A state-of-the-art review of direct observation tools for assessing competency in person-centred care

Nina Ekman\textsuperscript{a,b,c,*}, Charles Taft\textsuperscript{a,b}, Philip Moons\textsuperscript{a,b,c}, Åsa Mäkitalo\textsuperscript{d}, Eva Boström\textsuperscript{e}, Andreas Fors\textsuperscript{a,b,f}

\textsuperscript{a} Institute of Health and Care Sciences, Sahlgrenska Academy, University of Gothenburg, Box 457, 405 30 Gothenburg, Sweden
\textsuperscript{b} Centre for Person-Centred Care (GFCC), University of Gothenburg, Sweden
\textsuperscript{c} Academic Centre for Nursing and Midwifery, Faculty of Medicine, KU Leuven, Leuven, Belgium
\textsuperscript{d} Department of Education, Communication and Learning, University of Gothenburg, Gothenburg, Sweden
\textsuperscript{e} Department of Nursing, University of Örebro, Örebro, Sweden
\textsuperscript{f} Närhälsan Research and Development Primary Health Care, Region Västra Götaland, Sweden

\textbf{A R T I C L E    I N F O}

\textbf{Article history:}
Received 19 September 2019
Received in revised form 21 January 2020
Accepted 22 April 2020

\textbf{Keywords:}
Observation-based methods
Patient-centred care
Person-centred care
State-of-the-art review

\textbf{A B S T R A C T}

\textbf{Background:} Direct observation is a common assessment strategy in health education and training, in which trainees are observed and assessed while undertaking authentic patient care and clinical activities. A variety of direct observation tools have been developed for assessing competency in delivering person-centred care (PCC), yet to our knowledge no review of such tools exists.

\textbf{Objective:} To review and evaluate direct observation tools developed to assess health professionals' competency in delivering PCC.

\textbf{Design:} State-of-the-art review

\textbf{Data sources:} Electronic literature searches were conducted in PubMed, ERIC, CINAHL, and Web of Science for English-language articles describing the development and testing of direct observation tools for assessing PCC published until March 2017.

\textbf{Review methods:} Three authors independently assessed the records for eligibility. Duplicates were removed and articles were excluded that were irrelevant based on title and/or abstract. All remaining articles were read in full text. A data extraction form was developed to cover and extract information about the tools. The articles were examined for any conceptual or theoretical frameworks underlying tool development and coverage of recognized PCC dimensions was evaluated against a standard framework. The psychometric performance of the tools was obtained directly from the original articles.

\textbf{Result:} 16 tools were identified: five assessed PCC holistically and 11 assessed PCC within specific skill domains. Conceptual/theoretical underpinnings of the tools were generally unclear. Coverage of PCC domains varied markedly between tools. Most tools reported assessments of inter-rater reliability, internal consistency reliability and concurrent validity; however, intra-rater reliability, content and construct validity were rarely reported. Predictive and discriminant validity were not assessed.

\textbf{Conclusion:} Differences in scope, coverage and content of the tools likely reflect the complexity of PCC and lack of consensus in defining this concept. Although all may serve formative purposes, evidence supporting their use in summative evaluations is limited. Patients were not involved in the development of any tool, which seems intrinsically paradoxical given the aims of PCC. The tools may be useful for providing trainee feedback; however, rigorously tested and patient-derived tools are needed for high-stakes use.

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

* Corresponding author at: Institute of Health and Care Sciences, Sahlgrenska Academy, University of Gothenburg, Sweden. Box 457, 405 30 Gothenburg, Sweden.

E-mail address: nina.ekman@gu.se (N. Ekman).

https://doi.org/10.1016/j.ijnurstu.2020.103634

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)
What is already known about the topic?

- Person-centred care (PCC) has been designated and endorsed by professional bodies as a core competency needed for health professionals.
- Numerous direct observation tools have been developed for use specifically in assessing skills in delivering PCC.
- To our knowledge, no review has been conducted of direct observation tools for use in assessing PCC competency.

What this paper adds

- This state-of-the-art review identified 16 tools and found differences in their scope, coverage and content as well as a lack of consensus in defining PCC.
- Although all may serve formative purposes, evidence supporting their use in summative evaluations is limited.
- Paradoxically, given the aims of PCC, patients were not involved in the development of any of the tools.

1. Background

Widely acknowledged as an essential element of high quality care ([Institute of Medicine, 2001; World Health Organization, 2006; World Health Organization, 2008; Australian Commission on Safety and Quality in Health Care (ACSQHC), 2011; Goodrich and Cornwell, 2008; Agency for Healthcare Research and Quality (AHRQ), 2003; International Alliance of Patient’s Organizations, 2007; Socialstyrelsen, 2016]), person-centred care (PCC) has been designated and endorsed by professional bodies as one of a set of five core competencies needed for health professionals to meet the evolving challenges facing health care ([Institute of Medicine (US) 2003; World Health Organization 2005; World Health Organization 2003; Accreditation Council for Pharmacy Education (ACPE) 2015]). The particular importance of PCC competency was underscored when, in a follow-up report to their landmark Crossing the Quality Chasm, the US Institute of Medicine (IOM) positioned PCC as the central, overarching competency, recommending that All health professionals should be educated to deliver patient-centred care as members of an interdisciplinary team, emphasizing evidence-based practice, quality improvement approaches, and informatics. (9, chapter 3) Reflecting its importance for quality care, PCC is increasingly incorporated into education and training programs for health professionals (Dwamena et al., 2012).

Direct observation is a common assessment strategy in health education and training, in which trainees are observed and assessed while undertaking authentic patient care and clinical activities. Historically, such assessments have been implicit, unstandardized, and based on global, subjective judgments (Van der Vleuten, 1996); however, today a variety of checklists, rating scales and coding systems are available to guide, delineate and structure assessments of clinical competencies (Kogan et al., 2009). Likewise, numerous direct observation tools have been developed for use specifically in assessing skills in delivering PCC. As little agreement exists within and across professions on PCC nomenclature (Håkansson Eklund et al., 2018), definitions (Socialstyrelsen, 2016; Constand et al., 2014) and theoretical or conceptual frameworks (Mead and Bower, 2000; Lawrence and Kinn, 2012; Cronin, 2004; Scholl et al., 2014), it seems reasonable to assume that such tools differ significantly in their definitions and coverage of behaviours and behavioural domains indicative of PCC competency. To our knowledge no state-of-the-art review has been conducted of direct observation tools for use in assessing PCC competency. Given the increasing importance of PCC in healthcare education and training there is a need for guidance in selecting among existing direct observation tools for assessing trainees’ competency in delivering PCC. This review therefore aims to identify available tools and evaluate them with respect to: assessed competency domain or domains; existence of underlying theoretical or conceptual frameworks; coverage of recognized components of PCC; types of behavioural indicators; psychometric performance; and format (checklist, rating scale, coding system).

Although differences exist in the definitions of person-centred care and patient-centred care, the terms are frequently used interchangeably in the literature. For the sake of parsimony, the term person-centred care has systematically been used in this review.

2. Method

2.1. Search strategies for identification of studies

The search was performed in March 2017 using the databases PubMed, CINAHL and Scopus to identify relevant studies. Centred and centeredness (both UK and US spelling) and the following terms were identified and used in the search-string: (care OR healthcare) AND (“patient-centred” OR “patient centred” OR “person centred” OR “person centred” OR “patient centeredness” OR “patient centredness” “ OR “person centeredness” “ OR “person centredness”) AND (observ* OR video OR audio) AND (Humans[Mesh] AND (Danish[lang] OR Norwegian[lang] OR Swedish[lang] OR English[lang])).

2.2. Selection of studies

EndNote X7 software was used. Duplicates were removed and clearly irrelevant articles were excluded. Criteria for inclusion were: (i) direct observation tool (ii) reports and/or descriptions of any development or evaluation of an instrument that measures patient-centred care, PCC or person centredness (iii) not clinical encounters. The selection process started with examining the titles. After inclusion of relevant titles, the same procedure was performed with the abstracts. Abstracts that clearly did not match the inclusion criteria were removed. The articles with potential to be included were read in full text in order to determine whether to include or exclude. Snowballing was used to identify other potentially relevant records. Three authors (NE, CT and AF) independently assessed the records for eligibility and recorded the reasons for either inclusion or exclusion, which was documented in a PRISMA flowchart.

2.3. Data extraction

A data extraction form was developed to cover, identify and extract information about the direct observation tool including name of tool, main concept assessed, characteristics of the development sample, assessment format, assessed domains of person-centred, conceptual framework underpinning the tools, type of psychometric assessments performed (reliability and validity), and developer identified limitations.

3. Analyses

Each tool was examined against a standard framework for coverage of PCC dimensions. For this purpose the framework endorsed by the IOM (Institute of Medicine, 2001) was used. The framework includes six dimensions: Respect for patients’ values, preferences, and expressed needs, Coordination and integration of care, Information, communication, and education, Physical comfort, Emotional support—relieving fear and anxiety and Involvement of family and friends (Institute of Medicine, 2001). Articles were examined for evidence of theoretical or conceptual frameworks underpinning the development of the tools. If the articles referred to one specific framework and used phrases such as: “we operationalized PCC...” or
“person-centred care using the framework of...” or “building on previous conceptual basis we developed...” It was judged to have a theoretical or conceptual framework; those referring to several different frameworks were classified as unclear. Information about psychometric performance of the tools was obtained directly from the original articles and supplemented by a review of references from article bibliographies.

4. Result

4.1. Search results

An initial search yielded 2371 non-duplicated records (Fig. 1). After screening by titles, 91 abstracts were read. After excluding articles based on abstracts, 42 articles were read in full text. Thirteen full-texts met inclusion criteria and six additional articles were identified by snowballing and found to be eligible. In total, 19 papers describing 16 different direct observation tools for assessing PCC or a specific aspect of PCC were identified (Fig. 1).

4.2. Characteristics of the direct observation tools

Eleven of the 16 direct observation tools were coding systems (Bertakis and Azari, 2011; Henbest and Stewart, 1989; Zandbelt et al., 2005; Paul-Savoie et al., 2015; Braddock et al., 1997; Clayman et al., 2012; Dong et al., 2014; Sabee et al., 2020; Mjaaland and Finset, 2009; Krupat et al., 2006; D’Agostino and Bylund, 2014), three were rating scales (Elwyn et al., 2003; Shields et al., 2005; Gallagher et al., 2001) and two were checklists (Gaugler et al., 2013; Chesser et al., 2013) (Table 1). Eleven tools focused on specific aspects of PCC, namely communication and shared decision-making, whereas five tools purported to assess the general concepts PCC or person-centredness. Various conceptual frameworks were presented as starting points for developing the tools. A framework was clearly identified in seven tools (Henbest and Stewart, 1989; Zandbelt et al., 2005; Clayman et al., 2012; Elwyn et al., 2003; Chesser et al., 2013; Sabee et al., 2020; D’Agostino and Bylund, 2014), while the others referred to different definitions and concepts and no specific framework could be distinguished. The tools varied in their coverage of the six IOM domains (Table 2) from a single domain (D’Agostino and Bylund, 2011 October; Gallagher et al., 2001) to all six domains (Bertakis and Azari, 2011; Chesser et al., 2013), with an average of three domains per tool (Gaugler et al., 2013; Henbest and Stewart, 1989; Braddock et al., 1997; Clayman et al., 2012; Elwyn et al., 2003; Mjaaland and Finset, 2009; Krupat et al., 2006).

Inter-rater reliability was reported for all tools except two (Bertakis and Azari, 2011; Clayman et al., 2012) (Table 2). Reliability was estimated using a variety of methods, where the intra-class correlation (ICC) and Cohen’s kappa were most common. ICC coefficients were fair to excellent (Hallgren, 2012) ranging from 0.53 (PBCI subscale inhibiting behaviour) to 0.93 (PBCI subscale facilitating behaviour and SOS-PC). Kappa coefficients reflected generally moderate to substantial agreement (Hallgren, 2012), ranging between 0.46 (PISCH subscale enabling self-management) and 0.72 (PISCH subscale fostering relationships). Intra-rater reliability was reported for three tools, where stability in codings was high to perfect for the NAAS subscales ($r = 0.82–1.0$) and satisfactory for OPTION ($r = 0.66$).

Most of the tools (12 of 16) reported some evaluation of validity, where construct and content validity were most frequently assessed. About half of the tools were developed and evaluated in primary care settings (Bertakis and Azari, 2011; Henbest and Stewart, 1989; Braddock et al., 1997; Elwyn et al., 2003; Shields et al., 2005; Mjaaland and Finset, 2009; Krupat et al., 2006; Gallagher et al., 2001) and all but two (Zandbelt et al., 2005; Mjaaland and Finset, 2009) were developed in English speaking countries.

4.3. Description of the direct observation tools

The following section provides a brief overview of information about the included observation tools regarding their format, coverage, scoring, conceptual framework and psychometric evaluations or performance as reported in the original articles. The tools are organized in accordance with Table 1, i.e. tools for assessing global PCC/ person centredness, shared-decision making, person-centred communication and nonverbal person-centred communication, and alphabetically within each category.

4.3.1. Global person-centred care/ person centredness

**COT** is a 16-item checklist for assessing global PCC. The checklist assesses if the staff performs behaviours indicative of PCC, for example “speaks to a resident at least a total of 15 s during care interaction”. Items are scored as 1 if the behaviour is observed and 0 if not observed and scores are summed, where 16 represents maximum PCC. Inter-rater reliability, assessed across five raters (interdisciplinary reviewers) was satisfactory (ICC of all Kappa coefficients=0.77. Face validity included verbal and written feedback from scientific experts on earlier versions to refine and revise the tool. Content validity was assessed based on feedback from nine interdisciplinary scientific experts regarding 31 care worker-dementia patient interactions and open-ended feedback on items (Gaugler et al., 2013).

**Modified version of DOC** is a coding system for assessing global person-centred practice style and includes six different clusters (e.g., technical and health behaviour) of physician practice behaviours among the 20 DOC codes (e.g., structuring interaction, health education and health knowledge). For each DOC code, the number of intervals during which the associated behaviour is observed is recorded and is expressed as a percentage of the total of all DOC-coded behaviours noted during the visit. A total of 509 videotaped encounters between patients and family physicians or general internists were used for the development of the modified version of DOC. Reliability and validity are documented for the original DOC but have not been assessed for the modified version (Bertakis and Azari, 2011).

**Henbest and Stewart instrument** is a coding system assessing doctors’ global person-centred behaviour in primary care consultations. The method involves identifying patients’ offers, defined as any symptom, complaint, thought, feeling, expectation or observation expressed by the patient. The assessor then rates the doctors’ responses to these offers on a four-point scale (0–3). The total score for a consultation divided by the number of offers assessed gives the person-centredness score for that consultation. This tool showed high inter-rater (18 tapes analyzed) and intra-rater reliability after two weeks (two tapes analyzed) (Spearman correlation=0.91 and 0.88, respectively). After six weeks (12 tapes analyzed) the intra-rater coefficient decreased to 0.63. The tool was sensitive to differences among physicians and among physicians’ responses to different patient offers. Analysis of 12 tapes with two raters comparing this tool with that of Brown and colleagues showed moderate to high criterion validity (Spearman correlation 0.51–0.89) (Henbest and Stewart, 1989).

**PBCI** is a coding system assessing global person-centredness in two dimensions. The first dimension, Facilitating behaviours, has 11 categories (e.g., open and closed questions). The second dimension, Inhibiting behaviours, includes 8 categories (e.g., changing focus). These two dimensions are coded in relation to content (medical, psycho-social and other). Reliability and validity were evaluated in a sample of 323 videotaped appointments between residents
<table>
<thead>
<tr>
<th>Direct observation tool¹</th>
<th>Competency</th>
<th>Format/ content</th>
<th>Conceptual basis</th>
<th>IOM domains ³</th>
<th>Development setting</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The CARES Observational tool (COT) (Gaugler et al., 2013)</td>
<td>Person-centred Care</td>
<td>Checklist, 16 items</td>
<td>Unclear¹</td>
<td>X X X</td>
<td>Dementia home care, USA, Care workers-patients</td>
<td>Inter-rater reliability: Intraclass correlation coefficient (ICC)=0.77 N interactions: 5 N raters: 5</td>
<td>Face validity: PI with input from scientific advisors reviewed Content validity: panel of several interdisciplinary experts</td>
</tr>
<tr>
<td>Modified version of the Davis Observation Code (DOC) (Bertakis and Azari, 2011)</td>
<td>Person-centred care</td>
<td>Coding system, 6 categories</td>
<td>Unclear</td>
<td>X X X X X X</td>
<td>Primary care, USA, Physicians-patients</td>
<td>Not reported (NR)</td>
<td>NR</td>
</tr>
<tr>
<td>Henbest and Stewart instrument (Henbest and Stewart, 1989)</td>
<td>Person-centredness</td>
<td>Coding system identifying patients' offers defined as:</td>
<td>Yes</td>
<td>X X X</td>
<td>Primary care: UK, Physicians-patients</td>
<td>Inter-rater reliability: Spearman correlation=0.91 Intra-rater reliability: Spearman correlation=0.88 (after 2 weeks) and 0.63 (after 6 weeks). N interactions: 18 (inter-rater); 8 (intra-rater, 2 weeks); 12 (intra-rater, 12 weeks) N raters: 2</td>
<td>Criterion validity: Correlation with other measure (Patient-Centered Clinical Method) rs=0.51 and 0.89 (2 raters, 12 interactions)</td>
</tr>
</tbody>
</table>

¹ Direct observation tool
³ IOM domains: Respect, Coordination, Information, Physical comfort, Emotional support, Involvement of family

(Continued on next page)
<table>
<thead>
<tr>
<th>Direct observation tool</th>
<th>Competency</th>
<th>Format/ content</th>
<th>Conceptual basis</th>
<th>IOM domains</th>
<th>Development setting</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The patient-centred behaviour coding instrument (PBCI) (Zandbelt et al., 2005)</td>
<td>Person-centredness</td>
<td>Coding system</td>
<td>Facilitating behaviours Inhibiting behaviours</td>
<td>Yes</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>The Sherbrooke Observation Scale of Patient-Centered Care (SOS-PCC) (Paul-Savoie et al., 2015)</td>
<td>Person-centred care</td>
<td>Coding system</td>
<td></td>
<td>Unclear</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

(Continued on next page)
Table 1 (Continued).

<table>
<thead>
<tr>
<th>Direct observation tool</th>
<th>Competency</th>
<th>Format/ content</th>
<th>Conceptual basis</th>
<th>IOM domains</th>
<th>Development setting</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Respect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Coordination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Physical comfort</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Emotional support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Involvement of family</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informed Decision Making instrument (IDM) (Braddock et al., 1997)</td>
<td>Shared decision making</td>
<td>Coding system</td>
<td>Unclear</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Primary care, USA Physicians-patients</td>
</tr>
<tr>
<td>Detail of Essential Elements and Participants in Shared Decision Making (DEEP-SDM) (Clayman et al., 2012)</td>
<td>Shared Decision Making</td>
<td>Coding system, 10 categories</td>
<td>Yes</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Oncology consultations, USA Physicians-patients</td>
</tr>
</tbody>
</table>

(Continued on next page)
Table 1 (Continued).

<table>
<thead>
<tr>
<th>Direct observation tool</th>
<th>Competency</th>
<th>Format/ content</th>
<th>Conceptual basis</th>
<th>IOM domains</th>
<th>Development setting</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The OPTION (observing patient involvement) (Elwyn et al., 2003)</td>
<td>Shared decision making</td>
<td>Yes</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Primary care, United Kingdom Physicians-patients</td>
<td>Inter-rater reliability: ICC=0.82; Cohen’s kappa=0.71; Generalisability coefficient=0.68 Intra-rater reliability: Generalisability coefficient=0.66. Internal consistency reliability: Cronbach’s alpha= 0.79 N interactions: 186 N raters: 2</td>
</tr>
</tbody>
</table>

- Rating scale, 12 items
- The clinician:
- identifies a problem(s)
- states that there is more than one way...
- lists “options” ...
- explains the pros and cons...
- checks the patient’s preferred information...
- explores the patient’s expectations...
- explores the patient’s concerns...
- the patient has understood...
- provides opportunities...
- preferred level...
- An opportunity ...
- Arrangements are made...

(Continued on next page)
<table>
<thead>
<tr>
<th>Direct observation tool</th>
<th>Competency</th>
<th>Format/ content</th>
<th>Conceptual basis</th>
<th>IOM domains</th>
<th>Development setting</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Rochester Participatory Decision-Making Scale (RPAD) (Shields et al., 2005)</td>
<td>Shared decision making</td>
<td>Rating scale, 9 items</td>
<td>Unclear</td>
<td>X</td>
<td>Primary care, USA Physicians-patients</td>
<td>Inter-rater reliability: ICC=0.72 N interactions: 193 N raters: NR</td>
<td>Concurrent validity: correlation with other measure (MPCC, dimension finding common ground) $r = 0.19$. Correlation with standardized patient perceptions ($r = 0.32-0.36$) and patient survey measures ($r = 0.06-0.07$).</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>Direct observation tool</th>
<th>Competency</th>
<th>Format/ content</th>
<th>Conceptual basis</th>
<th>IOM domains</th>
<th>Development setting</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modified version of The Measure of Patient-Centered Communication (MPCC) (Dong et al., 2014)</strong></td>
<td>Person-Centred Communication</td>
<td>Coding system</td>
<td>Unclear</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Radiotherapy context, Australia Physicians-patients</td>
</tr>
<tr>
<td><strong>The patient-Centered Observation Form (PCOF) (Chesser et al., 2013; Schirmer et al., 2005)</strong></td>
<td>Person-centred communication</td>
<td>Checklist (form) 13 categories</td>
<td>Yes</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Family medicine residency centre, USA Physicians-patients</td>
</tr>
</tbody>
</table>

(Continued on next page)
Table 1 (Continued).

<table>
<thead>
<tr>
<th>Direct observation tool</th>
<th>Competency</th>
<th>Format/content</th>
<th>Conceptual basis</th>
<th>IOM domains</th>
<th>Development setting</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Respect</td>
<td>Coordination</td>
<td>Information</td>
<td>Physical comfort</td>
<td>Emotional support</td>
</tr>
<tr>
<td>The Process of Interactional Sensitivity Coding in Healthcare (PISCH) (Sabee et al., 2020)</td>
<td>Person-centred communication</td>
<td>Coding system, 7 categories:</td>
<td>Yes</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exchanging information</td>
<td>Fostering relationships</td>
<td>Managing uncertainty</td>
<td>Meta-communication</td>
<td>Recognizing and responding to emotions</td>
</tr>
<tr>
<td>Modified version of The Roter Interaction Analysis System (RIAS), ARCS (Mjaaland and Finset, 2009)</td>
<td>Person-centred communication</td>
<td>Coding system</td>
<td>Unclear</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Social talk</td>
<td>Biomedical questions</td>
<td>Biomedical information</td>
<td>Questions about lifestyle and psychosocial issues</td>
<td>Information about lifestyle and psychosocial issues</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Inter-rater reliability: Cohen’s kappa = 0.46–0.72; Scott’s pi = 0.44–0.72</td>
<td>Type 2 diabetes consultations, USA Physicians-patients</td>
<td></td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N interactions: 50</td>
<td>N raters: NR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Concurrent validity: correlation with other measure (RIAS). No misclassification between RIAS codes and ARCS codes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>Direct observation tool ¹</th>
<th>Competency</th>
<th>Format/ content</th>
<th>Conceptual basis</th>
<th>IOM domains ³</th>
<th>Development setting</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Habits Coding Scheme (4HCS) (Krupat et al., 2006; Frankel and Stein, 2001)</td>
<td>Person-centred communication</td>
<td>Coding system</td>
<td>Unclear</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Primary care, USA Physicians-patients</td>
</tr>
<tr>
<td>The Nonverbal Accommodation Analysis System (NAAS) (D’Agostino and Bylund, 2011 October; D’Agostino and Bylund, 2014)</td>
<td>Nonverbal person-centred communication</td>
<td>Coding system, 10 codes</td>
<td>Yes</td>
<td>X</td>
<td></td>
<td>Oncology consultations, USA Physicians-patients</td>
<td>Inter-rater reliability (Pearson correlation): paraverbal = 0.81–0.96 nonverbal=0.85–0.93 Intra-rater reliability (Pearson correlation): paraverbal = 0.82–1.0; non-verbal= 0.89–0.94</td>
</tr>
<tr>
<td>Adapted version of the Burgoon and Hale Relational communication scale for observational measurement (RCS-O) (Gallagher et al., 2001)</td>
<td>Nonverbal person-centred communication</td>
<td>Rating scale, 34 items, 6 categories:</td>
<td>Unclear</td>
<td>X</td>
<td></td>
<td>Primary care, USA Physicians-patients</td>
<td>Inter-rater-reliability (Cronbach alpha): Immediacy/affectation=0.62; Similarity/depth=0.51; Receptivity/trust=0.72; Composure=0.69; Formality=0.02; Dominance=0.34 Internal consistency (Cronbach alpha): Immediacy/affectation=0.95; Similarity/depth=0.84; Receptivity/trust=0.94; Composure=0.98; Formality=0.92; Dominance=0.60 Inter-rater-agreement (within group agreement coefficient): Immediacy/affectation=0.65; Similarity/depth=0.72; Receptivity/trust=0.86; Composure=0.74; Formality=0.58 Dominance=0.78</td>
</tr>
</tbody>
</table>

¹ Refers to several different frameworks.
and specialists in general medicine, internal medicine, rheumatology and gastro-enterology. Four raters (social scientists) coded the sessions. High inter-rater reliability was noted for the Facilitating behaviour dimension (relative agreement, ICC=0.93 and absolute agreement ICC=0.92) while it was moderate for the Inhibiting behaviour dimension (ICC=0.53 and ICC=0.53, respectively) (D’Agostino and Bylund, 2014). Internal consistency was assessed by Cronbach’s alpha (Facilitating behaviour dimension =0.64 and inhibiting behaviour dimension =0.50). Convergent validity was tested against the Eurocommunication scale where all correlations were in the expected directions (positive for the facilitating dimension $r=0.28$ and negative for the inhibiting dimension $r=-0.29$) (Zandbelt et al., 2005).

SOS-PCC is a 9-item coding system assessing global person-centred care in four dimensions (i.e. biological aspects, establish a therapeutic relationship and provides a treatment plan in col-
<table>
<thead>
<tr>
<th>Reliability</th>
<th>Internal consistency</th>
<th>Test-retest reliability</th>
<th>Inter-rater reliability</th>
<th>Concurrent content</th>
<th>Known groups</th>
<th>Face validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEEP-SDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PISCH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POCF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEEP-SDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEEP-SDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEEP-SDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEEP-SDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEEP-SDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEEP-SDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.3.2. Shared decision-making

IDM is a coding system assessing shared decision making and informed consent. It has six key elements (e.g., discussion of risks and benefits) and scores are determined for each consultation decision. Each decision is rated on each element from 0 to 2 and then aggregated over the six elements to a total score. Inter-rater reliability was assessed in a sample of 20 audiotaped encounters rated by three coders (one physician and two graduate students in medical ethics) and the complete agreement percentage on a given decision was 77%. The developers have reported that although preliminarily analyses indicate good validity, it needs further examination (Braddock et al., 1997).

DEEP-SDM is a coding system with ten coding frames (e.g., definition of option, patient preferences and values, and plan for follow-up) for assessing essential elements of shared decision making. Decision making is coded from 1 (doctor led) to 9 (patient led) for each decision. Coding was conducted by two research assistants using a sample of 20 video-recordings of 20 women at visits with their medical oncologist. Validity and reliability were not presented (Clayman et al., 2012).

OPTION is a 12-item checklist assessing how well clinicians involve patients in decision making. An example item is “The clinician identifies a problem(s) needing a decision making process”. Items are rated against a 5-point scale (strongly agree-disagree). Inter-rater reliability was acceptable in a random sample of 21 consultations rated by two non-clinical raters (inter-rater generalisability coefficient=0.68; Cohen’s kappa=0.71; intra-class coefficient=0.62), as was intra-rater reliability (0.66) (D’Agostino and Bylund, 2011 October). Content validity was based on literature reviews, qualitative studies and consultations with patients and clinicians. Construct validity was supported by showing that OPTION was sensitive to differences in clinician age, sex, qualification and clinical topic (Elwyn et al., 2003).

RPAD is a 9-item rating scale assessing patient-physician collaborative decision making. Example items include “Physician’s medical language matches patient’s level of understanding” and “Physician gives patient opportunity to ask questions and checks patient understanding of the treatment plan”. Items are rated against a 3-point Likert scale and item ratings are summed to a total score. Construct validity was tested against another measure (MPCC) resulting in a modest correlation \( r = 0.19 \). RPAD correlated with standardized patient’s perceptions of the physician-patient relationship \( r = 0.32 - 0.36 \) but less with the patient survey measures \( r = 0.06 \) to \( 0.07 \). Inter-rater reliability was shown to be good (ICC=0.72) in comparisons of ratings of 193 recordings from physician-patient encounters (Shields et al., 2005).

4.3.3. Person-centred communication

The modified version of MPCC is a coding system assessing person-centred communication. The tool comprises components and subcomponents which categorises physicians’ provision of information and responses to patients’ verbal ‘offers’ regarding symptoms, ideas, expectations, feelings and side effects of treatment and effect on function. The first subcomponent consists of
six items which are rated on a 6-point scale for depth of discussion and the second subcomponent consists of nine items. Reliability and validity were examined in 56 recorded radiotherapy consultations. Ten percent of the consultations were dual-coded and inter-coder reliability was high (Krippendorff’s α=0.86). Internal consistency reliability was also good (Cronbach’s alpha 0.48). The MPCC correlated weakly with patient-perceived person-centeredness (PPPC) (r = 0.01) (Dong et al., 2014).

PCOF is a 13-item checklist assessing person-centred communication in 13 different categories with three different options reflecting person-centred behaviours. For example the category Gathering Information has 3 options: uses open-ended questions, uses reflection statement, uses summary/clarifying statement). Scores represent the total number of options endorsed. Four raters (two physicians and two PhD social science researchers) observed 13 encounters in a family medicine residency centre resulting in an overall inter-rater reliability of 0.67 (Cronbach’s alpha). Reliability estimates also differed by discipline. Validity was not reported (Chesser et al., 2013; Schirmer et al., 2005).

PISCH is a coding system assessing interactional sensitivity in person-centred communication. It comprises 7 interaction units including multiple codes. Scores are expressed as a percentage of the units. Inter-coder reliability was assessed using a sample of 50 transcripts of conversations between physicians and patients with type 2 diabetes. Reliability for the PISCH units was estimated to be 0.46–0.72 (Cohens kappa) and 0.44–0.72 (Scott’s pi). Four of the categories (exchanging information, fostering relationships, managing uncertainty and meta-communication) reached acceptable inter-coder reliability. A panel of experts (health communication and/or discourse analysis) reviewed the coding scheme to test the face validity of the instrument (Sabee et al., 2020).

A modified version of RIAS, ARCS is a coding system assessing person-centred communication. The RIAS includes 10 categories (e.g. social talk, biomedical questions and empathy) and the ARCS supplements these with four additional person-centred categories: Attribution, Resources, Coping and Solution-focused techniques (ARCS). Verbal ‘utterances’ by both the doctor and patient are coded into each category. Scores represent the frequencies of utterances in the different 10 categories. The inter-rater reliability was fair (Cohens kappa = 0.52) between the five coders (researchers experienced in the communication skills training) using 145 videotaped consultations with 24 general practitioners (Arora, 2003). Construct validity was assessed by a correlation with the RIAS measure where no classifications across the general RIAS and ARCS codes were identified (Mjaaland and Finset, 2009).

4HCS is a coding system assessing clinicians’ communication skills. The 4HCS identifies 23 behaviours, each associated with one of the four habits: Invest in the beginning, (six items); Patient’s perspective (three items); Demonstrate empathy (four items) and Invest in the end (10 items). Items are rated on a 5-point scale. Construct validity was tested against other available measures. Inter-coder reliability was assessed in13 videotaped visits in primary care that were rated independently by two coders (health professions students). Reliability estimates were satisfactory (0.69–0.80 for the four habits (overall 0.72)). Internal consistency reliability was assessed using Cronbach’s alpha was satisfactory for two habits (Habit 1 = 0.71 and Habit 3 = 0.81) and unsatisfactory for the other two (Habit 2 = 0.51 and Habit 4 = 0.61) Evidence for construct validity was provided by correlations between 4HCS ratings, RIAS, back channel responses, and non-verbal measures (Krupat et al., 2006; Frankel and Stein, 2001).

4.3.4. Nonverbal person-centred communication

NAAS is a coding system assessing nonverbal communication. The tool consists of two codes (paraverbal and non-verbal) which include ten behaviour categories (e.g., talk-time, smiling, eye contact). NAAS categories are coded in one-minute segments. A baseline score is first calculated as (Institute of Medicine, 2001) an average of the physician’s NAAS behaviours during the first two segments in the consultation; and (World Health Organization, 2006) an average of the patient’s NAAS behaviours during the corresponding segments. Subsequent segments are compared with initial segments to reveal if the physician and patient changed within each NAAS behaviour and, if so, in what direction. Inter- and intra-rater reliability was assessed in ten randomly selected oncology consultations coded by two raters. Inter-rater reliability for the ten NAAS categories in the paraverbal and non-verbal codes was satisfactory (Pearson correlation=0.81–0.96 and 0.85–0.93, respectively). Intra-rater reliability was 0.82–1.0 (paraverbal) and 0.89–0.94 (non-verbal). Construct validity was assessed with the MIPS measure showing correlations with physician eye contact (r = 0.447) and patient eye contact (r = 0.623) (D’Agostino and Bylund, 2011 October; D’Agostino and Bylund, 2014).

This adapted version of the RCS-O is a 34-item rating scale assessing six dimensions of nonverbal communication. Items are rated on a 7-point Likert scale ranging from “strongly disagree” to “strongly agree” and item ratings are summed to dimension scores. Inter-rater reliability was assessed using a random sample of 20 videotaped interactions of medical students interviewing patients. Interactions were rated by three trained observers (social and behavioural scientists) at two time points separated by eight weeks. At time one, four of the six dimensions (immediacy/affectation, similarity/depth, receptivity/trust and composure) indicated fair to excellent internal consistency (0.84–0.98), inter-rater reliability (0.51–0.72), inter-rater agreement (0.65–0.86) and construct validity (correlation with the interview rating scale (IRS)) (0.50–0.76). The estimates were similar at time two (eight weeks later), except that the inter-rater reliability for receptivity/trust then was low (0.19). Internal consistency was also strong for the dimension validity (0.91–0.92) but moderate for the dominance dimension (0.60–0.66). Inter-rater reliability was low for both formality (–0.11–0.02) and dominance (0.34–0.50) and even if inter-rater agreement was good (0.58–0.86) these two dimensions were suggested to be dropped or revised. Formality and dominance both correlated negatively with the IRS (−0.26 to −0.31) (Gallagher et al., 2001).

5. Discussion and conclusion

This review identified a manifold of direct-observation methods for use in assessing PCC, varying in assessed main constructs; assessment formats; conceptual frameworks underpinning their development; coverage of core PCC dimensions; and psychometric evaluations and performance.

Given the multitude and variation in existing PCC frameworks and concepts, it is important that developers of assessment tools carefully articulate the conceptual underpinnings of their tool (Hallgren, 2012). Although clear descriptions were provided for some methods, in 9 of the 16 tools we were nevertheless unable to ascertain with any certainty what conceptual framework(s) was applied or developed to guide their construction. Furthermore, content analyses were seldom performed to affirm if the content of the methods adequately represents the construct(s) intended to be assessed. Similar shortcomings have previously been reported in quantitative measures for assessing patient-centred communication (Epstein et al., 2005). In the absence of information about the assumptions underlying a tool and theoretical and/or empirical support for those assumptions, its validity and utility as a measure for assessing PCC may be questioned.

Given that PCC assessment tools often embody different concepts, apply different nomenclature to designate often similar concepts (Arora, 2003; Mead and Bower, 2000 Jan) or seemingly sim-
ilar concepts that are operationalized differently (Epstein et al., 2005), we examined the content of each tool in relation to a standard framework in order to identify conceptual commonalities and differences between the tools. For this purpose we chose the six IOM-endorsed dimensions covering aspects of PCC (Institute of Medicine 2001). We found that only two tools, the DOC (Bertakis and Azari, 2011) and PCOF (Chesser et al., 2013), appeared to cover all six dimensions, whereas coverage by the remaining tools varied from as few as one to a maximum of four dimensions. Interestingly the DOC was the only tool of the five purporting to measure the holistic concept PCC or patient-centeredness to tap all IOM dimensions. The other four holistic tools were judged to include between three and four dimensions, indicating both notable conceptual differences between methods and significant gaps in assessment domains. This is in line with findings by Mead & Bower (Mead and Bower, 2000 Jan) showing that PCC assessed by means of three different direct-observation tools (an adapted RIAS, Eurocommunication and Henbest & Stewart instrument) (Mead and Bower, 2000 Jan) correlated poorly, which the authors suggested indicate that they tap distinct aspects of doctor-patient interactions.

Similarly, dimension coverage varied in the four tools (each tool included between two and three dimensions) designed specifically to assess shared decision making. For example, IDM (Braddock et al., 1997) was judged to tap Involvement of family and friends, whereas OPTION (Elwyn et al., 2003) instead covered Emotional support. This difference may in part account for weak correlations reported between these measures (Weiss and Peters, 2008). On a similar note, the five tools assessing patient-centred verbal communication also appeared to differ somewhat in dimension coverage, although commonalities in coverage were apparent. PCOF appeared to tap all six IOM dimensions and the others between three and four dimensions. With content suggesting three common underly ing concepts, substantial agreement between these communication assessment tools may be expected; however, to our knowledge the only head-on-head evaluation of these methods has been made between the MPCC and the 4HCS, where weak correlations were found (Clayton et al., 2011). The two identified tools for assessing nonverbal communication (NAAS (D’Agostino and Bylund, 2014) and RSC–O (Gallagher et al., 2001) were unsurprisingly, given their focus on behavioural aspects doctor-patient interactions, both judged to tap a single IOM dimension, namely Respect for patients’ values, preferences, and expressed needs.

Few of the tools appeared to cover the IOM dimensions Physical comfort (n = 4) or Involvement of family and friends (n = 3). It is noteworthy that these two dimensions have also been shown to be the least frequently covered IOM domains in patient-reported PCC measures (Tzelepis et al., 2015). This suggests that their importance as independent core elements of PCC may either be overstated in the IOM framework or undervalued in assessment tools. For example, providing patients with physical comfort may be seen to lie within the realm of standard professional ethics and good medical practice rather than as a component distinctly characterizing PCC. On the other hand, involving family and friends in patient care may serve to further the aims of PCC in that they may facilitate patient-clinician communication, provide emotional support to the patient during consultations and at home and help to negotiate medical decision making (Rosland et al., 2011). Despite its relatively meagre representation in the methods included in this review, involvement of family and friends has also been identified as one of the six most frequently recurring domains found in nine prominent PCC frameworks (Cronin, 2004).

The comprehensiveness of reported psychometric evaluations varied substantially between tools. No evaluations were reported for two tools (DEEP-SDM (Clayman et al., 2012) and DOC (Bertakis and Azari, 2011), whereas relatively rigorous reliability and validity evaluations were performed for the OPTION, MPCC and NAAS. Reliability was evaluated in terms of inter-observer reliability (IOR) for 12 of the 16 tools, all of which reported IOR estimates meeting conventional cutoffs for moderate to high agreement between/among raters. Cases of moderate IOR associated with the modified RIAS, PISC and RSC–O were attributed to shortfalls in the tool related to ambiguities in concept definitions or inadequacies in defining target behaviours, or inclusion of behaviours that are not readily observable (Sabee et al., 2020; Mjaaland and Finset, 2009; Gallagher et al., 2001). Although IOR is often understood to be a characteristic inherent to an assessment tool, it may be influenced by a number of methodological factors unrelated to the observation tool per se, such as sample homogeneity, or univariate analysis (Hallgren, 2012), extent of rater training (Mead and Bower, 2000 Jan), raters’ professional background (Chesser et al., 2013) and other rater characteristics/ idiosyncrasies (Kogan et al., 2017), etc. For example, although checklists are assumed to improve IOR because they provide relatively unambiguous behavioral definitions (Regehr et al., 1998) we noted that some of the coding systems, e.g., Henbest & Stewart and SOS–PCC actually outperformed the two checklists PCOF and COT. It is conceivable that in these cases extensive rater training in using the coding schemes may have offset the effect of greater clarity in target behaviour definitions.

It is noteworthy that other important types of reliability were less frequently reported for the various tools. For example, test-retest reliability, reflecting the stability of ratings over time, was reported for only 3 of the 16 tools, and homogeneity, as measured with e.g. Cronbach’s alpha, was reported for 7 of them. In the absence of comprehensive evaluations of reliability it is not possible to fully appraise the precision to be expected of codings. Also noteworthy is that power calculations for estimating samples sizes used in testing reliability (and validity) were reported for only one tool, the SOC–PCC, undermining confidence in the estimates.

Validity assessments were reported for all tools except DOC, IDM, DEEP-SDM and PCOF (Bertakis and Azari, 2011; Braddock et al., 1997; Clayton et al., 2012; Chesser et al., 2013). Concurrent validity was most commonly evaluated but notably many correlations were very low. Content validity was evaluated by and large by panels of experts; however, the experts’ areas of expertise were often not specified. Paradoxically, we found no mention of patients taking part in the evaluation process. Moreover, it is noteworthy that other measurement characteristics that are critical for appraising and interpreting outcomes, such as responsiveness and sensitivity were conspicuously absent in evaluations of the various methods. Likewise, predictive validity was not assessed for any method.

All of the identified methods have their own particular strengths and weaknesses and decisions about which one to use will invariably involve some amount of tradeoff. Ultimately, the choice of tool will depend on the purpose for which it is to be used. For formative use to guide learning by providing benchmarks to orient and give feedback to reinforce learners, all of the tools may in some way be useful. However, at a minimum evidence supporting content validity of the tool should be available to ensure that the sampled behaviours comprehensively and accurately represent the construct(s) of interest. On the other hand, for high stakes, summative use, tools with clear conceptual base and demonstrated strong validity and reliability are needed. We cannot recommend any of the tools without some reservations; however, of the tools assessing holistic PCC the Henbest & Stewart instrument (Henbest and Stewart, 1989) appears promising given its well-defined conceptual core and relatively good psychometric performance. It has not, however, been evaluated for content validity and only three IOM dimensions were covered. Of the tools for assessing shared decision making, OPTION (Elwyn et al., 2003) is one of the two with a defined conceptual framework and stands out as
5.1. Limitations

There are limitations to this review. The eligibility of tools for inclusion and their coverage of the IOM domains were subjectively judged but were thoroughly discussed by co-authors until agreement was reached. Our main search was limited principally to articles identified through the search string, subsequent evaluations of the included tools were not sought and no attempt was made to contact developers for additional information about their tools. On the other hand, after a tool was identified in the search, we searched the literature for earlier articles describing the development of the tools.

6. Conclusions

Differences in the scope, coverage and content of the identified direct observation tools likely reflect the complexity of PCC and lack of consensus in defining this concept. Greater clarity and transparency are needed in describing conceptualizations and assumptions underpinning of the tools to enable potential users to make informed decisions when selecting tools. Although all of the tools may serve formative purposes, evidence supporting their use in summative evaluations is scarce. Patients were not involved in the development of any of the assessment tools, which seems intrinsically paradoxical given the aims of PCC. All of the tools may be useful for providing trainee feedback; however, rigorously tested and patient-derived tools are needed for high-stakes use.

Funding sources

This work was supported by the centre for Person-centred Care at the University of Gothenburg (GPCC), Sweden. GPCC is funded by the Swedish Government’s grant for Strategic Research Areas, Care Sciences (Application to Swedish Research Council no. 2009-1088) and co-funded by the University of Gothenburg, Sweden.

Declaration of Competing Interest

None.

References

Rosland, A.-N., Piette, J.D., Choi, H.J., Heisler, M., 2011. Family and friend par-
ticipation in primary care visits of patients with diabetes or heart failure: patient and physician determinants and experiences. Med. Care 49 (1), 37–45.


