions:* This feasible and content-validated instrument shows potential to be used in investigations of physiotherapy students' and physiotherapists' clinical reasoning, however continued development and testing is needed.

**Keywords:** Behaviour change; clinical reasoning; instrument development; physiotherapy; validity

## Background

Clinical reasoning is a core competency of physiotherapists (APTA, 2014; Cross, Hicks, and Barwell, 2001). Based on evidence and worldwide expert agreement, the physiotherapy profession has been called upon to embrace a broad vision of health and illness and systematically apply methods that support clients' health-related behaviour changes rather than focusing primarily on physical symptoms and body functions (Dean et al, 2014; Foster and Delitto, 2011). Consequently, incorporating a holistic and behavioural approach in physiotherapists' clinical reasoning processes is fundamental to practice. Given this, universities and clinical settings need to focus on identifying the strengths and challenges in physiotherapists' clinical reasoning and develop educational strategies to maximise their reasoning proficiency (Ajjawi and Smith, 2010). Enabling such efforts requires that assessment of physiotherapists' clinical reasoning is robust which has been challenging (Durning, Artino, Schuwirth, and van der Vleuten, 2013; Kreiter and Bergus, 2009). For these reasons, no gold standard instrument for health professionals' clinical reasoning exists (van der Vleuten, Norman, and Schuwirth, 2008). In the profession of physiotherapy specifically, there is a lack of an efficient and well tested clinical reasoning measure whose content and structure are based on sound theory and evidence.

Physiotherapists' clinical reasoning includes a reflexive thinking and decision-making process in which the physiotherapist synthesizes and analyses findings from the assessment and consequently selects an appropriate intervention and evaluates its effectiveness (Edwards et al, 2004; Holdar, Wallin, and Heiwe, 2013). Knowledge as well as analytical and reflective capabilities and skills are interrelated

and used throughout the process (Smith, Higgs, and Ellis, 2008; Wainwright, Shepard, Harman, and Stephens, 2011). This reasoning is a collaborative process between the physiotherapist and the client with the process overall being influenced by the context (Edwards et al, 2004; Holdar, Wallin, and Heiwe, 2013). In addition, the evidence for applying cognitive and behavioural methods in client management (Brunner et al, 2013; Linton and Shaw, 2011; Williams, Eccleston, and Morley, 2012; Åsenlöf, Denison, and Lindberg, 2009) supports the incorporation of these methods also in the clinical reasoning process. Thus, the client's behaviours and goals as well as the use of behaviour change strategies are features of a theory- and evidence-based clinical reasoning process (Elvén, Hochwälder, Dean, and Söderlund, 2015). To support investigations of physiotherapists' clinical reasoning abilities, in accordance with theory- and evidence-based methods that promote health behaviour change, a new instrument should reflect the characteristics of such an approach. A model of clinical reasoning focusing on clients' behaviour change with specific reference to physiotherapists (CRBC-PT) has been reported (Elvén, Hochwälder, Dean, and Söderlund, 2015). The model describes the conceptual meaning of the construct 'physiotherapists' clinical reasoning focused on clients' activity-related behaviour and behaviour change'. Thus, this model provides an appropriate foundation for a new instrument.

In light of the complexity of clinical reasoning, core elements have been suggested in the design of valid clinical reasoning measures. Kreiter and Bergus (2009) proposed a design that views clinical reasoning as a chain of cognitive activities. This design provides assessments of clinical reasoning with a focus on the physiotherapist's understanding of how new client information affects and potentially changes hypotheses and decisions in the reasoning process. In addition to the view of clinical reasoning as a mental process, Durning, Artino, Schuwirth, and van der Vleuten (2013) described clinical reasoning as a process that involves behaviours that are visible through the physiotherapist's decisions. Both the mental processes and the decision behaviours depend on the physiotherapist's interaction with the client as well as the context overall. Based on these factors, multiple acceptable paths can emerge to address a clinical problem in the reasoning process. Thus, to improve validity, these core design elements, along with measurement precision, need to be incorporated in the assessment of clinical reasoning (Durning, Artino, Schuwirth, and van der Vleuten, 2013).

Assessment methods that are currently incorporated in physiotherapy education consist of evaluation of students' clinical competencies (in which clinical reasoning is included) in the clinical setting, simulated patients, and written exams. However, these methods are time consuming (Ladyshevsky, Baker, Jones, and Nelson, 2000; Yeung et al, 2015), lack standardization and precision (Dalton, Davidson, and Keating, 2012; Lewis, Stiller, and Hardy, 2008; Meldrum et al, 2008), and lack core attributes proposed for the design of the assessment of clinical reasoning. Two reliable and valid clinical reasoning measures that are frequently used for medical students (Dory, Gagnon, Vanpee, and Charlin, 2012; Farmer and Page, 2005) but less commonly used by other health professionals (Dory, Gagnon, Vanpee, and Charlin, 2012), are the Key Feature approach (Page, Bordage, and Allen, 1995) and the Script Concordance Test (Charlin, Leduc, Blouin, and Brailovsky, 1998). The Key Feature approach focuses on identification and interpretation of critical stimuli (key features) in the problem-solving process (Farmer and Page, 2005). The Script Concordance Test focuses on interpretation of information comprising a certain amount of imprecision, multiple judgements that are made in the clinical reasoning process, and concordance with judgements of a panel of reference experts (Charlin et al, 2000; Fournier, Demeester, and Charlin, 2008). However, with use of only one instrument, critical aspects of the clinical reasoning process such as multiple interrelated reasoning levels or systematic considerations of new client information might be insufficiently investigated. Thus, a new clinical reasoning instrument for physiotherapists based on theory, evidence and established recommendations in the design of clinical reasoning assessments is needed. With such an instrument our understanding of physiotherapy students' and physiotherapists' clinical reasoning could improve and consequently guide the development of their clinical reasoning competency.

The development and validation of instruments begins with a theory or model that clearly describes the phenomenon to be measured (DeVellis, 2012). In the following chain of procedures,

qualitative and quantitative methods are recommended to be incorporated (Luyt, 2012). Content validation is an accepted first procedure in the evaluation of the psychometric properties of a new instrument (DeVellis, 2012; Streiner and Norman, 2008). To the best of our knowledge, there is no conceptually grounded and validated instrument to assess physiotherapists' clinical reasoning in a standard manner. The aim of this study was to develop an instrument, based on the model of CRBC-PT (Elvén, Hochwälder, Dean, and Söderlund, 2015), to assess physiotherapy students' and physiotherapists' clinical reasoning focused on clients' activity-related behaviour and behaviour change, and initiate its evaluation, including feasibility and content validity. According to the Regional Ethical Review Board, Uppsala, Sweden, this study meets the ethical requirements stated in the Swedish law (SFS 2003:460) and was accepted to be conducted (Dnr 2013/020).

## **Methods**

### ***Design and overall procedure***

An exploratory design was used to develop the structure of the instrument and to generate its constituent items, and a concurrent mixed methods design was used to evaluate the instrument's content validity (Andrew and Halcomb, 2006). The construct to be measured, 'physiotherapists' clinical reasoning focused on clients' activity-related behaviour and behaviour change' was operationalized based on its definition described in the model of CRBC-PT (Elvén, Hochwälder, Dean, and Söderlund, 2015). The guidelines outlined by DeVellis (2012) informed the three phases of the current study.

### ***Phase 1 determination of the structure of the instrument and generation of its items***

The purpose of the first phase was to determine the structure of the instrument and to generate items that captured the domains of the construct along with current recommendations for measuring clinical reasoning. The development of the instrument structure and generation of items were conducted simultaneously to ensure their congruence (DeVellis, 2012).

### ***Procedure***

The model of CRBC-PT (Elvén, Hochwälder, Dean, and Söderlund, 2015) guided the instrument structure comprising four domains; (1) Physiotherapist; (2) Input from client; (3) Functional behavioural analysis; and (4) Strategies for behaviour change. Additionally, the structure was based on the following current recommendations to assess clinical reasoning: focus on the clinical reasoning process rather than its endpoint of problem solving and diagnosis; including the collection of important information from the client; including problem identification and decisions about client management; continuously adding new information about the client; including contextual factors to be considered; accepting a range of eligible performances in the client encounters; and including various cases with several items per case (Durning, Artino, Schuwirth, and van der Vleuten, 2013; Kreiter and Bergus, 2009; Norman, Bordage, Page, and Keane, 2006).

The generation of items was based on several sources (DeVellis, 2012; Netemeyer, Bearden, and Sharma, 2003; Streiner and Norman, 2008): (a) the conceptual model of CRBC-PT; (b) existing instruments assessing clinical reasoning; (c) self-assessment scales measuring individuals' perceptions of factors affecting their behaviours and health; (d) the instrument developers themselves; and (e) experts in the field of interest (phase 3). The primary investigator, in collaboration with the research team, constructed items for the first draft of the instrument. Discussions and consensus in the team informed all revisions of the instrument.

Items for the Physiotherapist domain were developed by adapting existing self-assessment scales measuring aspects of knowledge (Amemori et al, 2011), cognition (Huijg et al, 2014), metacognition (Schraw and Dennison, 1994), and guidelines for measuring self-efficacy (Bandura, 2006) and attitudes (Montano and Kasprzyk, 2008). The Key Feature approach (Hrynchak, Glover Takahashi, and Nayer,

2014; Page, Bordage, and Allen, 1995) and the Script Concordance Test (Charlin, Leduc, Blouin, and Brailovsky, 1998; Dory, Gagnon, Vanpee, and Charlin, 2012; Lubarsky et al, 2011) informed the item construction of domains two, three and four. The content of these items was informed by the descriptions of central elements of clinical reasoning, presented in Elvén, Hochwälder, Dean, and Söderlund (2015, pp 239-241). Furthermore, items related to the domain of Input from client were influenced by clinical measures assessing these central elements (Bergström et al, 1998; Jensen and Linton, 1993; Lundberg, Styf, and Carlsson, 2004; Marcus, Selby, Niaura, and Rossi, 1992; Sechrist, Walker, and Pender, 1987; Wójcicki, White, and McAuley, 2009). The domain of Strategies for behaviour change was developed by interventions and techniques to support clients' behaviour changes based on theory and evidence (Elvén, Hochwälder, Dean, and Söderlund, 2015) and was influenced by the use of the International Classification of Functioning, Disability and Health (World Health Organization, 2001) in a physiotherapy context (Josephson, Bülow, and Hedberg, 2011). The scoring system was based on the method of the Script Concordance Test. This method takes into account the variability of experts' interpretations and judgements in their clinical reasoning. This implies that the response options of each item are assigned a credit corresponding to the proportion of experts who selected that specific response option (Dory, Gagnon, Vanpee, and Charlin, 2012; Fournier, Demeester, and Charlin, 2008).

The cases in the instrument were influenced by authentic clinical client encounters and were modified to be unidentifiable. To reflect a biopsychosocial perspective of health and illness (Gatchel et al, 2007), physical, psychological, and social attributes were incorporated in the case descriptions. The most prevailing diseases, ill-health conditions, non-communicable diseases and associated life-style risk behaviours (Public Health Agency of Sweden, 2015; Statistic Sweden, 2013; World Health Organization, 2013), as well as common clinical fields of practice for the physiotherapy profession (APTA, 2014; Broberg and Tyni-Lenné, 2009), guided the selection of cases. The choices of scaling responses were dependent upon the nature of the items and followed recommendations for scale construction (Bandura, 2006; Farmer and Page, 2005; Lubarsky et al, 2013; Montano and Kasprzyk, 2008; Streiner and Norman, 2008).

A purposive sample of four practising physiotherapists reviewed the cases in the instrument. Their clinical experience ranged from 15 to 34 years and entailed experiences working within musculoskeletal, paediatric, neurology, and primary care fields of practice. The physiotherapists' task was to detect whether the cases were realistic and included features of a typical case and to provide suggestions for improvements.

### ***Phase 2 cognitive interviews with physiotherapy students***

The purpose of phase two was to evaluate how physiotherapy students comprehend the items and response scales and perceive the feasibility of the instrument to identify the main problems in the instrument.

#### ***Sample***

Respondents were recruited based on a purposive sampling strategy to ensure variability in the profile of the respondents with regard to knowledge and experience in a behavioural medicine approach in physiotherapy. The inclusion criteria were as follows: theoretical and practical experience of clinical reasoning; variation in expected knowledge about a behavioural medicine approach in physiotherapy; and good ability to understand and speak Swedish. These criteria resulted in the decision that the respondents should attend the last semester in a three-year physiotherapy programme in Sweden. The sample of five respondents (Buers et al, 2014; Spark and Willis, 2014) consisted of three students from a physiotherapy programme with a behavioural medicine profile and two students from two physiotherapy programmes without such a profile. Four of the respondents were women and one was a man.

## *Procedure*

The data collection method comprised individual cognitive interviews (Willis, 2005). Three physiotherapy programmes were contacted and assisted in the recruitment of respondents. Prospective student participants were asked about their interest in participating in the study. Four of the interviews were conducted in a designated room at the students' university, and one was conducted via a video call on the web. The interviews were audio-recorded and conducted by the primary investigator who had mainly developed the instrument (ME). Before the interview began, informed consent was obtained, followed by information about the importance of attempting to reproduce the thoughts that arose when answering the items and the understanding that no right or wrong answers existed. The instrument was in hard copy, and the respondents answered one domain at a time followed by the interview questions. The interview included probing questions based on the cognitive process of how respondents answer survey questions (Tourangeau, Rips, and Rasinski, 2000) as well as questions regarding the instrument's feasibility (Bowen et al, 2009; Yeung et al, 2015). The interview guide (adapted from Collins, 2003; Yeung et al, 2015) is presented in Table 1. Each interview lasted between two and two-and-a-half hours.

**Table 1.** Interview guide used in the cognitive interviews  
(Adapted from Collins (2003); Yeung et al (2015))

---

Questions regarding the question-and-response process and the instrument's feasibility
Comprehension and retrieval
What did you think when answering the questions?
How easy or difficult did you find the questions to answer? Why do you think that?
What does the term X (e.g., activity-related behaviour) mean to you?
How did you understand question X?
Judgement and responding
What did you think about the response options?
Feasibility
How clear did you find the instructions for answering the questions? Why?
How strenuous did you find the entire instrument to answer? Why?
What did you think about having different cases throughout the three parts of the instrument? What do you think about following one case throughout the instrument as an alternative structure?
Do you want to add any other comments about the instrument?
Additional questions for clarity
I noticed you hesitated before you answered - what were you thinking of?
Could you elaborate more about what you were thinking?

---

## *Data analysis*

The analysis was based on a coding system for classifying questionnaire problems (Willis and Lesser, 1999). First, the audio-recordings were transcribed verbatim. Thereafter, the primary investigator (ME) and the senior investigator (AS) read the excerpt of one interview and independently marked potential problems by using a coding scheme. After a dialog about agreement on coding, a few codes were modified to achieve consensus. The primary investigator coded the remaining four interviews. The results were compiled across interviews and categorized in an analysis spreadsheet of items and codes. Finally, the similarity, prevalence and severity of the identified problems were discussed in the research team, and the group agreed on the condensed main problems and revisions.

## ***Phase 3 content validation of the instrument using the Delphi technique***

The aim of phase three was to evaluate the instrument's content validity and agreement on the items in the instrument.

## Sample

Experts were selected using well-defined criteria based on appropriate knowledge in and experience of the topic (Baker, Lovell, and Harris, 2006) and the recommendations of a sample of between ten to 20 participants to construct a homogeneous expert panel (Baker, Lovell, and Harris, 2006; Skulmoski, Hartman, and Krahn, 2007). The inclusion criteria were as follows: registered physiotherapist in Sweden; holding a PhD or PhD student in physiotherapy; conducts research within physiotherapy with a behavioural medicine approach; has academic qualifications of at least five weeks of full-time studies in behavioural medicine at the postgraduate level or experience of teaching in such a course; has experience of teaching clinical reasoning for at least two years; and provided assurance that sufficient time could be dedicated to participate in the Delphi process. Discussion within the research team as well as a snowball sampling strategy (Polit and Beck, 2010) were used to identify the experts. This strategy implied that the research team identified clinicians, teachers and researchers who they were aware of and believed could fulfil the inclusion criteria. These individuals were contacted by phone to establish their interest in participating in the study and their eligibility. Further, each individual was asked to recommend other experts who might be eligible which resulted in new contacts. In total, 19 individuals appeared eligible for inclusion, and 18 of these were interested in participating in the study. One potential expert declined participation due to time constraints. To obtain initial consent to participate, an invitation letter was sent to the experts with written questions to ensure the experts met the inclusion criteria. All 18 experts fulfilled the criteria and were included. The experts were anonymous to each other but not to the primary investigator. The included experts were employed at six universities in Sweden. Sample characteristics are presented in Table 2.

**Table 2.** Demographic characteristics of participating experts (n=18)

Characteristics	n
Gender	
Female	16
Male	2
Age (y)	
30-39	3
40-49	5
50-59	7
60-69	3
Academic qualification	
PhD student	6
PhD	8
Associate professor	1
Professor	3
Current clinical practice	7
Current teaching practice	15
Experience in teaching clinical reasoning (y)	
2-3	5
4-5	2
6-10	4
11-20	6
21-30	1
Studies in behavioural medicine at postgraduate level or experience of teaching in such courses (weeks of full-time)	
5	11
6-10	6
> 10	1

## Procedure

A modified Delphi technique with a group of experts was used to collect information about the experts' opinions regarding the item relevance and guide refinement of the items. In contrast to a classical Delphi study, which uses a qualitative first round for idea generation of items, the Delphi technique in the

present study was modified by replacing the first round with scientific literature and earlier published measures as the basis for item generation (Keeney, Hasson, and McKenna, 2006; Keeney, Hasson, and McKenna, 2011). The number of Delphi rounds was based upon predefined consensus criteria. The consensus level was set to  $\geq 78\%$  (Polit and Beck, 2006) and once the pre-determined percentage of the expert group had come to agreement on the relevance of the items, the Delphi process was terminated. Before the study began, a pilot trial was conducted with two physiotherapists. They were asked to identify ambiguities in the instructions for the Delphi exercise and its questions. The Delphi round was conducted by mail. The instrument was sent to the experts with a cover letter including instructions and a definition of the construct to be measured (Grant and Davis, 1997), a demographic questionnaire, and a letter of informed consent. The experts were asked to judge the relevance of the items as a set for each domain, and to provide an overall rating of the relevance of the instrument as a whole. The relevance ratings were performed in relation to what the domains and the instrument intended to measure and were rated on a four-point ordinal scale (1=highly relevant; 2=quite relevant; 3=somewhat relevant; and 4=not relevant). The experts were also asked to provide qualitative comments about the items. After the data analysis of round one, the experts were informed about the results of the Delphi round, the accomplished revisions of the instrument and that the Delphi process was completed as consensus was achieved.

### *Data analysis*

The scores from the relevance ratings were analysed using descriptive statistics and a Content Validity Index (CVI) (Polit and Beck, 2006) for each domain (D-CVI) and for the instrument as a whole were computed. Furthermore, a Scale-CVI/Ave (the mean of all D-CVI) was computed. The IBM statistical software SPSS for Windows, Version 22.0 (Armonk, NY:IBM Corp) was used in the analysis. The 4-point response scale was dichotomized for the analysis. If the experts rated 1 or 2, they were categorized as being in agreement about the relevance of the items. The level of consensus was defined *a priori* as D-CVI  $\geq 0.78$  and S-CVI/Ave  $\geq 0.90$  (Polit and Beck, 2006).

Only comments concerning item relevance and suggestions for improvements and refinements were considered in the analysis of the qualitative comments. Comments were highlighted if at least 25% of the respondents shared a similar opinion. Additionally, comments from only one or few respondents were highlighted if they were well motivated and concerned issues that the research team had struggled with in the instrument development process. Discussion and consensus by the research team informed decisions about removing, adding or revising items, leading to the final instrument.

## **Results**

### *Phase 1*

The Physiotherapist domain (D1) included the subdomains Knowledge (D1.1), Cognition (D1.2), Metacognition (D1.3), Psychological factors (D1.4), and Contextual factors (D1.5), and comprised self-assessments scales. The domains Input from client (D2), Functional behavioural analysis (D3), and Strategies for behaviour change (D4) were developed from case scenarios (D2 and D4 comprised two cases, and D3 comprised four cases) incorporating a certain amount of imprecision to characterize real-life clinical encounters. The cases were gradually extended with new information throughout the response process. The functional behavioural analysis domain comprised four cases to emphasize its position as the core of the reasoning process (Elvén, Hochwalder, Dean, and Soderlund, 2015). Key features included in the cases should be identified, prioritized or interpreted in the following series of items. The key features pertained to components in the management of a case, such as initial data gathering: prioritizing among activity and participation problems, hypotheses concerning the client's target behaviour, and interventions: and interpretation of the data to guide decisions and actions aimed to support clients' behaviour change.

The generated items of D1 were composed of statements concerning capabilities and skills to focus on behaviour change in clients in clinical reasoning. The generated items of D2 covered the

interview with the client and assessments and measurements of the client's activity-related behaviour and associated physical, psychological and contextual factors. The items of D3 commenced with a case scenario in which the first item consisted of six hypotheses explaining what potentially causes, controls or maintains the client's activity-related target behaviour. The hypotheses focused on physical, psychological or contextual factors, and the examinees were asked to select the three most likely hypotheses according to their opinion. The following item of D3 consisted of three parts. In the first part, each hypothesis was presented. In the second part, new information, such as new physical examination findings, expectations, feelings, beliefs, influencing contextual factors, and consequences of a behaviour that may (or may not) have an effect on the given hypothesis, were presented. In the third part, the examinee judged the effect of the new information on the proposed hypothesis. D4 commenced with a hypothesis based on a functional behavioural analysis of the client's activity-related target behaviour. According to the client's progress or the accumulation of new information, the hypothesis was reformulated. The items assessed the examinees' ability to weigh the importance of physical, psychological and contextual factors in the case and to judge and prioritize among the interventions. The response scales comprised Likert scales, write-in formats, and lists of options.

The practising physiotherapists agreed that the cases represented clients in their fields of practice. However, they suggested revisions to emphasize typical features of the cases. Accordingly, symptoms were added or changed in six cases, sex was changed in two cases, and age was adjusted in two cases.

The first phase of the instrument development process yielded the first version of the instrument including 79 items, which were distributed as follows: D1 comprised 49 items, D2 comprised two cases with six items per case; D3 comprised four cases with two items per case; and D4 comprised two cases with five items per case.

## ***Phase 2***

The cognitive interviews resulted in the identification of 171 distinct problems, which were subsequently condensed to 15 main problems. According to the coding system (Willis and Lesser, 1999), 63 problems were related to the code 'clarity vague', 42 problems were related to the code 'response categories', and the remaining codes included 2-27 problems. Sixty-three percent of the problems resulted in revisions of the instrument. The reasons for not addressing the remaining problems were minor ambiguities regarding words, respondents' ability to respond despite a perceived problem, items were based on the guidelines for the construction of the Script Concordance Test or the Key Feature approach and therefore should not be changed, or respondents had been inattentive when reading the instructions or items. Overall, respondents perceived the instrument to be feasible. None of the respondents reported responding to the instrument to be strenuous, except for reading the long list of intervention options. Furthermore, the respondents perceived that the inclusion of various cases facilitated their ability to stay focused. The condensed main problems and related codes associated with the instrument and revisions to the instrument are presented in Table 3. Phase two resulted in the second version of the instrument.

## ***Phase 3***

The pilot trial of the Delphi round resulted in minor clarifications in the cover letter and improved layout of the Delphi questionnaire. The Delphi round had a 100% response rate. Consensus was achieved for the instrument as a whole as well as for the four domains and the five subdomains of D1. These results mean that the expert group agreed that the items were relevant according to what they intended to measure and that they should be included in the instrument. Because the predefined criteria were achieved after the first round, the Delphi process was terminated and the instrument was judged to have excellent content validity (Polit and Beck, 2006). The frequencies of the experts' ratings and CVIs are presented in Table 4.

**Table 3.** Condensed main problems of the first version of the instrument identified in the cognitive interviews, related codes and revisions

Condensed main problems	Codes	Revisions
Ambiguous definitions of central concepts in the initial instructions of the instrument	Clarity – concept	The concepts were revised, definitions were clarified, and an example was added to each concept.
D1 and D2: The concept "client's problem" was unclear	Clarity – concept	The concept was replaced by either "target-behaviour" or "complaints and life situation" to clarify the intended meaning.
D1: Eligible response options were missing for the items of psychological factors	Response categories	To increase opportunities for variability, the anchor of the scale 1 = Not at all important was replaced by 1 = Not very important and the number of response options was extended from 6 to 11.
D1: Irrelevant response option	Response categories	The response option "I don't work with client's specific behaviours" was removed.
D1: Multiple ways to interpret what contextual factors comprise	Clarity – vague	An overall instruction that urged the examinee to think about his or her clinical training placement/workplace in the last six months was added.
D1 and D2: Negatively worded items	Clarity – wording	The items were revised to positively worded items.
D2: Vague response options	Response categories	The anchors of the scale 1 = Not at all prioritized and 6 = Extremely prioritized were replaced by 1 = Not very important and 6 = Extremely important.
D2, D3, and D4: Items were insufficiently related to the specific client	Clarity – vague	The client's name was added to the items.
D3: Relevant information was missing in the case descriptions	Clarity – vague	Information was added to the case descriptions to serve as a better basis for the following six hypotheses.
D3: Difficulties in judging the impact of new information on a given hypothesis due to ambiguity in the information and inappropriate response categories	Clarity - vague and Response categories	The added new information was revised to better fit the given hypothesis. The response categories were changed to -2 = The hypothesis greatly weakened and +2 = The hypothesis greatly strengthened.
D4: Sensitivity for response bias	Sensitivity /bias	The first item was revised to include a write-in format in which the examinee could write four treatments in free text.
D4: The long list of behaviour change strategies was difficult to grasp due to overly detailed specifications and examples	Response categories	The specifications and examples of the behaviour change strategies were shortened.
D4: Information was missing regarding when treatment was planned to finish	Instructions	Addition of an item that focused on the last treatment session.
D2 and D4: Difficulties in selecting a limited number of assessments and behaviour change strategies	Response categories	No revision was made. Prioritization is included in clinical reasoning and the respondents always managed to respond to the items.
All domains included items and concepts that were ambiguous.	Clarity – vague and Clarity - concept	The items and concepts were clarified by changing words or word order in the sentences and the addition of examples.

D1=Physiotherapist domain; D2= Input from client domain; D3=Functional behavioural analysis domain; D4=Strategies for behaviour change domain

**Table 4.** The experts' ratings regarding the relevance of the items related to each domain and the relevance of the entire instrument: Frequencies and Domain Content Validity Index (n=18)

Domains of the instrument	Frequencies % (n)				Domain Content Validity Index
	1	2	3	4	
D1. Physiotherapist					
D1.1 Knowledge	44% (8)	50% (9)	6% (1)	-	0.94
D1.2 Cognition	44% (8)	50% (9)	6% (1)	-	0.94
D1.3 Metacognition	44% (8)	50% (9)	6% (1)	-	0.94
D1.4 Psychological factors	61% (11)	33% (6)	6% (1)	-	0.94
D1.5 Contextual factors	28% (5)	56% (10)	17% (3)	-	0.83
D2. Input	33% (6)	44% (8)	22% (4)	-	0.78
D3. Analysis	44% (8)	56% (10)	-	-	1.0
D4. Strategies for behaviour change	39% (7)	44% (8)	17% (3)	-	0.83
The entire instrument	33% (6)	50% (9)	17% (3)	-	0.83
Scale-CVI/Ave (the mean of all D-CVIs): 0.90					

1 = highly relevant, 2 = quite relevant, 3 = somewhat relevant, and 4 = not relevant.

The qualitative analysis resulted in the identification of four problems in the instrument according to at least 25% of the experts and five problems based on well-motivated comments. In addition, many experts recommended wording changes to improve precision and clarity. The research team's discussions resulted in the removal of three items (one in D1.1 and two in D1.5), the addition of two items (in D1.1 and D1.5), and clarification of a few items and instructions by improving sentence structure. Moreover, minor revisions, such as changing words, were made to several items without affecting their meanings. A more detailed descriptive summary of the qualitative analysis of the experts' comments and the revisions is presented in Appendix A.

In terms of developing the final instrument related to physiotherapists' clinical reasoning focused on clients' activity-related behaviour and behaviour change we identified 81 items distributed across four domains. The instrument was titled 'Reasoning 4 Change'. An illustration of the Reasoning 4 Change instrument is provided in Figure 1 and the characteristics of cases, items and response scales are described in Table 5.

<b>Domain 1</b> Physiotherapist	<b>Domain 2</b> Input from client	<b>Domain 3</b> Functional behavioural analysis	<b>Domain 4</b> Strategies for behaviour change
Five subdomains including statements that concern the physiotherapist's D1.1 Knowledge (6-point Likert scale) D1.2 Cognition (6-point Likert scale) D1.3 Metacognition (6-point Likert scale) D1.4 Psychological factors; the influences of attitudes towards and self-efficacy in using a behavioural approach in clinical reasoning (11-point Likert scale) D1.5 Contextual factors; the influences of the social and physical environment at the working place (6-point Likert scale)	Two case scenarios  Item construction based on the key feature approach  Interview and assessment  Write-in format* Short list format** Long list format*** Rating of agreements with statements on a 6-point Likert scale.	Four case scenarios  Item construction based on the Script concordance test.  Hypotheses  Judgements of the hypotheses on a 5-point Likert scale ranging from -2 to +2.	Two case scenarios  Item construction based on the key feature approach and the Script concordance test. Strategies to support the client's behaviour change  Write-in format* Long list format**** Judgements of the interventions on a 5-point Likert scale ranging from -2 to +2.
	12 items	8 items	12 items
49 items	Reflect the multiple inter-related levels of the clinical reasoning process		

\*The examinee supplies his or her responses in short note form. \*\*The examinee selects his or her responses from a short list of options (n=5). \*\*\*The examinee selects his or her responses from a long list of options (n=13 and n=17). \*\*\*\*The examinee selects his or her responses from a long list of options (n=42). The options comprise interventions or techniques, arranged in ten categories, to support behaviour change and maintain achieved behaviour (e.g., monitor and discuss how health is affected by life-style; exercise of balance; modification of previous set goal in light of achievement; enable self-monitoring; and reduce perceived

**Figure 1.** An overview of the reasoning 4 Change instrument.

**Table 5.** The Reasoning 4 Change instrument: domains; case characteristics; number of items; example items; and example response scales.

Domains				
Cases		Items (n)	Example items	Example response scales
D1	Physiotherapist			
D1.1	Knowledge	8	#1 I have very good knowledge of theories and models about how behaviours are learned (e.g. operant and respondent learning)	6-point Likert scale 1 = Do not agree at all 6 = Completely agree
D1.2	Cognition	8	#4 I have good skills in analysing how the client's physical and social environment affect the performance of the target behaviour.	6-point Likert scale 1 = Do not agree at all 6 = Completely agree
D1.3	Metacognition	8	#6 When implementing treatment, I often ask myself if I have considered all possible treatment strategies to help the client achieve their target behaviour.	6-point Likert scale 1 = Do not agree at all 6 = Completely agree
D1.4	Psychological factors	20	Two initiating questions; - How important do you think it is for you to use these methods in your practical work? and - How certain are you that you can use the following methods in your practical work?	11-point Likert scale 0 = Not very important 10 = Extremely important
D1.5	Contextual factors	5	#18 Together with the client formulate SMART goals for treatment. SMART = specific, measurable, activity-related, realistic and time-specific	11-point Likert scale 0 = Cannot use at all 10 = Highly certain can use
D1.5	Contextual factors	5	#2 There are factors at my clinical training placement/workplace, that make it difficult for me to focus on clients' target behaviour and behavioural change in my clinical reasoning	6-point Likert scale 1 = Do not agree at all 6 = Completely agree
D2	Input from client Man, age 48. Persistent low back pain. Woman, age 45. Multiple Sclerosis, and smoker.	6 per case	#2 Based on what you now know about the client, you need to collect more information in your interview/case history to understand the client's complaints and situation. Several areas can be important to get more information about, but which area would you prioritise as most important?	Short list format with five options: Physical and biomedical factors; Psychological factors; Contextual factors; Behavioural skills; or Lifestyle-related risk factors
D3	Functional behavioural analysis Boy, age 5. Cerebral Palsy. Woman, age 75. Knee disorder. Woman, age 40. Stroke, overweight, and high blood pressure. Man, age 70. Foot and costae fractures, and respiratory disorder.	2 per case	#1 Based on the information you now have about the client, which three hypotheses/assumptions do you think explain the most important causes for the client's difficulty in performing the target behaviour?	Short list format with six hypotheses focused on physical, psychological or contextual factors.
D3			#2 The item begins with a short case description followed by three parts. 1. Hypothesis/assumption explaining the cause for the client's difficulty performing the target behaviour	2. ... and if you get new information that.... 3. .... do you think that this strengthens or weakens the first hypothesis/assumption?

D3	<i>Continued</i>		#2.2 Leila's ability to walk without support while carrying something is made difficult because her husband often expresses worry in this situation, which leads to Leila rarely participating in the setting of the dinner table	Leila has very little confidence in her ability to set and clear the table. She rates 3 on how certain she is that she can do this on a scale of 0=not at all certain to 10=very certain	-2 -1 0 +1 +2	The hypothesis greatly weakened The hypothesis somewhat weakened The hypothesis neither weakened or strengthened The hypothesis somewhat strengthened The hypothesis greatly strengthened
D4	Strategies for behaviour change Man, age 23. Shoulder instability Woman, age 37. Diabetes, stress-related symptoms, and physically inactive	6 per case	#2 To help the client achieve the target behaviour, you need to use different interventions/treatments. What four interventions/treatments do you think are most important at this stage and those you want to prioritise beginning with?			Write-in format

## Discussion

The instrument development process in three phases resulted in an instrument to assess physiotherapy students' and physiotherapists' clinical reasoning focused on clients' activity-related behaviour and behaviour change based on theory and evidence. The initial evaluation of the instrument's psychometric properties demonstrated that the instrument is judged to be feasible and to have excellent content validity. The structure of the instrument and the generated items reflect the definition of the construct as it is described in the model of CRBC-PT (Elvén, Hochwälder, Dean, and Söderlund, 2015), and was informed by current guidelines related to clinical reasoning assessments (phase 1). In phase 2, qualitative data regarding the question-response process demonstrated good feasibility for completing the instrument and added important information about the validity process. In phase 3, quantitative data regarding item relevance demonstrated a high level of agreement among the experts in the field. In combination, these procedures provided support for the content validity of our newly developed instrument. This is the first documented instrument that has attempted to capture the spectrum of the clinical reasoning process with the integration of features of a biopsychosocial approach and behavioural considerations in relation to behaviour change.

By following the stepwise process recommended by DeVellis (2012), our study meets the requirements of systematics and accuracy in instrument development, which can be considered a strength. Furthermore, the use of mixed methods (Cook and Beckman, 2006; Luyt, 2012; Mastaglia, Toye, and Kirstjanson, 2003), specifically the combination of interviewing and judging item relevance, has been strongly recommended in studies validating new instruments (Caro-Bautista et al, 2015; Hagen et al, 2008). Thus, the use of cognitive interviews and the Delphi process increased the thoroughness of the development process. Cognitive interviews are recommended especially when new and complex concepts should be operationalized (Drennan, 2003), as was the case in our study. Important findings were the respondents' varying understandings of central concepts related to a behavioural approach in clinical reasoning. Because instrument validity includes the extent to which respondents' thought processes represent the construct to be measured (Cook and Beckman, 2006), unambiguous concepts were revised and clarified with examples to reduce the risk of multiple interpretations. Thus, the cognitive interviews contributed valuable information to improve the instrument's content validity.

The response rate of 100% in the Delphi process indicated that the applied inclusion criteria identified individuals with high interest in the topic and that the modified Delphi technique was an appropriate choice for our study. Expert commitment is emphasized to achieve the objectives of the Delphi process (Keeney, Hasson, and McKenna, 2006). Accordingly, the high engagement can be

considered a strength. All domains received a high level of consensus from the expert panel in the first round, which implied that the domains were relevant according to their definitions. The domain of Functional behavioural analysis, based on the Script Concordance Test, achieved full consensus, which indicates that analytical capabilities are assessed within the domain. This result is supported by a study by See, Lim, and Tan (2014), which demonstrated that Script Concordance Tests evaluate reasoning abilities rather than recall of factual knowledge. Despite a separation of the domains in the instrument, the experts also agreed that the instrument has the ability to generate an overall evaluation of physiotherapists' clinical reasoning focused on clients' activity-related behaviour and behaviour change. This is of particular relevance as this clinical reasoning process consists of multiple interrelated reasoning levels and not isolated parts (Elvén, Hochwälder, Dean, and Söderlund, 2015). Thus, our findings confirmed that the instrument was sufficiently comprehensive and represented the construct.

Because the levels of D-CVI and Scale-CVI/Ave were predefined and achieved after the first Delphi round, no further rounds were required. However, the experts' comments were valuable and resulted in changing words and clarifying some items and instructions. These revisions did not change the meaning of the items, rather they made the items more easily readable and further refined the construct operationalization. Thus, these changes did not affect the content significantly such that further relevance ratings were needed. One comment, expressed by most experts, concerned the instrument being perceived as overly comprehensive. Thus, reducing the number of cases was proposed. However, this view was not shared among the physiotherapy students who endorsed the instrument as feasible. In the development process, several aspects informed the decisions about the characteristics and numbers of cases in the instrument. For example, the accumulated evidence (Dory, Gagnon, Vanpee, and Charlin, 2012; Gagnon et al, 2009) regarding case specificity (an individual's ability to successfully solve one case is not related to success on another case) (Elstein, Shulman, and Sprafka, 1978) was considered. Furthermore, the cases needed to be sufficiently numerous and diverse to ensure coverage of the content of the construct (Dory, Gagnon, Vanpee, and Charlin, 2012) and to allow generalization of the results across clinical settings (Kazdin, 2010). These considerations resulted in the research team's informed decision to maintain the number of cases in the final version of the instrument. Further testing of the instrument is needed to confirm or disconfirm its proposed construction.

Assessments of latent constructs such as clinical reasoning need to be based on clear construct definitions to ensure measurement precision (DeVellis, 2012; Streiner and Norman, 2008). A particular strength of the Reasoning 4 Change instrument lies in the solid base of the construct, described in the model of CRBC-PT (Elvén, Hochwälder, Dean, and Söderlund, 2015). The model is based on theory, the exploration of existing research, and the input of experts, which are recommended to enhance the accuracy and comprehensiveness of construct definitions (Mastaglia, Toye, and Kirstjanson, 2003; Netemeyer, Bearden, and Sharma, 2003). Durning, Artino, Schuwirth, and van der Vleuten (2013) emphasized that for the phenomenon of clinical reasoning to be captured, various theoretical views need to be considered. From a situated perspective, the clinical reasoning process is shared among the client, the therapist, and the context, and the process possesses non-linearity decisions. From the behaviourist perspective, clinical reasoning entails observable behaviours exhibited in terms of informed decisions and the selection of clinical procedures. Furthermore, clinical reasoning relies on physiotherapists' cognitive processes and information organization. These perspectives are in line with the theoretical framework described in the model of CRBC-PT (Elvén, Hochwälder, Dean, and Söderlund, 2015), which further reinforces the basis of the instrument content. A critical aspect of content validity is that the items should capture the dimensions of the phenomenon that are identified in the definition of the construct and not other dimensions, even though they might be conceptually related (DeVellis, 2012). In the current study, clients' activity-related behaviour and behaviour change informed the boundaries of the clinical reasoning construct. For example, the items of the instrument covered the identification of key features to efficiently assess and analyse clients' activity-related target behaviour, the interpretation of data to make functional behavioural analyses, the tests of hypotheses, and the selection of strategies to support clients' behaviour change. Cognitive processes are central to reasoning and need

to be embraced in a valid instrument (Kreiter and Bergus, 2009). In our instrument, non-analytical as well as analytical reasoning were included and the physiotherapists rated their knowledge, cognition and metacognition in relation to a behavioural approach in clinical reasoning. Furthermore, clinical reasoning as context dependent and collaborative was accounted for by including contextual aspects in the cases and the extent to which physiotherapists perceived that contextual factors influenced their clinical reasoning. Because collaboration is a prerequisite for applying a behavioural approach in clinical reasoning, considerations of biopsychosocial factors and the clients' perspective on their life situation were incorporated in the cases and items. For example, the clients' values and beliefs about the problem and preferences for treatment were expressed in the continually added information in the reasoning process. Accordingly, the construct underlying the instrument was distinct from but related to clinical reasoning in general.

Given the low number of reported studies that examine instruments measuring physiotherapists' clinical reasoning (Dalton, Davidson, and Keating, 2012; Ladyshevsky, Baker, Jones, and Nelson, 2000; Lewis, Stiller, and Hardy, 2008; Meldrum et al, 2008; Yeung et al, 2015) and the importance of physiotherapists' clinical reasoning competency (Cross, Hicks, and Barwell, 2001), it is likely that non-tested or poorly tested assessments are used in the physiotherapy context. In addition, based on the identified deficiencies of known measures there is a risk that physiotherapists' clinical reasoning is measured too superficially or with a lack of standardization. These shortcomings were prevented in the Reasoning 4 Change instrument by the use of a well-tested format, the inclusion of theory- and evidence-based clinical reasoning capabilities and skills, standardized case scenarios, and predetermined acceptable reasoning paths. Furthermore, the use of complementary methods, such as the Script Concordance Test and the Key Feature approach, may provide a more comprehensive appraisal of the characteristics of clinical reasoning than the use of either of these measures alone (Amini et al, 2011; Groves, Dick, McColl, and Bilszta, 2013).

Although the current study provided evidence of the instrument's feasibility and content validity, some limitations should be noted. The sample size in the cognitive interviews could be seen as a limitation. However, there is no consensus about the choice of sample size in cognitive interviewing. Beatty and Willis (2007) stated that approximately five to fifteen respondents would be a sufficient sample, whereas Blair and Conrad (2011) recommended much larger samples. Due to the small sample size, there was a risk of an incomplete identification of problems. However, studies have demonstrated that even small samples are suitable to identify significant problems and contribute to improved questionnaires (Buers et al, 2014; Spark and Willis, 2014). Furthermore, cognitive interviewing was used as part of multiple pre-test methods, which might justify the small sample size (Blair and Conrad, 2011). There are no uniform guidelines for how to collect and analyse data on cognitive interviews, and this method has been criticized for being too subjective (Conrad and Blair, 2009; Drennan, 2003). By using a standardized procedure with probing questions, transcribed audio-recordings, the use of systematic coding, and agreement among the investigators in the data analysis, the objectivity of the procedure was improved. Despite the stated limitations, the cognitive interviews provided an understanding of the respondents' responses to the items that other methods in the validation process did not illuminate. Another limitation was the potential uncertainty that not all experts in the Delphi process achieved the predetermined level of an expert in a behavioural medicine approach in physiotherapy. This uncertainty is due to the lack of a consensus definition of an expert in the literature. However, knowledge and experience are frequently used in definitions (Baker, Lovell, and Harris, 2006) and informed our inclusion criteria. The specified criteria resulted in a homogenous sample that might ensure that the 'true experts' were identified. (Baker, Lovell, and Harris, 2006). Furthermore, similarities among the experts supported the rather small sample size (Baker, Lovell, and Harris, 2006; Skulmoski, Hartman, and Krahn, 2007).

Because validity is a property of the interpretations and inferences of test scores, it must be established for each proposed condition or situation (Streiner and Norman, 2008). In the current study, physiotherapy students and physiotherapists informed the development process as well as the validation.

Thus, the established content validity applies to physiotherapy students in the latter part of their education as well as practitioners. Accordingly, the results provide a foundation to enable valid statements about physiotherapists and physiotherapy students based on their scores on the Reasoning 4 Change instrument. The present results indicate that the instrument may be useful in future investigations of physiotherapists' and physiotherapy students' clinical reasoning. Based on new knowledge generated by the instrument, factors that impact clinical reasoning may be explored, and variations in the incorporation of a holistic and behavioural approach may be revealed. However, instrument development is an ongoing process. Based on further evaluations of the instrument's psychometric properties, refinements may be necessary. Specifically, item analyses, construct validity and test-retest reliability are required to increase the accuracy of generated results and conclusions. In addition, to improve the feasibility of comprehensive investigations, an efficient and user-friendly web-based format of the instrument is warranted. This is the aim of our next study.

## Conclusions

The present study constructed the first systematically developed theory- and evidence-informed instrument to assess physiotherapy students' and physiotherapists' clinical reasoning focused on clients' activity-related behaviour and behaviour change, titled Reasoning 4 Change. The construct to be measured rests on a solid theoretical foundation, which is reflected in the instrument's structure and items, and its design is consistent with current evidence underpinning the assessment of clinical reasoning. Based on the findings of the cognitive interviews and the Delphi process, the content and feasibility of the instrument were supported, and important feedback to improve it was provided. This triangulation of methods reinforced the likelihood that only relevant content was included in the instrument, thus supporting the content validity of the instrument. The delineation of the instrument development process reveals the complexity of physiotherapists' clinical reasoning in integration with clients' behaviour change and provides keys to assess this competency. This enhanced understanding could support teaching and learning of clinical reasoning in education and practice. The instrument shows promise to be used as an assessment tool to investigate physiotherapy students' and physiotherapists' abilities to integrate a biopsychosocial approach and clients' activity-related target behaviours, analyse target behaviours and select strategies to support a behaviour change intervention. The Reasoning 4 Change instrument fills a gap in the literature and may contribute to improve physiotherapy students' and physiotherapists' clinical reasoning abilities to facilitate clients' behaviour change for the benefit of their health and wellbeing.

## Declaration of interest

The authors report no declarations of interest.

## References

- Ajjawi R, Smith M 2010 Clinical reasoning capability: Current understanding and implications for physiotherapy educators. *Focus on Health Professional Education: A Multi-disciplinary Journal* 12: 60-73.
- Amemori M, Murtomaa H, Michie S, Korhonen T, Kinnunen TH 2011 Assessing implementation difficulties in tobacco use prevention and cessation counselling among dental providers. *Implementation Science* 6: 1.
- Amini M, Moghadami M, Kojuri J, Abbasi H, Dehbozorgian M, Jafari M, Abadi AA, Molaee N, Pishbin E, Javadzade H, Kasmaee V, Shafaghi A, Vakili H, Akbari R, Sadat MA, Omidvar B, Monajemi A, Arabshahi K, Adibi P, Charlin B 2011 An innovative method to assess clinical reasoning skills: Clinical reasoning tests in the second national medical science Olympiad in Iran. *BMC Research Notes* 4: 418.
- Andrew S, Halcomb EJ 2006 Mixed methods research is an effective method of enquiry for community health research. *Contemporary Nurse: A Journal for the Australian Nursing Profession* 23: 145-153.

- APTA 2014 American Physical Therapy Association. Guide to Physical Therapist Practice 3.0. <http://guidetoptpractice.apta.org/>.
- Åsenlöf P, Denison E, Lindberg P 2009 Long-term follow-up of tailored behavioural treatment and exercise based physical therapy in persistent musculoskeletal pain: A randomized controlled trial in primary care. *European Journal of Pain* 13: 1080-1088.
- Baker J, Lovell K, Harris N 2006 How expert are the experts? An exploration of the concept of 'expert' within Delphi panel techniques. *Nurse Researcher* 14: 59-70.
- Bandura A 2006 Guide for constructing self-efficacy scales. In: Pajares F, and Urdan T (eds) *Self-efficacy beliefs of adolescents*, pp 307-337. Greenwich CT, Information age Publishing.
- Beatty PC, Willis GB 2007 Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly* 71: 287-311.
- Bergström G, Jensen IB, Bodin L, Linton SJ, Nygren AL, Carlsson SG 1998 Reliability and factor structure of the Multidimensional Pain Inventory-Swedish Language Version (MPI-S). *Pain* 75: 101-110.
- Blair J, Conrad FG 2011 Sample size for cognitive interview pretesting. *Public Opinion Quarterly* 75: 636-658.
- Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, Bakken S, Kaplan CP, Squiers L, Fabrizio C, Fernandez M 2009 How we design feasibility studies. *American Journal of Preventive Medicine* 36: 452-457.
- Broberg C, Tyni-Lenné R 2009 Sjukgymnastik som vetenskap och profession. <http://www.fysioterapeuterna.se/Professionsutveckling/Om-professionen/>
- Brunner E, De Herdt A, Minguet P, Baldew S-S, Probst M 2013 Can cognitive behavioural therapy based strategies be integrated into physiotherapy for the prevention of chronic low back pain? A systematic review. *Disability and Rehabilitation* 35: 1-10.
- Buers C, Triemstra M, Bloemendal E, Zwijnenberg NC, Hendriks M, Delnoij DMJ 2014 The value of cognitive interviewing for optimizing a patient experience survey. *International Journal of Social Research Methodology* 17: 325-340.
- Caro-Bautista J, Martín-Santos FJ, Morilla-Herrera JC, Villa-Estrada F, Cuevas-Fernández-Gallego M, Morales-Asencio JM 2015 Using qualitative methods in developing an instrument to identify barriers to self-care among persons with type 2 diabetes mellitus. *Journal of Clinical Nursing* 24: 1024-1037.
- Charlin B, Leduc C, Blouin D, Brailovsky C 1998 The diagnosis script questionnaire: A new tool to assess a specific dimension of clinical competence. *Advances in Health Sciences Education* 3: 51-58.
- Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C 2000 The Script Concordance test: A tool to assess the reflective clinician. *Teaching and Learning in Medicine* 12: 189-195.
- Collins D 2003 Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research* 12: 229-238.
- Conrad FG, Blair J 2009 Sources of error in cognitive interviews. *Public Opinion Quarterly* 73: 32-55.
- Cook DA, Beckman TJ 2006 Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal Of Medicine* 119: 166.e7-166.e16.
- Cross V, Hicks C, Barwell F 2001 Comparing the importance of clinical competence criteria across specialties: impact on undergraduate assessment. *Physiotherapy* 87: 351-367.
- Dalton M, Davidson M, Keating JL 2012 The Assessment of Physiotherapy Practice (APP) is a reliable measure of professional competence of physiotherapy students: a reliability study. *Journal of Physiotherapy* 58: 49-56.
- Dean E, Dornelas de Andrade A, O'Donoghue G, Skinner M, Umereh G, Beenen P, Cleaver S, Afzalzada D, Fran Delaune M, Footer C, Gannotti M, Gappmaier E, Figl-Hertlein A, Henderson B, Hudson MK, Spiteri K, King J, Klug JL, Laakso EL, LaPier T, Lomi C, Maart S, Matereke N, Meyer ER, M'Kumbuzi VRP, Mostert-Wentzel K, Myezwa H, Fagevik Olsén M, Peterson C, Pétursdóttir U, Robinson J, Sangroula K, Stensdotter A-K, Yee Tan B, Tschoepe BA, Bruno S, Mathur S, Wong WP 2014 The Second Physical Therapy Summit on Global Health: developing an action plan to promote health in daily practice and reduce the burden of non-communicable diseases. *Physiotherapy Theory and Practice* 30: 261-275.
- DeVellis RF 2012 *Scale Development. Theory and Applications*, 3rd edn. Thousands Oaks, SAGE Publications
- Dory V, Gagnon R, Vanpee D, Charlin B 2012 How to construct and implement script concordance tests: insights from a systematic review. *Medical Education* 46: 552-563.
- Drennan J 2003 Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing* 42: 57-63.
- Durning SJ, Artino Jr AR, Schuwirth L, van der Vleuten C 2013 Clarifying assumptions to enhance our understanding and assessment of clinical reasoning. *Academic Medicine* 88: 442-448.
- Edwards I, Jones M, Carr J, Braunack-Mayer A, Jensen GM 2004 Clinical reasoning strategies in physical therapy. *Physical Therapy* 84: 312-330.
- Elstein AS, Shulman LS, Sprafka SA 1978 *Medical Problem Solving: An analysis of clinical reasoning*. Cambridge, Massachusetts, Harvard University Press.

- Elvén M, Hochwälder J, Dean E, Söderlund A 2015 A clinical reasoning model focused on clients' behaviour change with reference to physiotherapists: Its multiphase development and validation *Physiotherapy Theory and Practice* 31: 231-243.
- Farmer EA, Page G 2005 A practical guide to assessing clinical decision-making skills using the key features approach. *Medical Education* 39: 1188-1194.
- Foster NE, Delitto A 2011 Embedding psychosocial perspectives within clinical management of low back pain: integration of psychosocially informed management principles into physical therapist practice-challenges and opportunities. *Physical Therapy* 91: 790-803.
- Fournier J, Demeester A, Charlin B 2008 Script Concordance Tests: Guidelines for construction. *BMC Medical Informatics and Decision Making* 8: 18.
- Gagnon R, Charlin B, Lambert C, Carrière B, Van der Vleuten C 2009 Script concordance testing: more cases or more questions? *Advances in Health Sciences Education* 14: 367-375.
- Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC 2007 The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychological Bulletin* 133: 581-624.
- Grant JS, Davis LL 1997 Selection and use of content experts for instrument development. *Research in Nursing and Health* 20: 269-274.
- Groves M, Dick ML, McColl G, Bilszta J 2013 Analysing clinical reasoning characteristics using a combined methods approach. *BMC Medical Education* 13: 144.
- Hagen NA, Stiles C, Nikolaichuk C, Biondo P, Carlson LE, Fisher K, Fainsinger R 2008 The Alberta Breakthrough Pain Assessment Tool for Cancer Patients: A validation study using a delphi process and patient think-aloud interviews. *Journal of Pain and Symptom Management* 35: 136-152.
- Holdar U, Wallin L, Heiwe S 2013 Why do we do as we do? Factors influencing clinical reasoning and decision-making among physiotherapists in an acute setting. *Physiotherapy Research International* 18: 220-229.
- Hrynchak P, Glover Takahashi S, Nayer M 2014 Key-feature questions for assessment of clinical reasoning: a literature review. *Medical Education* 48: 870-883.
- Huijg JM, Gebhardt WA, Dusseldorp E, Verheijden MW, van der Zouwe N, Middelkoop BJC, Crone MR 2014 Measuring determinants of implementation behavior: Psychometric properties of a questionnaire based on the theoretical domains framework. *Implementation Science* 9: 33.
- Jensen IB, Linton SJ 1993 Coping Strategies Questionnaire (CSQ): Reliability of the Swedish version of the CSQ. *Scandinavian Journal of Behaviour Therapy* 22: 139-145.
- Josephson I, Bülow P, Hedberg B 2011 Physiotherapists' clinical reasoning about patients with non-specific low back pain, as described by the International Classification of Functioning, Disability and Health. *Disability and Rehabilitation* 33: 2217-2228.
- Kazdin AE 2010 *Research design in clinical psychology*, 4th edn. Boston, Allyn and Bacon.
- Keeney S, Hasson F, McKenna H 2006 Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *Journal of Advanced Nursing* 53: 205-212.
- Keeney S, Hasson F, McKenna H 2011 *The Delphi technique in nursing and health research*. West Sussex, Wiley-Blackwell.
- Kreiter CD, Bergus G 2009 The validity of performance-based measures of clinical reasoning and alternative approaches. *Medical Education* 43: 320-325.
- Ladyshevsky R, Baker R, Jones M, Nelson L 2000 Evaluating clinical performance in physical therapy with simulated patients. *Journal of Physical Therapy Education* 14: 31-37.
- Lewis LK, Stiller K, Hardy F 2008 A clinical assessment tool used for physiotherapy students - is it reliable? *Physiotherapy Theory and Practice* 24: 121-134.
- Linton SJ, Shaw WS 2011 Impact of psychological factors in the experience of pain. *Physical Therapy* 91: 700-711.
- Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CPM 2011 Script concordance testing: a review of published validity evidence. *Medical Education* 45: 329-338 310p.
- Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B 2013 Script concordance testing: From theory to practice: AMEE Guide No. 75. *Medical Teacher* 35: 184-193.
- Lundberg MKE, Styf J, Carlsson SG 2004 A psychometric evaluation of the Tampa Scale for Kinesiophobia - from a physiotherapeutic perspective. *Physiotherapy Theory and Practice* 20: 121-133.
- Luyt R 2012 A framework for mixing methods in quantitative measurement development, validation, and revision: A case study. *Journal of Mixed Methods Research* 6: 294-316.
- Marcus BH, Selby VC, Niaura RS, Rossi JS 1992 Self-efficacy and the stages of exercise behavior change. *Research Quarterly for Exercise and Sport* 63: 60-66.
- Mastaglia B, Toye C, Kirstjanson LJ 2003 Ensuring content validity in instrument development: challenges and innovative approaches. *Contemporary Nurse* 14: 281-291.
- Meldrum D, Lydon A-M, Loughnane M, Geary F, Shanley L, Sayers K, Shinnick E, Filan D 2008 Assessment of undergraduate physiotherapist clinical performance: investigation of educator inter-rater reliability. *Physiotherapy* 94: 212-219.

- Montano DE, Kasprzyk D 2008 Theory of reasoned action, Theory of planned behaviour, and the Integrated behavioural model. In: Glanz K, Rimer B K, and Viswanath K (eds) Health behaviour and health education. Theory, research, and practice, 4th edn, pp 67-96. San Francisco, Jossey-Bass.
- Netemeyer RG, Bearden WO, Sharma S 2003 Scaling Procedures. Issues and Applications. Thousands Oaks, Sage Publications.
- Norman G, Bordage G, Page G, Keane D 2006 How specific is case specificity? *Medical Education* 40: 618-623.
- Page G, Bordage G, Allen T 1995 Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine* 70: 194-201.
- Polit DF, Beck CT 2006 The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in Nursing and Health* 29: 489-497.
- Polit DF, Beck CT 2010 *Essentials of nursing research. Appraising evidence for nursing practice*, 7th edn. Philadelphia, Lippincott Williams & Wilkins.
- Public Health Agency of Sweden 2015 Nationella folkhälsoenkäten- Hälsa på lika villkor. <http://www.folkhalsomyndigheten.se/amnesomraden/statistik-och-undersokningar/enkater-och-undersokningar/nationella-folkhalsoenkaten/>
- Schraw G, Dennison RS 1994 Assessing metacognitive awareness. *Contemporary Educational Psychology* 19: 460-475.
- Sechrist KR, Walker SN, Pender NJ 1987 Development and psychometric evaluation of the exercise benefits/barriers scale. *Research in Nursing and Health* 10: 357-365.
- See KC, Lim TK, Tan KL 2014 The script concordance test for clinical reasoning: Re-examining its utility and potential weakness. *Medical Education* 48: 1069-1077.
- SFS 2003:460. The Swedish Code of Statutes. The Ethical Review Act. Department of Education and Research. Stockholm. [https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/lag-2003460-om-etikprovning-av-forskning-som\\_sfs-2003-460](https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/lag-2003460-om-etikprovning-av-forskning-som_sfs-2003-460)
- Skulmoski GJ, Hartman FT, Krahn J 2007 The Delphi method for graduate research. *Journal of Information Technology Education* 6: 1-21.
- Smith M, Higgs J, Ellis E 2008 Characteristics and processes of physiotherapy clinical decision making: a study of acute care cardiorespiratory physiotherapy. *Physiotherapy Research International* 13: 209-222.
- Spark MJ, Willis J 2014 Application of cognitive interviewing to improve self-administered questionnaires used in small scale social pharmacy research. *Research in Social and Administrative Pharmacy* 10: 469-474.
- Statistic Sweden 2013 Living Condition Surveys (ULF/SILC) 2012-2013. [http://www.scb.se/sv/\\_Hitta-statistik/Statistik-efter-amne/Levnadsforhallanden/Levnadsforhallanden/Undersokningarna-av-levnadsforhallanden-ULFSILC/-c\\_li\\_354235](http://www.scb.se/sv/_Hitta-statistik/Statistik-efter-amne/Levnadsforhallanden/Levnadsforhallanden/Undersokningarna-av-levnadsforhallanden-ULFSILC/-c_li_354235)
- Streiner DL, Norman GR 2008 *Health measurement scales. A practical guide to their development and use.*, 4th edn. Oxford, University Press.
- Tourangeau R, Rips LJ, Rasinski K 2000 *The psychology of survey response*. Cambridge, Cambridge University Press.
- van der Vleuten CPM, Norman GR, Schuwirth LT 2008 Assessing clinical reasoning. In: Higgs J, Jones M A, Loftus S, and Christensen N (eds) *Clinical reasoning in the health professions*, 3rd edn, pp 413-421. Amsterdam, Butterworth Heinemann.
- Wainwright SF, Shepard KF, Harman LB, Stephens J 2011 Factors that influence the clinical decision making of novice and experienced physical therapists. *Physical Therapy* 91: 87-101.
- Williams A, Eccleston C, Morley S 2012 Psychological therapies for the management of chronic pain (excluding headache) in adults. *Cochrane Database of Systematic Reviews* 11:CD007407.
- Willis GB 2005 *Cognitive Interviewing. A tool for improving questionnaire design*. Thousands Oaks, CA, SAGE Publications, Inc.
- Willis GB, Lesser JT 1999 The Question Appraisal System QAS-99. <http://appliedresearch.cancer.gov/areas/cognitive/guides.html>
- Wójcicki TR, White SM, McAuley E 2009 Assessing outcome expectations in older adults: The multidimensional outcome expectations for exercise scale. *Journals of Gerontology: Psychological Sciences and Social Sciences* 64: 33-40.
- World Health Organization 2001 *International Classification of Functioning, Disability, and Health (ICF)*. <http://www.who.int/classifications/icf/en/>
- World Health Organization 2013 *Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013-2020*. [http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236\\_eng.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236_eng.pdf?ua=1)
- Yeung E, Woods N, Dubrowski A, Hodges B, Carnahan H 2015 Sensibility of a new instrument to assess clinical reasoning in post-graduate orthopaedic manual physical therapy education. *Manual therapy* 20: 303-312.

**Appendix A.** A summary of the qualitative analysis of the experts' comments regarding item relevance and the revisions of the instrument

Domain	Similar comments expressed by $\geq 25\%$ of the respondents	Well motivated comments	Previous content or formulation	Revised content or formulation
D1. Physiotherapist	The initial instruction for knowledge, cognition and metacognition are unambiguous and can be misinterpreted.	-	<i>Relate to what you think is a reasonable level for a newly graduated physiotherapist when you answer the questions.</i>	The instruction was removed
D1.1 Knowledge	-	Add an item that concern knowledge about pathology and diseases.	-	New item 5. <i>"I have very good theoretical knowledge about how injuries and diseases can affect bodily and mental functions"</i>
		Items 7 and 8 are overlapping	Item 7. <i>"I know which treatments are effective for supporting clients' behavioural change based on evidence (research and clinical experience)"</i>  Item 8. <i>"I know which treatments are effective for supporting clients' behavioural change based on my own practical experience"</i>	Clarification of item 7 <i>"I know which methods are effective for supporting clients' behavioural change based on evidence (i.e. best available scientific knowledge)"</i>  Clarification of item 8. <i>"I know which methods are effective for supporting clients' behavioural change based on my own practical experience"</i>
				Due to changes in items 7 and 8, the item 6 became redundant and was removed: <i>"I have very good theoretical knowledge about how I support clients' behaviour change"</i>
D1.2 Cognition	-	Unclear if the client's thoughts and feelings and psychological factors should be perceived as synonyms or different constructs.		Thoughts and feelings were replaced by psychological factors throughout the instrument.
D1.3 Metacognition	-	-		
D1.4 Psychological factors	-	-		

(Continued)

(Continued)

D1.5 Contextual factors	Item 3 is not relevant		<i>Item 3. "Many clients don't want to take the responsibility that is needed to accomplish a behaviour change"</i>	Item 3 was removed
	Doubtful whether the items cover all relevant aspects of contextual factors.		Item 6. <i>"My supervisors'/colleagues' way of performing clinical reasoning affects my way of reasoning a lot"</i>	Item 6 was removed. A new item 5 was added: <i>"At my clinical training placement/workplace, there is not the time needed to focus on clients' target behaviour and behavioural change in my clinical reasoning"</i>
		Item 2 includes a value labels that is unnecessary.	Item 2 <i>"many factors..."</i>	The value label many was removed
D2. Input from client	Two factors are not enough.	Item 1 is unambiguous and the words "at this time-point" should be clarified.	Item 1. <i>"What two factors in the interview do you think are the most important to focus on at this time-point to better understand the clients' complaints and situation?"</i>	Clarifications of item 1. <i>"What three factors in the interview/case history do you think are the most important to ask more about in your first consultation in order to understand the clients' complaints and situation?"</i>
		Item 4. The most important target behaviour, based on the client's perspective, is not clearly described in the case	Item 4 Josph: <i>"He misses the daily walks...."</i> Item 4 Caroline: <i>"It feels like a major problem for Caroline because cooking is...."</i>	Clarifications of item 4 Josph: <i>"What he missis the most is his daily walks..."</i> Clarifications of item 4 Caroline: <i>"Right now Caroline feels that this is the biggest problem because cooking is..."</i>
D3. Functional behavioural analysis	-	-		
D4. Strategies for behaviour change	-	-		
Overall judgement of the instrument	The instrument may be too long and comprehensive. There is a risk that the respondents become wearied. Consider reducing the number of items or cases.	-		The research team considered these comments. However, currently the instrument is used for research purposes and in order to provide reliable scores and avoid the risk of case specificity several cases are needed (Dory, Gagnon, Vanpee, and Charlin, 2012). Therefore, the number of cases was not diminished.