

Observed Interprofessional Collaboration (OIPC) During Interdisciplinary Team Meetings: Development and Validation of a Tool in a Rehabilitation Setting

**Emmanuelle Careau, OT, PhD; Claude Vincent, OT, PhD;
& Bonnie R. Swaine, PT, PhD**

Abstract

Background: Despite all the efforts made in the past few years, interprofessional collaboration (IPC) in clinical settings is not always optimal. In addition, there are only a few instruments that healthcare managers and practitioners can use to evaluate the quality of IPC practice. Therefore, we developed an observation-based tool to evaluate IPC interactional factors occurring during interdisciplinary team meetings, and we examined the initial validation of the tool in a rehabilitation setting.

Methods and Findings: The items were developed and pre-tested iteratively by construct experts ($N = 7$) and non-experts ($N = 4$). Interrater reliability was determined between two observers, following the analysis of 30 video recordings of meetings in two rehabilitation centres involving a total of 152 participants. An observation grid (OIPC) consisting of 20 items that can be answered on a three-point scale and demonstrating acceptable interrater reliability was developed.

Conclusions: The OIPC is a tool aimed at evaluating IPC interactional factors during interdisciplinary meetings based on team performance rather than individual behaviours. It can be useful for healthcare managers and practitioners who want to evaluate the quality of IPC practices.

Keywords: Interprofessional collaboration; Measurement; Observation-based tool; Interdisciplinary team meetings

Introduction

Interprofessional collaboration (IPC) is considered an efficient, effective, and satisfying way to provide healthcare and social services [1]. In various countries, policy reviews have highlighted the importance of collaboration and promoted changes in healthcare to enhance IPC [2-5]. Although IPC practice is sometimes problematic in clinical settings [3,6,7], little attention has been paid to evaluating the quality of these practices.

In the healthcare field, quality has been defined as “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge” [8, p. 232]. Donabedian maintained that quality could be formally evaluated by the concordance between the process and the standards for different aspects of the process [9]. “Standards” refer to what constitutes “goodness” for the phenomenon being studied: normative standards can be developed according to experts’ recommendations, and empirical standards can be derived from patterns of care observed in practice [10]. One of the limitations in

evaluating IPC quality is the lack of knowledge concerning what IPC processes are exactly. In fact, many researchers seem to adopt a “black box” approach, with more emphasis on IPC determinants and results and ignoring the processes. This makes it difficult to formulate IPC quality standards and develop accurate measurement tools to evaluate collaborative practices in order to improve deficient ones and encourage exemplary ones. To address this situation, we conducted two previous studies to develop a framework illustrating IPC processes, which we called interactional factors [11], and to describe how the factors could be operationalized in the context of interdisciplinary team meetings [12]. We chose to focus on interdisciplinary team meetings because they are an IPC modality linked to “best practices” in many contexts [13-15]. The present study aimed 1) to develop an observation-based instrument to evaluate IPC interactional factors occurring in interdisciplinary team meetings, and 2) to examine the initial validation of the tool in a rehabilitation setting.

As depicted in the IPC framework we developed [11], IPC interactional factors emerge from a particular situation: Practitioners have to adjust the type and intensity of their collaborative practice to the biopsychosocial needs of clients and family members. In general, the more complex the individual’s needs are, the more intense the collaboration should be. Thus, practitioners need to create a partnership to address these needs, and underlying these partnerships is an *intention* that motivates them to collaborate with each other. According to this perspective, interdisciplinary team meetings refer specifically to what we have identified as the “shared healthcare practice.” In this type of collaborative practice, the specific intention of all partners should be to share decisions and actions regarding common objectives.

There does not appear to be a tool that can be used to specifically evaluate IPC interactional factors related to the shared healthcare practice. The majority of IPC measurement tools aim to evaluate attitudes and beliefs about IPC [19-23]. Based on a review of 35 articles reporting the development or use of surveys measuring teamwork processes in healthcare, Valentine et al. found that the most commonly assessed teamwork process dimensions are “communication,” “co-ordination,” “use of all members’ expertise,” and “respect” [24]. These instruments were often developed to assess self-perception about team processes and/or performance [25-28]. Although self-report and subjective data offer valuable insight into respondents’ perceptions regarding IPC, these tools do not provide an objective description of the actual collaboration among professionals [29].

The use of observation-based instruments is generally recognized as an appropriate way to evaluate the quality of collaborative interactional factors in complex and dynamic systems [30]. However, only a few observation-based instruments aimed at evaluating IPC interactional factors in health and social care can be found in the literature. One of these, the Interprofessional Collaborator Assessment Rubric (ICAR) from Curran et al., aims to evaluate individual IPC competencies according to six dimensions: communication, collaboration, role and responsibility, client-centred practice, team functioning, and conflict resolution [31]. The observer decides what level of competence each team member has demonstrated for a set of behavioural indicators/performance statements on a four-point scale

(minimal, developing, competent, mastery). Initially, the ICAR was developed and validated in an academic setting to assess students' IPC competences, making it difficult to use in a clinical context. Moreover, this tool is based on the observation of individual behaviours, so it cannot be used to assess the performance of a *team* of individuals.

Another observation-based instrument, the Observational Teamwork Assessment for Surgery from Healey et al., allows real-time evaluation of collaboration occurring in a surgical team [32]. The authors developed a list of exemplary behaviours related to five dimensions: communication, co-operation, co-ordination and back-up behaviour, leadership, and team monitoring/situation awareness. Observers score each of these dimensions on a seven-point scale (0 = problematic behaviour/team function severely hindered to 6 = exemplary behaviour/very highly effective in enhancing team function). Finally, the Mayo High Performance Teamwork Scale developed by Malec et al. is a behavioural checklist focusing on the frequency of key teamwork behaviours occurring in medical crisis resource management [33]. Observers score the frequency of these behaviours on a three-point scale (0 = never or rarely, 1 = inconsistently, 2 = consistently). To develop the items, the authors relied on four behavioural dimensions: co-operation/communication, leadership/management, situation awareness, and decision-making. Although these two instruments have good psychometric and pragmatic properties, they are specific to a particular context (surgical team or medical crisis resource management), making them difficult or inappropriate to use to evaluate IPC during interdisciplinary team meetings. For this reason, we believed it was important to develop an observation-based instrument to measure the quality of IPC interactional factors during interdisciplinary team meetings.

Method

This research was approved by the ethics committee of the *Institut de réadaptation en déficience physique de Québec* (project #2008-145). The tool was developed in several phases following a standardized method [33,34].

Phase 1: Identification of dimensions

In a previous study, the construct of IPC interactional factors that occurred during interdisciplinary meetings was defined as

situated at the extremity of a continuum of collaborative practice involving multiple stakeholders (practitioners, clients, family members, community partners, etc.) and characterized by interdependence between these individuals to develop a cohesive care plan that meets clients' needs. ... During each interdisciplinary meeting, two major temporal phases can be observed. First, those involved exchange information to build a common vision of the situation and then they discuss it to develop a common cohesive action plan. During each phase, behaviors that positively or negatively influence IPC can be observed. Some behaviors are linked to fundamental aspects of IPC (client-cen-

tred practice, sharing a common vision, power relationship based on expertise, decision-making based on real consensus and adoption of a common action plan), while other elements support and facilitate IPC (mediation/arbitration/regulation, problem-solving, open and respectful environment, size and composition of the team, vocabulary used, frame of reference used). [12, p. 300]

In addition to this definition, literature on IPC was searched to find articles and documents to help authors identify the construct dimensions. This search was conducted using Medline via Pubmed and CINAHL between 2008 and 2012 and the following keywords: interprofessional*, interdisciplin*, multidisciplin*, pluridisciplin*, transdisciplin* in combination with collaboration, team*, interaction, practice and definition, concept analysis, clarification, conceptual framework, and model. A manual search using reference lists was also performed. Only documents presenting an original definition based on a concept analysis, systematic literature review, or expert consultation were included. Documents from domains other than health and social care were excluded. Around 50 scientific papers, reports, and books were consulted. The three developers (E.C., C.V. & B.S.) then met to decide consensually upon the ten dimensions to be included in the tool being developed. They were: 1) purpose of the meeting, 2) team composition, 3) expertise affirmation and recognition, 4) attainment of consensus, 5) person-centred practice, 6) communication, 7) respectful attitude, 8) facilitation/mediation, 9) shared decision-making, and 10) adoption of a common action plan.

Phase 2: Generation of items and pre-testing

Based on the previous discussions around the dimensions, the first author (E.C.) generated an initial pool of 22 items that were subsequently validated by the second and third authors. A subsequent version was developed following successive and iterative steps of item generation/modification through consultation with experts and pre-testing with expert and non-expert observers (see Table 1 for more details concerning the expert and non-expert participants). These steps are presented in Figure 1, which illustrates the procedure followed to obtain the final version of the OIPC (version 2.4), as well as the number of participants involved in each step and the modifications made throughout the process.

The consultation involved six IPC construct experts from academic ($N = 3$) and clinical settings ($N = 3$) in Quebec (Canada) to ensure the content validity of the tool to be developed (see Table 1). Experts were selected based on their known interest and expertise in IPC. By email, they were asked to judge the relevance of each item with regard to the construct definition and the dimensions on a four-point scale (from very relevant to very irrelevant). They were also invited to make comments and suggestions to improve them. This procedure was repeated after each item modification step. Although the experts were asked to send back their evaluation within a month, some ($N = 3$) did not always send it back, but the process continued; thus, the number of participating experts varied during the process (see

Figure 1). To test the items' applicability, pre-testing was conducted with two of the authors (E.C. & C.V.), the project's research assistant, another research assistant, and a graduate student, who were not directly involved in this research, and one construct expert, who was also involved in the consultation (see Table 1 for more details about these experts). For each step in the pre-test, these participants observed one real interdisciplinary clinical meeting previously recorded on video and scored what they observed using the OIPC. Then, semi-structured individual interviews with these participants were conducted using the principles of cognitive interviews [36], during which they were asked to explain carefully how they interpreted each item and how they would score each item based on what they observed. This allowed misunderstandings or potential problems with the tool's items to be detected. Afterward, the content validity was explored by comparing OIPC's dimensions with those identified in a review of 24 questionnaires aiming to assess the self-perception of IPC interactional factors occurring in healthcare teams [24].

Table 1
Participants' profile and their involvement in the item generation and pre-testing phases

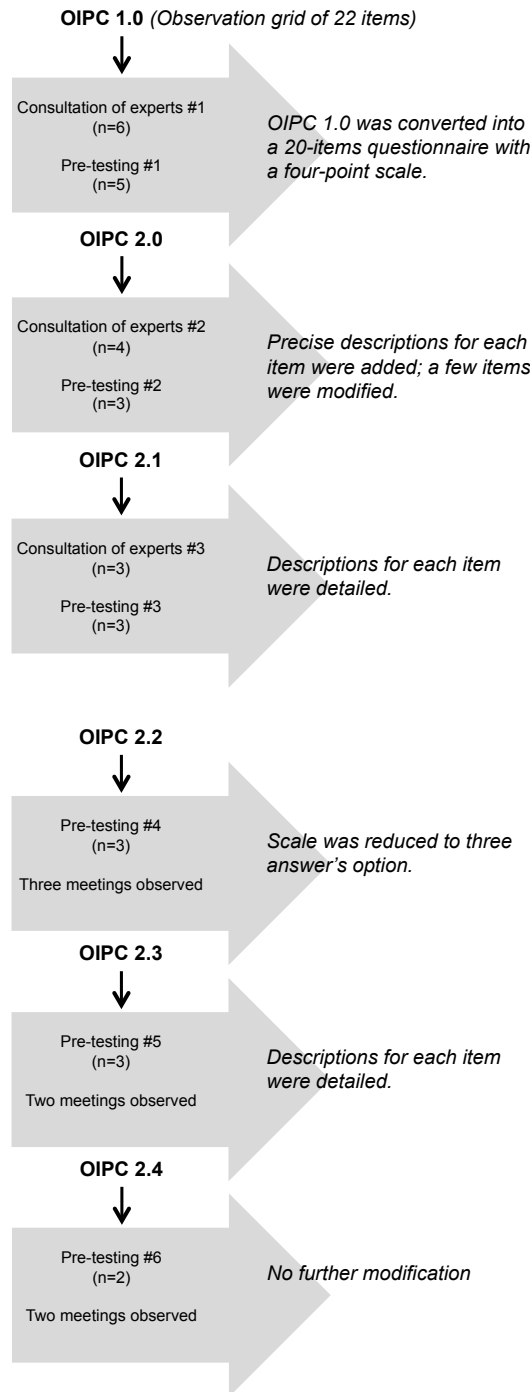
| Participants | Profile | Experts' consultation | | | Pre-testing | | | | | |
|---------------|--|-----------------------|---|---|-------------|---|---|---|---|---|
| | | 1 | 2 | 3 | 1 | 2 | 3 | 4 | 5 | 6 |
| Expert 1* | Occupational therapist, IPC clinical consultant, PhD student | | | | | X | X | X | X | X |
| Expert 2 | Social worker, researcher/professor | X | X | X | | | | | | |
| Expert 3 | Social worker, researcher/professor, director of Université Laval's IPC office | X | X | X | | | | | | |
| Expert 4 | Occupational therapist, rehabilitation healthcare practitioner involved in the development of IPC guidelines | X | | | | | | | | |
| Expert 5 | Social worker, rehabilitation healthcare practitioner involved in the development of IPC guidelines | X | | | | | | | | |
| Expert 6 | Occupational therapist, researcher/professor | X | X | | | | | | | |
| Expert 7 | Psychologist, IPC consultant for a primary healthcare facility | X | X | X | X | | | | | |
| Non-expert 1 | Undergraduate student in occupational therapy; research assistant for the project | | | | X | X | X | X | X | X |
| Non-expert 2* | Occupational therapist, researcher/professor | | | | X | X | X | X | X | |
| Non-expert 3 | Athletic therapist, PhD student in neurorehabilitation | | | | X | | | | | |
| Non-expert 4 | Undergraduate student in occupational therapy, research assistant for another project | | | | X | | | | | |

* Expert 1 was the first author of this article (E.C.) and non-expert 2 was the second author (C.V.).

Phase 3: Examination of interrater reliability

Because the OIPC is an observation-based tool, it was important to document its interrater reliability [37]. In the province of Québec, interdisciplinary meetings are used with almost every rehabilitation client, so we decided to form a convenience

Figure 1
Detailed procedure followed to develop the OIPC



sample of clients in two urban rehabilitation centres (Québec City and Montréal). We tried to record as many interdisciplinary meetings as possible in a six-month period in 2009 with various teams and different patient populations from different clinical programs (head injury, spinal cord injury, musculoskeletal injury, and amputation). To do so, we followed this procedure: After receiving the weekly schedule of all the meetings at the participating centres, the first author (E.C.) or the project's research assistant contacted the clinical co-ordinators to determine which meetings could be recorded. For ethical reasons, it was the clinical co-ordinator from each program who decided which meetings could be recorded, and unfortunately we do not know exactly how this decision was made because the clinical co-ordinator did not have to explain the reason to the researchers. Once identified, a research assistant contacted all of the potential meeting participants to sign the consent form. The client (i.e., rehabilitation service user) had to provide his or her consent even if he or she would not be present at the meeting. If the client and all of the participants provided consent, the research assistant placed a camera in the room, started recording, and left the room a few minutes before the meeting started.

Two raters watched the 30 videotaped meetings and independently

scored the team’s performance using the OIPC version 2.4, thus providing data to examine the tool’s interrater reliability. The first rater was one of the authors (E.C.) and was considered a construct expert. The second rater was an undergraduate student in occupational therapy and was considered a non-expert with regard to the IPC construct. Nevertheless, she participated in the last five pre-tests as an observer, so she was very familiar with the tool. An interrater reliability analysis using Cohen’s Kappa was performed to determine consistency among observers. Cohen’s Kappa is generally best suited for dichotomous/ordinal scales, especially when the discrepancy between response options is not equal [38]. We also generated graphs from the cross-frequency tables for each item. These graphs allowed us to examine the entire score distribution for each item.

Results

The first version of the instrument (OIPC 1.0) resembled a grid that an observer could use to collect quantitative and qualitative data. Figure 2 presents examples of initial items for the dimensions facilitation/regulation, shared decision-making, and adoption of a common plan. Experts felt the initial items were generally relevant but it was difficult to judge the quality of IPC interactional factors based only on the data collected related to these items. Following the first pre-test of version 1.0, we

Figure 2
Examples of items in the OIPC 1.0

#21- Which regulation tasks were used by the moderator?

| Tasks | Used often | Used moderately | Used rarely | Comments |
|--------------|------------|-----------------|-------------|----------|
| Stimulation | | | | |
| Coordination | | | | |
| Facilitation | | | | |

#9- How were the decisions made?

By consensus By the majority Some people imposed their idea(s)

If applicable, which professionals imposed their ideas?

#7- What was the action plan?

| Objectives | Involve only one professional | Involve more than one professional | Involve all the professionals |
|------------|-------------------------------|------------------------------------|-------------------------------|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

also noted that this version was hard to complete, requiring the observer to watch the videotaped meeting twice to adequately score all items. Consequently, the subsequent version of the tool (version 2.0) was a questionnaire with a four-point scale measuring team performance in terms of IPC interactional factors (very adequate to very inadequate). Because we wanted the tool to be simple enough to allow a quick measure of IPC, we decided to generate a pool of one or two items per dimension, except for communication, which comprised six items. Communication was the only dimension considered too complex to assess using only one or two items. Six items were developed to document various aspects of communication (e.g., vocabulary used, clinical content of the discussion, its duration and efficacy) for each of the temporal phases.

After the subsequent steps of pre-testing and expert consultation, the number of items remained the same but the wording had been modified, and precise descriptions of what was expected in terms of a team's performance (quality standards) were provided to assist in item scoring (version 2.1). The third step in the experts' consultation and pre-testing allowed further refinement of the quality standards (version 2.2). Since further pre-testing focused on improving the scoring procedure and no major modifications were made to the wording of the items, we stopped the consultation process after generating version 2.2. After the fourth step of pre-testing, we found that the 4-point scale was too hard for a non-expert to use, so we decided to reduce the scale to three response options (inadequate, more or less adequate, adequate) (version 2.3). Following the last pre-test, the descriptions of several items were again reworded to facilitate the scoring procedure (version 2.4).

The final version of the OIPC (version 2.4) was composed of 20 items measuring ten dimensions related to IPC interactional factors that occurred during interdisciplinary team meetings. The first ten items were related to the meeting's first temporal phase (building a common vision of the situation), while the remainder were related to the second temporal phase (development of a common action plan). An additional section, identified as potentially useful when evaluating the quality of IPC, was included in the tool to allow observers to record general information about the meeting (e.g., number of participants, total duration of meeting, general relevance of meeting, etc.). The OIPC was developed and validated in the Canadian French language. However, for the purpose of this article, the tool was translated into English. To ensure the validity of the items' translation, we conducted a simple back translation (French to English, then English to French). Click here to consult the OIPC 2.4 in its original language (Canadian French) and here to consult the translated version of the tool.¹ However, please note that only the Canadian-French version has been validated and reported here.

Validity

According to DeVellis [34] and Streiner and Norman [35] there are four types of validity: face validity, content validity, construct validity, and criterion-related validity. Face validity "indicates whether, on the face of it, the instrument appears to be assessing the desired qualities" [35, p. 6], whereas content validity consists of a

“judgement whether the instrument samples all the relevant or important content or domains” [35, p. 6]. Construct validity can be defined as “the extent to which a measure ‘behaves’ the way that the construct it purports to measure should behave with regard to established measures of other constructs” [34, p. 53], and criterion-related validity is “the correlation of a scale with some other measure of the [phenomenon] under study, ideally, a ‘gold-standard’ which has been used and accepted in the field” [35, p. 254]. We believe the OIPC has good face validity and content validity since items were generated following formal consultation with IPC experts. With respect to the other types of validity, it is difficult to affirm the validity of IPC instruments, because there is no consensus in the scientific community around the definition of the IPC construct. Because of the limited theory about the construct, and the fact that the OIPC is the only available tool that specifically measures IPC interactional factors during interdisciplinary meetings through external observation, determining the tool’s validity is challenging. However, the OIPC’s content validity can be examined by comparing its measured dimensions with those of other instruments measuring a similar construct. Table 2 highlights the consistency between the OIPC’s measured dimensions and the dimensions identified by Valentine et al., following a review of 24 questionnaires aiming to assess the self-per-

Table 2
Comparison of OIPC measured dimensions with those measured in 24 questionnaires on self-perception of IPC interactional factors as identified by Valentine et al. [24]

| Dimensions of “IPC interactional factors” construct as measured by the OIPC | Dimensions of “Teamwork processes” construct in health and social care as identified by Valentine et al. (2012) | | | | | | | | | | | | | |
|---|---|-----------------------------------|---------------|--------------------------------|--------------------------------|------------------------|----------------------------|-----------------|----------|--------------------------------|----------------|----------------------|-----------------------------------|-------------------|
| | Behavioral processes | | | | | | | Emerging states | | | | | | |
| | Communication* | Coordination (mutual adjustment)* | Collaboration | Use of all members’ expertise* | Help each other/share workload | Shared decision-making | Active conflict management | Effort | Respect* | Group cohesion/shared identity | Social support | Psychological safety | Role responsibility understanding | Shared objectives |
| Purpose of meeting | | | | | | | | | | | | | | + |
| Team composition | | | | | + | | | | | | | | + | |
| Expertise affirmation and recognition | | | | ++ | | | + | | | | | | + | |
| Attainment of consensus | | + | + | + | | | | | | | | | | |
| Person-centered practice | | | | | | | | | | | | | | |
| Communication | ++ | | | | | | | | | | | | | |
| Respectful attitude | | | | | | | + | | ++ | | + | + | | |
| Facilitation/mediation | | + | | | | | ++ | | | | | | | |
| Shared decision-making | | | + | | | ++ | | | | | | | | |
| Adoption of a common action plan | | | + | | ++ | | | | | | | | + | |

* Dimensions identified by authors as the most fundamental and the most often measured

ception of IPC interactional factors occurring in healthcare teams [24]. The symbol “++” indicates the dimensions measured were exactly the same, while a “+” indicates the OIPC’s dimensions were similar but covered only some attributes of the dimensions identified by Valentine et al. [24]. We should note that the OIPC’s dimensions cover almost all of the dimensions identified by Valentine et al. [24]. The dimension “group cohesion/shared identity” is not assessed by the OIPC, and the dimensions “shared objectives,” “role responsibility understanding,” “psychological safety,” “social support,” “effort,” and “help each other/share workload” are covered only partially.

Interrater reliability

Because of the challenge involved in obtaining ethical consent from all of the participants in time before the meeting, we were able to record 30 meetings over six months to examine the interrater reliability. These meetings lasted between 17 and 75 minutes and involved 3 to 14 people per meeting ($M = 8$ people per meeting). Clients were present at 67% of the meetings. Clients had various conditions such as traumatic brain injury, musculoskeletal injury, spinal cord injury, or amputation. Overall, 29 clients, 12 relatives, 106 practitioners, and 5 insurance agents participated in this phase of the study. Figure 3 presents the percentage of meetings attended by each discipline. For example, a clinical co-ordinator attended all of the meetings, while a psychologist was present at 40% of the 30 meetings.

Figure 3
Percentage of meetings attended by each discipline
(N = 106 healthcare workers from two urban rehabilitation centres)

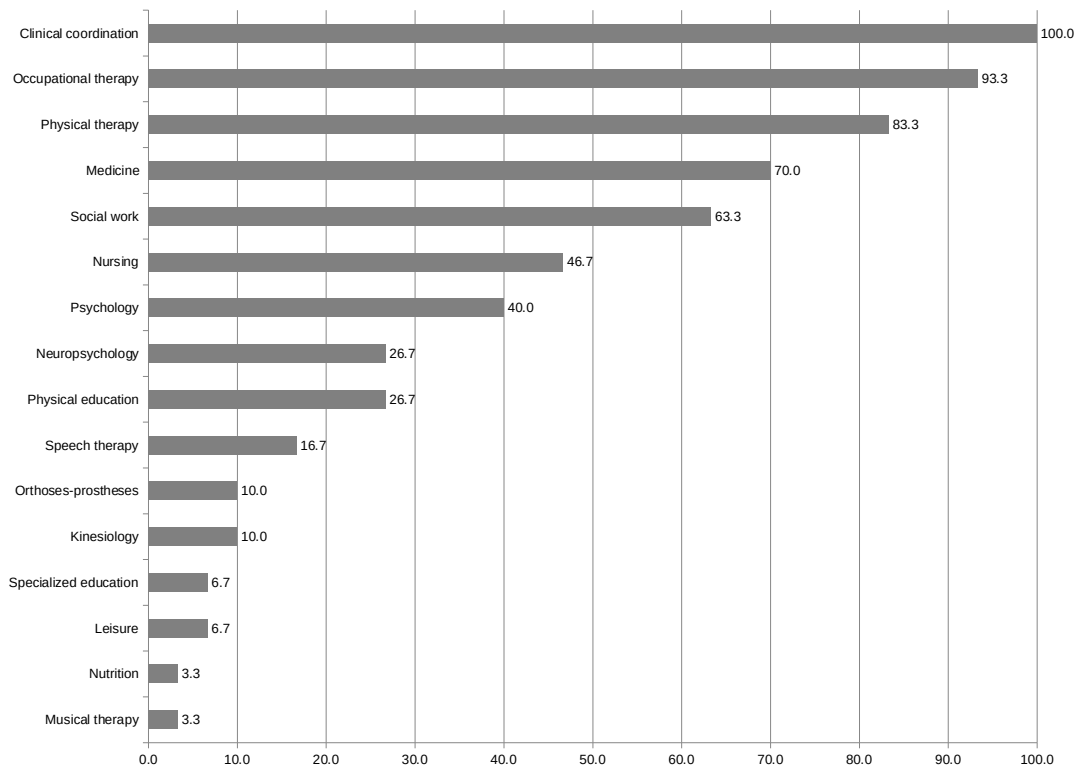


Table 3 presents the items of the OIPC version 2.4 and their associated dimensions for each temporal phase along with the results and interpretation of interrater reliability. The percentage agreement varied from 53.5 to 100%. It also shows that interrater reliability was quite variable, ranging from “good” to “slight” depending on the item (k from .79 to .15). Three items out of 20 could not be interpreted because of a low variance.

Table 3
Items of the OIPC version 2.4, their corresponding dimensions and the interrater reliability associated with its scoring

| Items referring to the first phase "building a common vision of the situation" | Dimension measured | % agreement | κ (p value) | Interpretation* |
|---|-----------------------------------|-------------|-------------|-----------------|
| Q1 - How do you rate the way the meeting's objectives were discussed? | Purpose of meeting | 86.7 | .76 (< .01) | Good |
| Q2 - How do you rate the duration of the discussion on the situation? | Communication | 70.0 | .18 (.08) | Slight |
| Q3 - Based on the issues identified, how do you rate team composition? | Team composition | 86.7 | .73 (< .01) | Good |
| Q4 - How do you rate the nature of information discussed to build the common vision of the situation? | Communication | 66.7 | .33 (.03) | Fair |
| Q5 - How do you rate team members' participation in the discussion on the situation? | Expertise affirmation/recognition | 93.3 | .79 (< .01) | Good |
| Q6 - How do you rate the consensus on the situation? | Attainment of consensus | 90.0 | .63 (< .01) | Good |
| Q7 - How do you rate the person-centered practice in the discussion on the situation? | Person-centered practice | 80.0 | .58 (< .01) | Moderate |
| Q8 - How do you rate the language used in the discussion on the situation? | Communication | 90.0 | .61 (< .01) | Moderate |
| Q9 - How do you rate the participants' attitude in the discussion on the situation? | Respectful attitude | 100.0 | - | - |
| Q10 - How do you rate the facilitation/regulation in the discussion on the situation? | Facilitation/regulation | 53.3 | .17 (< .01) | Slight |
| Items referring to the second phase "development of a common action plan" | Dimension measured | % agreement | κ (p value) | Interpretation* |
| Q11 - How do you rate the duration of the discussion on the action plan? | Communication | 86.7 | .60 (< .01) | Moderate |
| Q12 - Based on the interventions discussed, how do you rate team composition? | Team composition | 83.3 | .66 (< .01) | Good |
| Q13 - How do you rate the nature of information discussed about the action plan? | Communication | 66.7 | .27 (.04) | Fair |
| Q14 - How do you rate team members' participation in the discussion on the action plan? | Expertise affirmation/recognition | 70.0 | .31 (.05) | Fair |
| Q15 - How do you rate the decision-making processes? | Shared decision-making | 86.7 | .43 (< .01) | Moderate |
| Q16 - How do you rate the action plan developed? | Adoption of a common plan | 60.0 | .15 (< .01) | Slight |
| Q17 - How do you rate the person-centered practice in the discussion on the action plan? | Person-centered practice | 90.0 | .52 (< .01) | Moderate |
| Q18 - How do you rate the language used in the discussion on the action plan? | Communication | 100.0 | - | - |
| Q19 - How do you rate the participants' attitude in the discussion on the action plan? | Respectful attitude | 100.0 | - | - |
| Q20 - How do you rate the facilitation/regulation in the discussion on the action plan? | Facilitation/regulation | 70.0 | .44 (< .01) | Moderate |

* Interpretation based on Landis & Koch (1977) [39].

Finally, examination of the graphs from the cross-frequency tables indicated that the expert rater was more severe in her scoring than the non-expert rater for only three items out of 20 (items #1, 6, and 11). There was no difference between the severity of the raters' scoring for the other items.

Pragmatic properties

Pragmatic properties refer to the practical aspect related to an instrument's use. Auger et al. developed a hierarchical grouping of pragmatic criteria, grouped under the umbrella term "applicability," in a paper about measurement tools in geriatric rehabilitation [40]. Among these criteria are "respondent burden," "examiner burden," "score distribution," and "format compatibility." In this study, respondent burden and format compatibility were not formally examined. Concerning examiner burden, observers involved in Phase 3 of this study (examination of interrater reliability) reported that they only had to watch the meeting once to be able to complete the OIPC; however, they reported needing to pause the recording, during the first three meetings observed, to gather their thoughts and think about their scoring. Concerning score distribution, graphs from the cross-frequency tables showed that the response option "inadequate" was rarely chosen while "adequate" was the most frequent response. Also, in the OIPC, some items were developed to measure exactly the same aspect but for the two different temporal phases of an interprofessional team meeting. For example, item #5 ("How do you rate team members' participation in the discussion on the situation?") is paired with item #14 ("How do you rate team members' participation in the discussion on the action plan?"). Of the eight pairs of similar items, observers reported that for four of them they did not feel it was necessary to score them for each temporal phase (items #3 and #12, #7 and #17, #9 and #19, #10 and #20). Examination of the graphs from the cross-frequency tables confirmed this assumption since, for these four pairs of items, the score distribution was exactly the same for items measuring the first phase as that for items measuring the second phase. Figure 4 shows examples of what we mean by "different" and "similar" score distributions obtained in the graphs from the cross-frequency tables.

Discussion

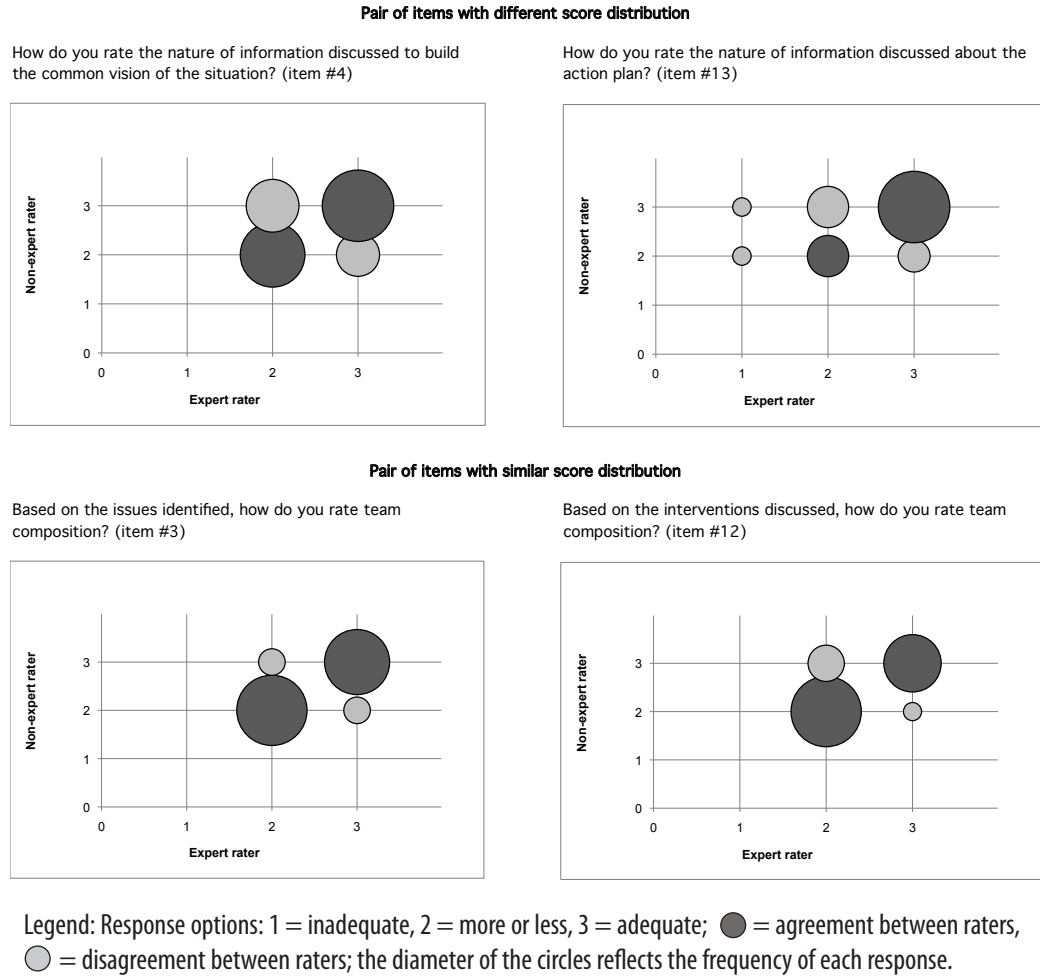
The objectives of this study were to develop an observation-based instrument to evaluate IPC interactional factors during interdisciplinary team meetings and to examine the initial validation of the tool in a rehabilitation setting. The OIPC tool was developed following iterative steps of consulting with experts and pre-testing. Interrater reliability was determined between two observers with a convenience sample of 30 video recordings of interdisciplinary meetings in two urban rehabilitation centres.

Validity of OIPC

Valentine et al. [24] divided the dimensions of IPC processes into two categories: "emerging states" and "behavioural processes." Emerging states refer to the internal

Figure 4

Examples of one pair of items measuring the same aspect in two different temporal phases with the same response distribution and one pair with a different response distribution



processes such as the psychological safety within the group, role understanding, and group cohesion. These processes are difficult to assess directly by observing participants' behaviour. So, since the OIPC is an observation-based instrument, it is not surprising to note that all of these dimensions are covered only partially or not at all by the OIPC. In fact, emerging state dimensions would probably best be measured by a tool aimed at documenting attitudes and beliefs about IPC. Dimensions related to behavioural processes are more suitable for assessment by an observation-based instrument like the OIPC, which refers to the team's behaviours and actions. Results also indicate that the dimension co-ordination as identified by Valentine et al. [23] is only partially covered by the OIPC. This is because, as depicted in our framework developed in a previous study, we considered the concept of co-ordination to underlie the type of collaborative practice called "concerted practice" [11]. In contrast, the OIPC was anchored in another type of collaborative practice we called "shared healthcare practice" since it was developed to document interdisciplinary

team meetings. Finally, the dimension collaboration identified by Valentine et al. [24] can be defined as a collective action aimed at co-constructing knowledge and reaching a consensus concerning a common goal [1,41,42]. All of these attributes are covered by the OIPC through items related to the dimensions “attainment of consensus,” “shared-decision making,” and “adoption of a common action plan.” As such, we believe the OIPC demonstrates a certain degree of content validity, although a formal validation study was not conducted.

Interrater reliability of OIPC

Results show that the interrater reliability is quite variable, ranging from “slight” to “good,” depending on the item. Even so, the reliability of the scoring of 11 out of 20 items can be considered moderate to good, which is acceptable considering the complexity of the construct measured.

One of the problematic items (i.e., with the lowest kappa) was the item associated with the dimension *facilitation/regulation* (item #10). In fact, the observers found the two items associated with this dimension difficult to score, because it measures many different aspects all related to facilitation and regulation. To develop quality standards for these items, we relied on the work of Landry [43]. According to this author, the facilitation functions can be divided into three components: co-ordination, stimulation, and regulation. Co-ordination refers to actions such as speaking time management, agenda formulation and reminders, and time management. Stimulation refers to actions aimed at ensuring that the participants work adequately toward the right objectives, so the facilitator should stimulate each person’s participation, facilitate interactions, pick up forgotten ideas, synthesize, etc. Regulation refers to actions such as conflict resolution, clarification of the team’s emotions, etc. Ideally, a good facilitator should be able to accomplish all these tasks when needed. In the OIPC, only two items assess facilitation (one for each temporal phase), and it proved very difficult to evaluate all three of the components with only these items. For example, the facilitator was sometimes very effective in accomplishing co-ordination functions but was ineffective with stimulation functions. Also, regulation functions were sometimes not accomplished at all since no conflict arose during the meeting. Thus, observers had to score those items based on their subjective judgment, which inevitably lowered interrater reliability. Because the items associated with the facilitation/regulation dimension were quite difficult to score, and considering that it is not a dimension found in the literature review of Valentine et al. [24], we suggest removing items #10 and #20 from the tool.

Two other items—“How do you rate team members’ participation in the discussion on the action plan?” (item #14), and “How do you rate the action plan developed?” (item #16)—were also problematic. It was noted that observers needed to have clinical experience and good knowledge of practitioners’ roles and responsibilities to be able to adequately score these two items. One rater was an undergraduate student with limited clinical experience, which undoubtedly reduced observer agreement. One method suggested to improve reliability, when observers have different levels of expertise, is to provide them with more training in scoring the items [35].

However, it is not clear how training could improve interrater reliability of these two items, because the disagreement between observers was not related to limited expertise concerning the use of the OIPC or even their knowledge of the construct. Rather, it was related to limited clinical expertise in the rehabilitation field. Thus, results show that, even if the OIPC was developed to be a generic instrument assessing IPC interactional factors, it probably should be used by observers who have sufficient and specific expertise related to the context in which the meeting is held.

It is interesting to note that three problematic items, and one that could not be analyzed because of low variance, were associated with the dimension communication (items #2, #4, #13, and #18). These items aimed to assess the vocabulary used (specialized discipline-specific vocabulary versus vocabulary that can be understood by every team member), and the relevance and duration of the discussion, which are related to the efficacy of communication. Again, it was hard for an observer who does not have intimate knowledge of the clinical unit's mandate and client's situation to judge these aspects. The scoring of the communication items would probably be easier for an observer who has specific clinical expertise about the context in which the meeting is held. Furthermore, communication is a multidimensional concept itself, which is not easy to assess with only a few items.

The two items related to the dimension respectful attitude (#9 and #19) could not be analyzed because of the absence of variance, with both raters consistently attributing scores of "adequate." What may explain this situation is a methodological limitation associated with the sample composition. Meetings were not recorded following random sampling, since it was the clinic co-ordinators who decided whether to record a particular meeting or not. This procedure introduced an important selection bias. We noted that potentially problematic meetings, such as involving ethical deliberations or practitioners with contradictory opinions, were automatically discarded by the clinic co-ordinators who thought the situation might deteriorate if the meeting were filmed. Nevertheless, maintaining a respectful attitude during a meeting represents a real challenge, particularly when there are differences of opinion.

Finally, the results of the OIPC's interrater reliability clearly illustrate the dilemma of "bandwidth versus fidelity," indicating that a generic tool that can be used with a variety of populations often shows worse reliability than a very specific tool [35]. Because the OIPC is a generic tool measuring a complex and multidimensional construct, it would be surprising to obtain exceptionally good reliability, despite improvements that could be made to the items or scoring system. Assessment of IPC interactional factors would always depend to a certain extent on the observer's subjective judgement concerning optimal team performance with regard to a clinical unit's mandate and a client's situation. It is important to remember that the notion of "good" or "bad" reliability cannot be absolute; rather, it is related to the nature of the tool and its context of use [35]. Since the OIPC was developed to help practitioners improve IPC practices occurring during interdisciplinary meetings, we are relatively satisfied with the tool demonstrating moderate reliability.

Instrument's pragmatic properties

As the OIPC was initially developed to be used by healthcare managers and practitioners to promote good IPC practices and improve deficient ones, it is important to discuss its pragmatic properties. First, even if IPC interactional factors represent a complex and multidimensional construct, we kept the number of items generated to a minimum to be able to complete the scoring, while watching a meeting only once. This decision reduced the examiner's burden, as well as the respondents' burden, since the OIPC could eventually be completed by a non-biased observer in real time during an actual team meeting. On the other hand, this decision also increased the scoring complexity, as it requires the observer to study many aspects of behaviour simultaneously. Even if, at first, the OIPC seems simple to use, it requires the examiner to have good observation skills and the ability to analyze an often-complex situation. The response distribution also showed that the response "inadequate" was rarely chosen. We think this situation could be explained by a selection bias, as explained earlier. Indeed, meetings identified by the clinic co-ordinator as potentially problematic or having the potential for conflicts were not included in the sample. Results also suggest that pairs of items measuring team composition (items #3 and #12), client-centred practice (items #7 and #17), the participants' attitude (items #9 and #19), and facilitation/regulation (items #10 and #20) could be merged together since there was no need to measure these aspects for each temporal phase of an inter-professional team meeting.

Future research

The OIPC was initially developed to be used in any healthcare setting. However, in this article we examined its interrater reliability only in a rehabilitation setting. To consider the OIPC to be a generic tool measuring IPC interactional factors in any interdisciplinary team meeting, its psychometric properties need to be examined in various healthcare settings. Moreover, because interesting IPC instruments have been published in the last two years, it will also be possible to determine criterion-related validity by using an instrument that fits the construct definition and has good psychometric properties.

Conclusion

The OIPC is the only generic tool aimed at evaluating IPC interactional factors during interdisciplinary meetings. It also appears to be the only tool that can evaluate team performance rather than individual behaviours in terms of IPC practices. Multiple iterative consultation steps with IPC construct experts and pre-testing with expert and non-expert observers allowed us to develop a generic tool that can be easily used in real clinical settings despite the complexity of the construct. Although the results showed moderate reliability, the OIPC version 2.4 can be useful for healthcare managers and practitioners who want to evaluate the quality of IPC practices. However, it is important to understand that the OIPC was developed to be used during interdisciplinary team meetings. Nevertheless, team meetings are only one kind of IPC modality, so we cannot judge the quality of IPC practices

occurring in a clinical setting based only on the observation of team meetings. Moreover, some emerging states related to IPC attitudes and beliefs are nearly impossible to document by observing practitioners' behaviours. For these reasons, it is important to use various approaches and different types of instruments to obtain a complete evaluation of IPC practices occurring in a given clinical setting.

Acknowledgements

The authors thank Émilie Légaré and Frédérique Poncet for their help with sample constitution and data collection. We also thank the practitioners, clinic co-ordinators, and administrators from the Institut de réadaptation en déficience physique de Québec (IRD PQ) and the Institut de réadaptation Gingras-Lindsay de Montréal (IRGLM) for their participation in this project.

Emmanuelle Careau thanks the Canadian Institutes of Health Research for the scholarship that she held at the time of this work. The authors thank the Centre for Interdisciplinary Research in Rehabilitation and Social Integration (CIRRISS) for its support and funding.

Note

1. This link (<http://www.cirris.ulaval.ca/fr/emmanuelle-careau>) connects to a French language version of this article as well as a translated version of the tool.

References

1. Oandasan, I., Ross Baker, G., Barker, K., Bosco, C., & D'Amour, D., Jones, L., Kimpton, S., Lemieux-Charles, L., Nasmith, L., San Martin Rodriguez, L., Tepper, J. & Way, D. (2006). *Teamwork in healthcare: Promoting effective teamwork in healthcare in Canada. Policy synthesis and recommendations*. Ottawa, ON: Canadian Health Services Research Foundation.
2. Curran, V., Hollett, A., Casimiro, L., McCarthy, P., & Banfield, V., Hall, P., Lackie, K., Oandasan, I., Simmons, B., & Wagner, S. (2011). Development and validation of the interprofessional collaborator assessment rubric (ICAR). *Journal of Interprofessional Care, 25*, 339-344.
3. Health Council of Canada. (2008). *Rekindling reform: Health care renewal in Canada, 2003-2008*. Toronto, ON: Health Council.
4. Romanow, R.J. (2002). *Building on values: The future of health care in Canada: Final report*. Saskatoon, SK: Commission on the Future of Health Care in Canada.
5. World Health Organization. (2010). *Framework for action on interprofessional education & collaborative practice*. Geneva: WHO.
6. Reeves, S., Lewin, S., Espin, S., & Zwarenstein, M. (2010). *Interprofessional teamwork for health and social care*. Oxford: Wiley-Blackwell.
7. Réseau de collaboration sur les pratiques interprofessionnelles en santé et services sociaux. (2010). *Étude des besoins de formation continue en lien avec la collaboration interprofessionnelle dans les établissements du RUIS-UL: Rapport d'analyse*. Québec, Canada. URL: http://www.rcpi.ulaval.ca/fileadmin/media/documents/RCPI_2010_Rapport%20Etude%20de%20besoins%20CIP.pdf.
8. Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press.
9. Donabedian, A. (1966). Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly, 44*(3), 166-206.
10. Donabedian, A. (1968). Promoting quality through evaluating the process of patient care. *Medical Care, 6*(3), 181-202.
11. Careau, E., Vincent, C., & Swaine, B.R. (2011). Consensus group session of experts to describe interprofessional collaboration processes in team meetings. *Journal of Interprofessional Care, 25*(4), 299-301.
12. Abreu, B.C., Zhang, L., Seale, G., Primeau, L., & Jones, J.S. (2002). Interdisciplinary meetings: Investigating the collaboration between persons with brain injury and treatment teams. *Brain Injury, 16*(8), 691-704.

13. Careau, E., Briere, N., Houle, N., Dumont, S., Vincent, C. & Swaine, B.R. (in press). Interprofessional Collaboration: Development of a tool to enhance knowledge translation. *Infirma Healthcare: Disability & Rehabilitation*.
14. Boxer, M.M., Vinod, S.K., Shafiq, J., & Duggan, K.J. (2011). Do multidisciplinary team meetings make a difference in the management of lung cancer? *Cancer*, 117(22), 5112-5120.
15. Fiddler, M., Borglin, G., Galloway, A., Jackson, C., McGowan, L., et al. (2010). Once-a-week psychiatric ward round or daily inpatient team meeting? A multidisciplinary mental health team's experience of new ways of working. *International Journal of Mental Health Nursing*, 19(2), 119-127.
16. Machare Delgado, E., Callahan, A., Paganelli, G., Reville, B., Parks, S.M., & Marik, P.E. (2009). Multidisciplinary family meetings in the ICU facilitate end-of-life decision making. *American Journal of Hospice & Palliative Care*, 26(4), 295-302.
17. Navarro, A.E., Wilber, K.H., Yonashiro, J., & Homeier, D.C. (2010). Do we really need another meeting? Lessons from the Los Angeles County Elder Abuse Forensic Center. *Gerontologist*, 50(5), 702-711.
18. Wittenberg-Lyles, E., Parker Oliver, D.P., Demiris, G., & Courtney, K.L. (2007). Assessing the nature and process of hospice interdisciplinary team meetings. *Journal of Hospice & Palliative Nursing*, 9(1), 17-21.
19. Ducanis, A., & Golin, A. (1979). *The interdisciplinary health care team*. Germantown, MD: Aspen Systems.
20. Lindqvist, S., Duncan, A., Shepstone, L., Watts, F., & Pearce, S. (2005). Development of the 'Attitudes to Health Professionals Questionnaire' (AHPQ): A measure to assess interprofessional attitudes. *Journal of Interprofessional Care*, 19(3), 269-279.
21. Kenaszchuk, C., Reeves, S., Nicholas, D., & Zwarenstein, M. (2010). Validity and reliability of a multiple-group measurement scale for interprofessional collaboration. *BMC Health Services Research*, 10, 83.
22. King, G., Shaw, L., Orchard, C.A., & Miller, S. (2010). The interprofessional socialization and valuing scale: A tool for evaluating the shift toward collaborative care approaches in health care settings. *Work*, 35(1), 77-85.
23. Wolf, K.N. (1999). Allied health professionals and attitudes toward teamwork. *Journal of Allied Health*, 28(1), 15-20.
24. Valentine, M.A., Nembhard, I.M., & Edmondson, A.C. (2012). *Measuring teamwork in health care settings: A review of survey instruments*. Boston, MA: Harvard Business School.
25. Bronstein, L.R. (2002). Index of interdisciplinary collaboration. *Social Work Research*, 26(2), 113-123.
26. Parker-Oliver, D., Wittenberg-Lyles, E.M., & Day, M. (2007). Measuring interdisciplinary perceptions of collaboration on hospice teams. *American Journal of Hospice and Palliative Medicine*, 24(1), 49-53.
27. Larkin, C., & Callaghan, P. (2005). Professionals' perceptions of interprofessional working in community mental health teams. *Journal of Interprofessional Care*, 19(4), 338-346.
28. Temkin-Greener, H., Gross, D., Kunitz, S.J., & Mukamel, D. (2004). Measuring interdisciplinary team performance in a long-term care setting. *Medical Care*, 42(5), 472-481.
29. Thannhauser, J., Russell-Mayhew, S., & Scott, C. (2010). Measures of interprofessional education and collaboration. *Journal of Interprofessional Care*, 24(4), 336-349.
30. Mackenzie, C.F., & Xiao, Y. (2003). Video techniques and data compared with observation in emergency trauma care. *Quality and Safety in Health Care*, 12(S2), ii51-ii57.
31. Curran, V. (2007). *Collaborative care: Synthesis series on sharing insights*. Ottawa, ON: Health Canada.
32. Healey, A.N., Olsen, S., Davis, R., & Vincent, C.A. (2007). A method for measuring work interference in surgical teams. *Cognition, Technology & Work*, 10, 305-312.
33. Malec, J.F., Torsler, L.C., Dunn, W.F., Wiegmann, D.A., Arnold, J.J., Brown, D.A., & Phatak, V. (2007). The Mayo high performance teamwork scale: Reliability and validity for evaluating key crew resource management skills. *Simulation in Healthcare*, 2(1), 4-10.
34. DeVellis, R.F. (2003). *Scale development: Theory and applications*, 2nd edition. Thousand Oaks, CA: Sage Publications.
35. Streiner, D., & Norman, G. (2008). *Health measurement scales: A practical guide to their development and use*. New York: Oxford Medical Publications.
36. DeWillis, G.B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.
37. Norimatsu, H. (2008). Initiation aux techniques d'observation. In Norimatsu, H. & Pigem, N. (Eds.), *Les techniques d'observation en sciences humaines*. (pp. 5-18). Paris: Armand-Colin.

38. Chicchetti, D.V. (1976). Assessing inter-rater reliability for rating scales: Resolving some basic issues. *The British Journal of Psychiatry*, 129, 452-456.
39. Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
40. Auger, C., Demers, L., & Swaine, B. (2006). Making sense of pragmatic criteria for the selection of geriatric rehabilitation measurement tools. *Archives of gerontology and geriatrics*, 43(1), 65-83.
41. D'Amour, D., Ferrada-Videla, M., San Martin Rodriguez, L., & Beaulieu, M.-D. (2005). The conceptual basis for interprofessional collaboration: Core concepts and theoretical frameworks. *Journal of Interprofessional Care*, 19(2), 116-131.
42. Henneman, E.A., Lee, J.L., & Cohen, J.I. (1995). Collaboration: A concept analysis. *Journal of Advanced Nursing*, 21, 103-109.
43. Landry, S. (2007). *Travail, affection et pouvoir dans les groupes restreints*. Québec: Les presses de l'Université du Québec.