

## Activity qualifiers using an argument-based construction

Esteban Guerrero<sup>1</sup>  · Juan Carlos Nieves<sup>1</sup> ·  
Marlene Sandlund<sup>2</sup> · Helena Lindgren<sup>1</sup>

Received: 1 November 2016 / Revised: 1 November 2016 / Accepted: 21 September 2017 /  
Published online: 7 October 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Based on an argumentation theory approach, we present a novel method for evaluating complex goal-based activities by generalizing a notion of *qualifier* defined in the health domain. Three instances of the general qualifier are proposed: Performance, Actuation and Capacity; the first one evaluates what a person does, the second how an individual follows an action plan, and the third one how “well” or “bad” an activity is executed. Qualifiers are intended to be used by autonomous systems for evaluating human activity. We exemplify our approach using a health domain assessment protocol. Main results of this test show a partial correlation between ambiguities assessed by experts and our argument-based approach; and a multi-dimensional perspective how an activity is executed when a combined evaluation of qualifiers is used. This last outcome was interesting for some therapists consulted. Results also show differences between values of qualifiers using different argumentation semantics; two scenarios were proposed by therapist for using different semantics: preliminary activity screening and time-span follow-up evaluation.

**Keywords** Human activity · Argumentation theory · Logic programming · Performance · Capacity · Activity evaluation

---

✉ Esteban Guerrero  
esteban@cs.umu.se

Juan Carlos Nieves  
jcnieves@cs.umu.se

Marlene Sandlund  
marlene.sandlund@umu.se

Helena Lindgren  
helena@cs.umu.se

<sup>1</sup> User, Interaction and Knowledge Modelling Group, Department of Computing Sciences, Umeå University, 901 87 Umeå, Sweden

<sup>2</sup> Department of Community Medicine and Rehabilitation, Physiotherapy Unit, Umeå University, 901 87 Umeå, Sweden

# 1 Introduction

There are different notions regarding the concept of activity and how can be evaluated. In the literature, different methods for recognize complex activities have been introduced for instance, statistical approaches from sensor-based data. They are especially suitable when recognizing sequential activities but they require enough training data to manage incomplete or uncertain data (see [1] for a review of different statistical approaches). In order to reason about complex activities, argument-based methods such as the framework introduced by Nieves et al. [25] can be used. However, such approach is abstract and methods about how to capture knowledge from sensor data or deal with inconsistencies is disregarded. Moreover, although statistical methods can deal with these data issues, some scenarios such as the health domain, not always can be managed through these approaches given the lack of enough training data.

Therefore, this work aims at develop generic methods for detecting and evaluating hierarchical human activities, which can be instantiated using for example, logic-based approaches. We introduce an activity *qualifier* that is based on an International Classification of Functioning, Disability and Health<sup>1</sup> (ICF) concept, to record the presence and severity of a problem in functioning. ICF defines two main qualifiers: *Performance* and *Capacity* [34]. In a previous work, we define these qualifiers in terms of a magnitude function [19]. We extend previous work by generalizing the notion of qualifier to evaluate different aspects of an activity, such as goals, actions and observations using the hierarchical approach of Activity Theory [22, 24].

Our approach rests on an argumentation-based process in order to: (1) generate hypothesis about the current individual's execution of an activity; (2) provide a non-monotonic behavior of the system, i.e., with the right of conclusions retraction when more information is available; and (3) obtain consistent sets of hypothesis explaining pieces of the activity.

We extend and generalize our previous work about methods for evaluating human activities [19, 25]. We use data from a previous pilot study based on an assessment protocol: the Short Physical Performance Battery (SPPB) test [20] for demonstrating our contributions. The main aims of the pilot study were:

- *Aim 1*: explore how qualifiers in an actual situation can be used and demonstrate their applicability.
- *Aim 2*: evaluate effects on qualifiers empirically, using different argumentation semantics.

The contributions of this paper are:

- A generalization for the notion of an activity qualifier.
- Three instances of the general qualifier are defined: *Capacity*, *Actuation* and *Performance* qualifiers.
- We demonstrate the applicability of qualifiers based on “classical” argumentation semantics [11], using data from an experimental pilot.
- Results of the pilot test show: (1) partial correlation between ambiguities assessed by experts and our argument-based approach; and (2) usefulness when qualifiers are combined.

The rest of the paper is structured as follows: in Sect. 2 we present the theories and methods used. In Sect. 3, we present key definitions of argumentation theory as well as we present some contributions for reasoning about complex activities using an argumentation approach.

<sup>1</sup> The International Classification of Functioning, Disability and Health was officially endorsed by the World Health Organization as the international standard to describe and measure health and disability. <http://www.who.int/classifications/icf>.

The notions of a general qualifier as well as some instances as Performance, Actuation and Capacity qualifiers are defined in Sect. 3. In Sect. 4, we present results from a pilot study. Our contributions are discussed in Sect. 5 and some conclusions and future work are summarized in Sect. 6.

## 2 Background

In this section some relevant background regarding Argumentation Theory and the underlying language is introduced. The focus of this section is on Phase 1 and Phase 2 of the information processes represented in Fig. 1. In the next section, Phase 3, the qualifier generation is introduced. For the current section, we assume that the reader is familiar with basic terms of Argumentation Theory [28].

### 2.1 Underlying logical language

The type of physical activities considered in this paper has different elements such as *observations* as well as *goals* which can be captured by logic programs with negations as failure (NAF) represented by *not*.

We use a propositional logic with a syntax language constituted by propositional symbols:  $p_0, p_1, \dots$ ; connectives:  $\wedge, \leftarrow, \neg, not, \top$ ; and auxiliary symbols:  $(, )$ , in which  $\wedge, \leftarrow$  are 2-place connectives,  $\neg, not$  are 1-place connectives and  $\top$  is a 0-place connective. Propositional symbol  $\top$  and symbols of the form  $\neg p_i (i \geq 0)$  stand for indecomposable propositions which we call *atoms*, or *atomic propositions*. Atoms of the form  $\neg a$  are called *extended atoms* in the literature. An *extended normal clause*,  $C$ , is denoted:  $a \leftarrow b_1, \dots, b_j, not\ b_{j+1}, \dots, not\ b_{j+n}$  where  $j + n \geq 0$ ,  $a$  is an atom and each  $b_i (1 \leq i \leq j + n)$  is an atom. When  $j + n = 0$  the clause is an abbreviation of  $a \leftarrow \top$  such that  $\top$  always evaluates true. An *extended normal program*  $P$  is a finite set of extended normal clauses. By  $\mathcal{L}_P$ , we denote the set of atoms which appear in a program  $P$ . ELP use both strong negation  $\neg$  and *not*, representing common-sense knowledge through logic programs. On programs with NAF, the consequence operator:  $\leftarrow$  is not monotonic, which means that the evaluation result, may change as more information is added to the program. Two major semantics for ELP have been defined: (1) answer set semantics [16], an extension of *Stable model semantics*, and (2) a version of the Well-Founded Semantics (WFS) [30]. WFS performs a *skeptical* reasoning approach being both polynomial time computable and always defined. In contrast to Stable, WFS satisfies the *relevance* property which allows

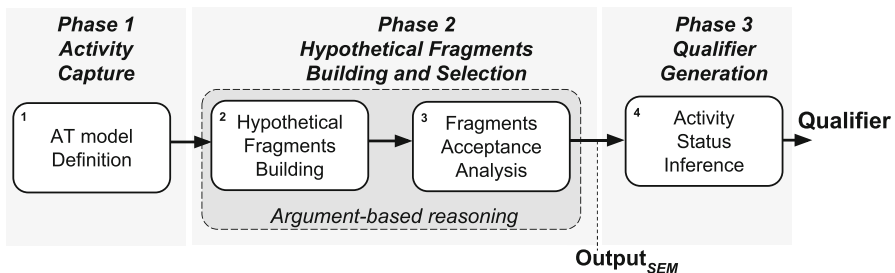


Fig. 1 Diagram of an argument-based system for generating qualifiers

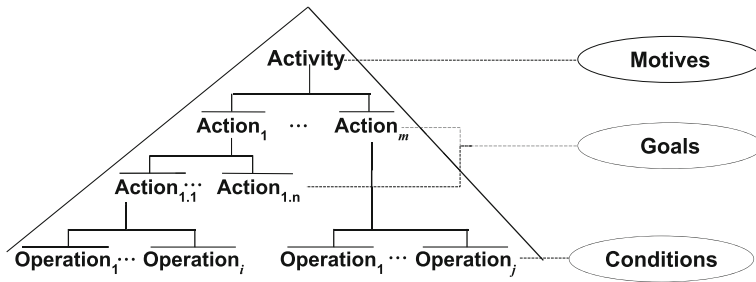


Fig. 2 Activity Theory hierarchical structure. Adapted from [22]

us infer consistent conclusion and avoiding problems associated with the so-called *conflict propagation* [14].

### 2.2 Activity evaluation

We follow a general and systemic approach from Social Sciences suitable for the representation of different human activities: Activity Theory [24]. Activity theory defines a three-layer hierarchy (hereinafter *AT model*): *activity*, consisting of a set of *actions*, which in turn may consist of actions and *operations* in a nested structure such is presented in Fig. 2.

Our study is framed on the evaluation of activities based on the AT model where the following assumptions hold: (1) activities consists in sets of actions directed to a goal;<sup>2</sup> (2) goals and actions cannot exist outside of an activity; and (3) goals have subgoals with more granularity in a goal, more unconsciousness level is reached. In this setting, an AT model can be defined as follows:

**Definition 1** Let  $P$  be a logic program capturing the behavior rules of an activity.  $\mathcal{L}_P$  denotes the set of atoms which appear in a program  $P$ . An **AT model** is a tuple of the form  $\langle Ax, Go, Op \rangle$  in which:

- $Ax = \{ax_1, \dots, ax_j\}(j > 0)$  is a set of atoms such that  $Ax \subseteq \mathcal{L}_P$ .  $Ax$  denotes the set of actions in an AT model.
- $Go = \{g_1, \dots, g_k\}(k > 0)$  is a set of atoms such that  $Go \subseteq \mathcal{L}_P$ .  $Go$  denotes the set of goals in an AT model.
- $Op = \{o_1, \dots, o_l\}(l > 0)$  is a set of atoms such that  $Op \subseteq \mathcal{L}_P$ .  $Op$  denotes the set of goals in an AT model.

An AT model as is defined in Definition 1 establishes the information setting in which a human activity is performed.

In order to exemplify the notion of a structured activity let us introduce a running example.

*Example 1* (A Short Physical Performance Battery—SPPB test) Consider a physical activity described by a well-known assessment protocol SPPB test [20]. This assessment protocol evaluates the lower-extremity function by measuring the standing balance, walking speed, and ability to rise from a chair of a person, see Table 1. The SPPB scenario can be captured

<sup>2</sup> Actions have goals and are executed by a human agent at a conscious level, in contrast with observations which do not have a goal of their own and which are executed at the lowest level as automated, unconscious processes.

**Table 1** SPPB test as a structured activity

Activity	Modified Short Physical Performance Battery – SPPB				
	Side-by-side	Semi-tandem	Tandem	Sit-to-stand	Sway
Goals (Go)	Goal 1 correct ( $g_1$ ) incorrect ( $\neg g_1$ )	Goal 2 correct ( $g_2$ ) incorrect ( $\neg g_2$ )	Goal 3 correct ( $g_3$ ) incorrect ( $\neg g_3$ )	Goal 4 correct ( $g_4$ ) incorrect ( $\neg g_4$ )	Goal 5 correct ( $g_5$ ) incorrect ( $\neg g_5$ )
Actions (Ax)	<ul style="list-style-type: none"> <li>• be standing (<math>ax_1</math>)</li> <li>• holding feet position (<math>ax_2</math>)</li> </ul>			<ul style="list-style-type: none"> <li>• rising up (<math>ax_3</math>)</li> <li>• sitting down (<math>ax_4</math>)</li> <li>• fold arms across the chest (<math>ax_5</math>)</li> </ul>	<ul style="list-style-type: none"> <li>• swing back and forth (<math>ax_6</math>)</li> </ul>
Operations (Op)	<ul style="list-style-type: none"> <li>• Much sway (<math>o_1</math>)</li> <li>• Little sway (<math>o_2</math>)</li> <li>• ...</li> </ul>	<ul style="list-style-type: none"> <li>• Slow sitting down (<math>o_3</math>)</li> <li>• Fast sitting down (<math>o_4</math>)</li> <li>• ...</li> </ul>	<ul style="list-style-type: none"> <li>• Slow rising up (<math>o_5</math>)</li> <li>• Fast rising up (<math>o_6</math>)</li> <li>• ...</li> </ul>		

by a logic program defining rules which governing the behavior of an agent<sup>3</sup> supporting the SPPB execution. In order to simplify the presentation, a goal such as *correct side-by-side test* will be expressed as  $g_1$ , observations, e.g., *slow rising up* as  $o_5$ , and actions noted by atoms, e.g.,  $ax_1$ , as is presented in Fig. 3.

In Example 1, the logic program defines the knowledge base of an agent, describing the behavior with all the possible decisions that an agent can take in reference to an AT model. The behavior of the agent is ruled by a goal which is shared with a person, and information from the environment. For example, a clause such as:  $\neg g_1 \leftarrow o_1 \wedge ax_1 \wedge ax_2 \wedge h_3$ , can be read as: “if the agent observes that person is standing ( $ax_1$ ), holding the feet position ( $ax_2$ ) and it is detected strong sway ( $o_1$ ) and it decides do nothing ( $h_3$ ), it can be the cause for the not achievement of a correct side-by-side test ( $\neg g_1$ )”

The notion of *qualifier* was first introduced in [19] with an emphasis to physical activities in health domain. Let us start recall the notion of an activity framework [19] as follows:

**Definition 2** (Activity framework) An activity framework  $ActF$  is a tuple of the form  $\langle P, \mathcal{H}_A, \mathcal{G}, \mathcal{O}, AT \rangle$  in which:

- $P$  is a logic program.  $\mathcal{L}_P$  denotes the set of atoms which appear in  $P$ .
- $\mathcal{H}_A = \{h_1, \dots, h_i\}$  is a set of atoms such that  $\mathcal{H}_A \subseteq \mathcal{L}_P$ .  $\mathcal{H}_A$  denotes the set of hypothetical actions which an agent can perform in a world.
- $\mathcal{G} = \{g_1, \dots, g_j\}$  is a set of atoms such that  $\mathcal{G} \subseteq \mathcal{L}_P$ .  $\mathcal{G}$  denotes a set of goals of an agent.
- $\mathcal{O} = \{o_1, \dots, o_k\}$  is a set of atoms such that  $\mathcal{O} \subseteq \mathcal{L}_P$ .  $\mathcal{O}$  denotes a set of world observations of an agent.
- $AT$  is an activity model of the form:  $\langle Ax, Go, Op \rangle$ , following Definition 1.

Example 1 shows how an intelligent agent can support a person in the execution of an activity, dealing with uncertainty by using negation as failure. For instance, an intuitive reading of clause  $o_2 \leftarrow not\ o_1$  (in Table 2) can be understood as “if there are no evidence that

<sup>3</sup> We assume here that a software-based agent assists or it is used by a physician in an assessment procedure, such as the SPPB test.

$$P := \left\{ \begin{array}{l} g_1 \leftarrow ax_1 \wedge ax_2 \wedge o_2 \wedge h_1 \\ \neg g_1 \leftarrow o_1 \wedge ax_1 \wedge ax_2 \wedge h_3 \\ g_2 \leftarrow ax_1 \wedge ax_2 \wedge o_2 \wedge ax_1 \\ \neg g_2 \leftarrow o_1 \wedge ax_1 \wedge ax_2 \wedge h_3 \\ g_3 \leftarrow ax_1 \wedge ax_2 \wedge o_2 \wedge h_1 \\ \neg g_3 \leftarrow o_1 \wedge ax_1 \wedge ax_2 \wedge h_3 \\ g_4 \leftarrow g_5 \wedge ax_3 \wedge ax_4 \wedge ax_5 \wedge o_3 \wedge o_6 \wedge h_1 \\ g_5 \leftarrow ax_6 \wedge o_2 \wedge h_1 \\ \neg g_4 \leftarrow ax_3 \wedge ax_4 \wedge ax_5 \wedge o_4 \wedge o_5 \wedge h_1 \\ \neg g_5 \leftarrow ax_6 \wedge o_1 \wedge h_2 \\ \\ o_1 \leftarrow not\ o_2 \\ o_2 \leftarrow not\ o_1 \\ o_3 \leftarrow not\ o_4 \\ o_4 \leftarrow not\ o_3 \\ o_5 \leftarrow not\ o_6 \\ o_6 \leftarrow not\ o_5 \end{array} \right.$$

$$\mathcal{O} := \left\{ \begin{array}{l} o_1, o_2, o_3, o_4, o_5, o_6, \\ ax_1, ax_2, ax_3, ax_4, ax_5 \end{array} \right\}$$

$$\mathcal{G} := \left\{ \begin{array}{l} g_1, g_2, g_3, g_4, g_5, \\ \neg g_1, \neg g_2, \neg g_3, \neg g_4, \neg g_5 \end{array} \right\}$$

$$\mathcal{H}_A := \{h_1, h_2, h_3\}$$

Being  $h_1$  : positive feedback;  $h_2$  : warning message;  $h_3$  : do nothing

**Fig. 3** Partial structure of a logic program capturing the sit-to-stand test of the SPPB for an individual

a person has a much sway, the agent assumes that there are little sway in the operation". This example shows a default rule capturing the lack of knowledge about the sway observation.

In [19] a method to generate explanations of an activity based on the information of an activity frameworks was introduced. Hypothetical fragments<sup>4</sup> of an activity can be built by using an answer set programming approach for building arguments [18].

**Definition 3** (Hypothetical fragment of an activity) Let  $ActF = \langle P, \mathcal{H}_A, \mathcal{G}, \mathcal{O}, AT \rangle$  be an activity framework. A hypothetical fragment of an activity is of the form  $HF = \langle S, O', h, g \rangle$  such that:

1.  $S \subseteq P, O' \subseteq \mathcal{O}, h \in \mathcal{H}_A, g \in \mathcal{G}$ ,
2.  $S \cup O' \cup \{h\}$  is consistent,<sup>5</sup>

<sup>4</sup> Hypothetical fragments are based on the notion arguments. In argumentation theory literature, an argument is usually a tuple  $\langle support, conclusion \rangle$ .

<sup>5</sup> In argumentation literature, a set  $S \subseteq P$  is consistent iff  $\nexists \psi, \phi \in P$  such that  $\psi = \neg\phi$ , Definition 6. Consistent set in [10]

**Table 2** Set of fragments describing the chair-standing test of the SPPB scenario

$F_1 =$	$\langle \{g_1 \leftarrow ax_1 \wedge ax_2 \wedge o_2 \wedge h_1; o_2 \leftarrow not\ o_1\},$ $\{o_2, ax_1, ax_2\},$ $h_1,$ $g_1 \rangle$ ...
$F_7 =$	$\langle \{g_4 \leftarrow g_5 \wedge ax_3 \wedge ax_4 \wedge ax_5 \wedge o_3 \wedge o_6 \wedge h_1;$ $g_5 \leftarrow ax_6 \wedge o_2 \wedge h_1; o_2 \leftarrow not\ o_1;$ $o_3 \leftarrow not\ o_4; o_6 \leftarrow not\ o_5\},$ $\{o_2, o_3, o_6, ax_3, ax_4, ax_5, ax_6\},$ $h_1,$ $g_4 \rangle$
$F_8 =$	$\langle \{g_5 \leftarrow ax_6 \wedge o_2 \wedge h_1; o_2 \leftarrow not\ o_1\},$ $\{o_2, ax_6\},$ $h_1,$ $g_5 \rangle$ ...
$F_{10} =$	$\langle \{\neg g_5 \leftarrow o_1 \wedge h_2; o_1 \leftarrow not\ o_2\},$ $\{o_1, 6\},$ $h_2,$ $\neg g_5 \rangle$ ...

3.  $g \in T$  such that  $WFS(S \cup O' \cup \{h\}) = \langle T, F \rangle$ ,
4.  $S$  and  $O'$  are minimal w.r.t. set inclusion.

$WFS(S)$  is a function inferring the Well-Founded Semantics (WFS) [30].

A fragment  $HF = \langle S, O', h, g \rangle$  (Definition 3) establishes conditions of the agent’s behavior adapted to the execution of a human activity. Fragments are goal-oriented considering the uncertainty to achieve such goal through an action given a set of captured observations. Fragments can be seen as *conjectures* that an agent creates to evaluate an activity.

Let us denote by  $\mathcal{HF}$  the set of all the hypothetical fragments obtained by applying Definition 3 to a logic program  $P$ .  $\mathcal{HF}$  can be seen as a set of activity explanations in a given situation. The support of a fragment can be an assembled substructure of other fragments, the so-called *subfragments*. In order to define this concept, we introduce some auxiliary functions **Supp** and **Concl** which return the support and conclusion of a given fragment, respectively, e.g., given the fragment  $HF = \langle S, O', h, g \rangle$ , we have  $Supp(HF) = S \cup O' \cup \{h\}$  and  $Concl(HF) = \{g\}$ .

*Example 2* Consider the SPPB scenario introduced in Example 1. Using Definition 3 we can build 11 fragments (see Table 2).

**Definition 4 (Subfragment)** Let  $HF_1 = \langle S_1, O'_1, h_1, g_1 \rangle, HF_2 = \langle S_2, O'_2, h_2, g_2 \rangle$  be two fragments of an activity.  $HF_1$  is a subfragment of  $HF_2$  iff  $Supp(HF_1) \subseteq Supp(HF_2)$ .

The notion of subfragment in Definition 4 allows us provide hypotheses regarding subgoals in the hierarchical structure of an activity.

Using Definition 3 an agent can build different explanations, even with opposed or contradictory conclusions. For instance, let us suppose that the agent used in Example 1 uses multiple sensors, one of them incorrectly working stating that an individual is having too much sway ( $o_1$  in Table 1) in the sit-to-stand test. At the same time, another sensor may not have evidence about much swaying, inferring so that the person has little sway ( $o_2 \leftarrow not\ o_1$  in Table 1). In this case, two different and contradictory fragments are built by the agent ( $F_8$  and  $F_{10}$  in Table 2). Different types of contradictory relationship among fragments and subfragments can be defined. In the argumentation literature, this relationships are called *attacks*.

**Definition 5** (*Contradictory relationships among fragments*)

Let  $ActF = \langle P, \mathcal{H}_A, \mathcal{G}, \mathcal{O}, Acts \rangle$  be an activity framework. Let  $HF_1 = \langle S_1, O'_1, a_1, g_1 \rangle$ ,  $HF_2 = \langle S_2, O'_2, a_2, g_2 \rangle$  be two fragments such that  $HF_1, HF_2 \in \mathcal{HF}$ .  $WFS(\text{Supp}(HF_1)) = \langle T_1, F_1 \rangle$  and  $WFS(\text{Supp}(HF_2)) = \langle T_2, F_2 \rangle$  denote the semantic evaluation of the support, then  $HF_1$  attacks  $HF_2$  if one of the following conditions hold: (1)  $\alpha \in T_1$  and  $\neg\alpha \in T_2$ ; (2)  $\alpha \in T_1$  and  $\alpha \in F_2$ .

In argumentation theory literature, an *argumentation framework* can be seen as a directed graph where vertices are arguments and the edges are attack/support relationships. An argumentation framework is a pair  $\langle Args, att \rangle$  in which  $Args$  is a finite set of arguments and  $att \subseteq Args \times Args$ . In [19] an argumentation-based activity framework for reasoning about activities was proposed, by considering argumentation as inference method:

**Definition 6** (*Activity argumentation framework*) Let  $ActF$  be an activity framework of the form  $\langle P, \mathcal{H}_A, \mathcal{G}, \mathcal{O}, Acts \rangle$ ; let  $\mathcal{HF}$  be the set of fragments w.r.t.  $ActF$  and  $Att_{\mathcal{HF}}$  or simply  $Att$  the set of all the attacks among  $\mathcal{HF}$ . An activity argumentation framework AAF with respect to  $ActF$  is of the form:  $AAF = \langle ActF, \mathcal{HF}, Att \rangle$

In argumentation theory literature, Dung in his seminal work [11] introduced a set of *patterns of selection* of arguments called *argumentation semantics*. Intuitively, an argumentation semantics SEM is a formal method to identify conflict outcomes from argumentation frameworks (AF). The sets of arguments suggested by an argumentation semantics are called *extensions* which can be regarded as “the best” explanation for the current situation. Let  $SEM()$  be a function returning a set of extensions, given an AF such as an AAF. In this sense, we can denote  $SEM(AAF) = \{Ext_1, \dots, Ext_k\}$  as the set of  $k$  extensions generated by an argumentation semantics w.r.t. an activity argumentation framework AAF. In this setting, from the perspective of an intelligent agent what it is expected to have is: (1) no contradictory or conflicting sets of fragments sets explaining what is happening in the ongoing activity, and (2) fragments sets defending/supporting a hypothesis about the activity from other fragments. These two notions defines two main concepts in Dung’s argumentation semantics: acceptable and admissible arguments.

**Definition 7** (1) An fragment  $HF_A \in \mathcal{HF}$  is *acceptable* w.r.t. a set  $S$  of fragments *iff* for each fragment  $HF_B \in \mathcal{HF}$ : if  $HF_B$  attacks  $HF_A$ , then  $HF_B$  is attacked by  $S$ . (2) *conflict-free* set of fragments  $S$  in an activity is *admissible* *iff* each fragment in  $S$  is acceptable w.r.t.  $S$ .

Using these notions of fragment admissibility, different argumentation semantics can draw given an activity argumentation framework:

**Definition 8** Let  $AAF = \langle ActF, \mathcal{HF}, Att \rangle$  be an activity argumentation framework following Definition 6. An admissible set of fragments  $S \subseteq \mathcal{HF}$  is: (1) *stable* if and only if  $S$  attacks each fragment which does not belong to  $S$ ; (2) *preferred* if and only if  $S$  is a maximal



(w.r.t. inclusion) admissible set of AAF; (3) *complete* if and only if each fragment, which is acceptable with respect to  $S$ , belongs to  $S$ ; and (4) the *grounded* extension of AAF if and only if  $S$  is the minimal (w.r.t. inclusion) complete extension of AAF.

It is easy to see that the output of an argumentation-based system (ABS) depends on directly of the selected argumentation semantics (see Fig. 1). In order to encapsulate the behavior of an ABS, we can define two useful functions:

**Definition 9** (*ABS output and conclusions set*) Let  $AAF = \langle ActF, \mathcal{HF}, Att \rangle$  be an activity argumentation framework and SEM be an argumentation semantics, then if  $SEM(AAF) = \{E_1, \dots, E_n\} (n \geq 0)$ , then:  $Concs(E_i) = \{Concl(HF) \mid HF \in E_i\} (1 \leq i \leq n)$ ; and  $Output_{SEM} = \bigcap_{i=1, \dots, n} Concl(E_i)$ .

In Definition 9, we can differentiate the scope of **Concs** and **Output**. The later has a skeptical approach, being *skepticism* a term related to make more or less committed evaluations about the justification state of fragments in a given situation: a more skeptical attitude corresponds to less committed evaluations of an activity [5]. Intuitively, a skeptical view is opposed to a credulous position.

### 3 Qualifiers for evaluating activity

Intuitively, a qualifier is an argument-based evaluation of the current status of activity elements: operations, actions or goals, w.r.t. a reference value.<sup>6</sup>

In Definition 9,  $Output_{SEM}$  suggests a consistent explanation about the current status of an activity. A reference value of activity elements, for example goals or observations can be defined as a set of hypothetical fragments conclusions  $Concs(\mathcal{HF}_{ref})$ . A qualifier is a metric comparing  $Output_{SEM}$  w.r.t.  $Concs(\mathcal{HF}_{ref})$ . In this regard, we can define a general *similarity function* to compare these sets, for example the current goal achievement status of an activity ( $G_C$ ) w.r.t. a set of reference goal values ( $G_R$ ) as follows:  $Sim(G_C, G_R) : 2^G \times 2^G \rightarrow \mathbb{R}$ , where  $G \subseteq Go$  and  $Sim(G_C, G_R) = n \in \mathbb{R}$ , with  $G_C, G_R \in 2^G$ . In this setting, a qualifier is defined by:

**Definition 10** (*Qualifier*)

Let  $AAF = \langle ActF, \mathcal{HF}, Att \rangle$  be an activity argumentation framework, let SEM be an argumentation semantics with  $SEM(AAF) = \{E_1, \dots, E_n\} (n \geq 1)$  and HF be a fragment. Let  $Concs(E_i) = \{Concl(HF) \mid HF \in E_i\} (1 \leq i \leq n)$  be the conclusions of an argument-based system; let  $Output_{SEM} = \bigcap_{i=1, \dots, n} Concl(E_i)$  be its output; and let  $Concs(\mathcal{HF}_{ref})$  be a set of reference values. A *qualifier* is defined as:

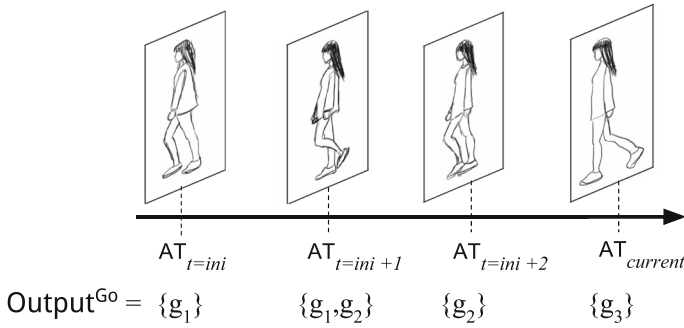
$$Q = Sim(Output_{SEM}, Concs(\mathcal{HF}_{ref}))$$

where  $\mathcal{HF}_{ref} \subseteq \mathcal{HF}$  s.t. if  $a, b \in \mathcal{HF}_{ref}$  then  $(a, b) \notin Att$ .

Roughly speaking,  $Concs(\mathcal{HF}_{ref})$  in Definition 10 is information taken as an external criterion<sup>7</sup> for validating/comparing the *correctness* of the current status of an activity:

<sup>6</sup> A qualifier resembles the type of activity evaluation performed by therapist in health care, particularly *descriptive assessments* which provide a snapshot of the person's function, circumstances and needs at one point in time only and are often used to provide a baseline [15].

<sup>7</sup> In health literature this external criterion is sometimes called a *gold standard* and it is used in different types of human assessments. In [15] different principles of assessments protocols are defined.



**Fig. 4** Example 3 output of an argument-based system in the SPPB test

$Output_{SEM}$ . In this definition, a minimal requirement of the set  $\mathcal{HF}_{ref}$  is to be conflict-free, i.e., there are no fragments  $a$  and  $b$  in  $\mathcal{HF}_{ref}$  such that  $a$  attacks  $b$ . Considering Example 1, a reference set of non-conflicting observations can be  $Concs(\mathcal{HF}_{ref}) = \{o_6, o_3\}$  which intuitively says that a person rises from a chair quickly and sits down slowly.

### 3.1 Instantiating qualifiers

According to the ICF, Performance describes what an individual “really does”. In this setting, Performance can be seen as a measurement of activity goal achievement, as follows:

**Definition 11** (*Performance qualifier*)

Let  $Sim$  be a similarity function; and let  $Output_{SEM}^G$  and  $Concs^G$  be output sets from an argument-based process in terms of goals as it was defined in Definition 10. Performance qualifier  $Perf$  is given by:

$$Perf = Sim(Output_{SEM}^G, Concs^G(\mathcal{HF}_{ref}))$$

The left side of  $Perf$  in Definition 11 evaluates the current status of goal achievement in an activity. Right side of  $Perf$  can be seen as a “correct” way to achieve such activity in terms of goals.

*Example 3* Let us consider Example 1 with the SPPB test being an activity composed by four tasks aimed to achieve five goals:  $Go = \{g_1, g_2, g_3, g_4, g_5\}$ . Let us suppose that a therapist considers that the achievement of the three first goals is a good indication of an acceptable activity performance. The system takes observations four times during the activity as is presented in Fig. 4. In this setting, a *reference* value can be the achievement of  $\{g_1, g_2, g_3, g_4, g_5\}$  goals for the SPPB as an activity model  $AT$ .

The scenario depicted in Fig. 4 can be expressed as follows:

$$Perf = Sim(Output_{SEM}^G, Concs^G(\mathcal{HF}_{ref}))$$

$$Perf = Sim(\{g_1, g_2, g_3\}, \{g_1, g_2, g_3, g_4, g_5\})$$

In Example 3, similarity function  $Sim$  can produce different values depending on normalization factors, for instance if we consider a simple goal counting evaluation, Performance in Example 3 would be:  $Perf = Sim(|\{g_1, g_2, g_3\}|, |\{g_1, g_2, g_3, g_4, g_5\}|) = Sim(3, 5) = 0.6$  or 60% of goal achievement. In the health literature, particularly considering physical assessment protocols, there are no a consensus about the range for a quantitative measurement [15].

The ICF Capacity qualifier measures how “well” or “bad” an individual executes an activity [33]. We instantiate the generic qualifier  $Q$  (Definition 10) in terms of observations. In this manner, a Capacity qualifier provides us with a tool to evaluate operative movements which are captured in the execution of goal-based action.

**Definition 12** (*Capacity qualifier*)

Let  $\text{Sim}$  be a similarity function; and let  $\text{Output}_{\text{SEM}}^O$  and  $\text{Concs}^O$  be output sets from an argument-based process in terms of activity observations. Capacity qualifier  $Cap$  is given by:

$$Cap = \text{Sim}(\text{Output}_{\text{SEM}}^O, \text{Concs}^O(\mathcal{HF}_{\text{ref}}))$$

Similar to Performance qualifier, the left side in Definition 12 evaluates the current operations as observations of an activity. Right side of  $Cap$  is a reference set of operational processes in such activity.

*Example 4* Let us consider an extension of Example 3 obtaining as output the following set of observed operations:

- $t_{ini}$   $\text{Output}_{\text{SEM}}^O = \{o_2\}$
- $t_{ini+1}$   $\text{Output}_{\text{SEM}}^O = \{o_2, o_3, o_6\}$
- $t_{ini+2}$   $\text{Output}_{\text{SEM}}^O = \{o_2\}$
- $t_{current}$   $\text{Output}_{\text{SEM}}^O = \{o_1\}$

A therapist can consider that a person can execute well an activity if precise operational movements are observed, a *reference* value in terms of operations for this particular case can be  $\{o_2, o_3, o_6\}$ . The capacity qualifier can be expressed as follows:

$$Cap = \text{Sim}(\text{Output}_{\text{SEM}}^O, \text{Concs}^O(\mathcal{HF}_{\text{ref}}))$$

$$Cap = \text{Sim}(\{o_2, o_3, o_6\}, \{o_2, o_3, o_6\})$$

Assuming a simple version of  $\text{Sim}$  function as an observation counting evaluation in Example 4, we have that:  $Cap = 1$  or in other words, the Capacity qualifier suggests a 100% of observed operations executed considering the reference set.

Instantiating the general qualifier  $Q$  w.r.t. the set of actions governing an activity, we can obtain a tool for quantifying if a person executes or not a defined set of actions in order to achieve a given goal set. We call this instance of  $Q$  the *actuation* qualifier, and is defined as follows:

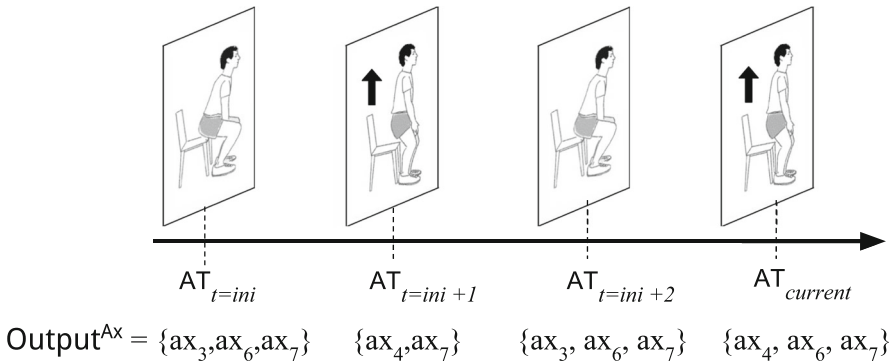
**Definition 13** (*Actuation qualifier*)

Let  $\text{Sim}$  be a similarity function; and let  $\text{Output}_{\text{SEM}}^A$  and  $\text{Concs}^A$  be output sets from an argument-based process in terms of actions (A) framed on a particular activity. An actuation qualifier *Actuate* is given by:

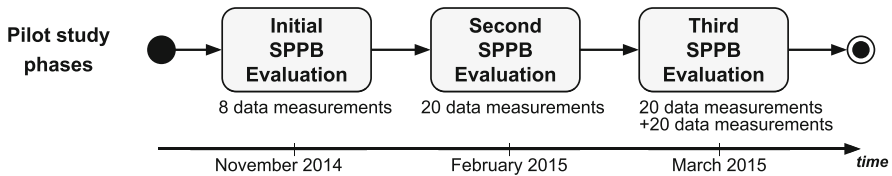
$$Actuate = \text{Sim}(\text{Output}_{\text{SEM}}^A, \text{Concs}^A(\mathcal{HF}_{\text{ref}}))$$

*Example 5* Let us consider an extension of Example 3 using the SPPB scenario. In Table 1, a set of actions were defined oriented to five goals in the SPPB. Let us suppose an extra action  $ax_7 = \neg ax_5$  the action of not folding arms across the chest during the execution of the sit-to-stand test. And let us suppose that a therapist considers the set of actions:  $\{ax_3, ax_4, ax_5, ax_6\}$  as a reference set. Figure 5 shows the updated scenario.

- $t_{ini}$   $\text{Output}_{\text{SEM}}^A = \{ax_3, ax_7, ax_6\}$
- $t_{ini+1}$   $\text{Output}_{\text{SEM}}^A = \{ax_4, ax_7\}$



**Fig. 5** Example 5 output of an argument-based system considering the set of actions executed in the SPPB scenario



**Fig. 6** Pilot study workflow

- $t_{ini+2}$   $Output_{SEM}^A = \{ax_3, ax_7ax_6\}$
- $t_{current}$   $Output_{SEM}^A = \{ax_4, ax_7ax_6\}$

This scenario can be expressed as follows:

$$Actuate = Sim(Output_{SEM}^A, Concs^A(\mathcal{H}\mathcal{F}_{ref}))$$

$$Actuate = Sim(\{ax_3, ax_4, ax_6, ax_7\}, \{ax_3, ax_4, ax_5, ax_6\})$$

Intuitively, Actuate qualifier (Definition 13) can be oriented to the evaluation of action supervision in a goal-oriented scenario. In Example 5, the reference action set is close enough to the evaluation of the individual at “current” time.

### 4 Demonstration of qualifiers in a use case

In this section, we demonstrate our approach through data obtained in a pilot study [19]. We extend previous results with a further analysis of qualifiers and the argument semantics governing the behavior of our argument-based system.

#### 4.1 Pilot study setting

The pilot study presented in [19], included three data collection phases using a modified version of the SPPB test, as is represented in Fig. 6. Sensor data were collected in the three phases by using a sensor-based mobile application called *Balansera* (<https://github.com/esteban-g/Balansera-mob>) which collects acceleration data. The mobile phone with the Balansera application was placed in the lower back using a belt for maintaining its position. The population sample inclusion criteria was: people over 65 years, ability to rise from a



**Fig. 7** Balansera mobile application. The mobile application was tested and used by individuals native Swedish speaking and the dialogues, video demonstrations and texts are in Swedish

chair with a seat height of 45 cm., with his arms crossed over her/his chest and the ability to understand instructions on Swedish language, all the participants were assessed in laboratory conditions. The first group, 8 participants (7 female, 1 male) was assessed in balance and chair-standing using the SPPB. Similarly, the second group of 20 participants (11 female 9 male) was assessed in balance and chair-standing tasks; and, a third group of 20 participants (11 female 9 male) was assessed in chair-standing twice. In each occasion data measurements were obtained, in total, the data set contains 68 measurements from 28 different older adults.

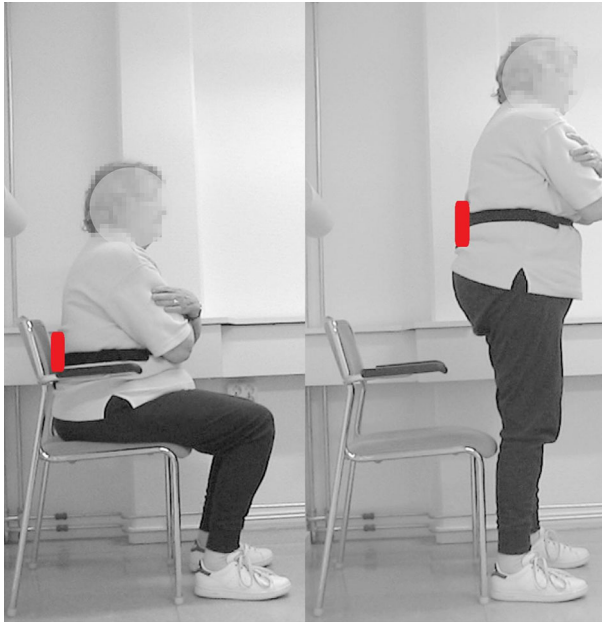
## 4.2 Data acquisition

Assessments were carried out by a physiotherapist in laboratory conditions using an adaptation of the SPPB test as follows:

- Test of standing balance included *tandem*, *semi-tandem* and *side-by-side stands*. For each stand, the interviewer first demonstrated the task using a video displayed in the Balansera mobile application (Fig. 7). The individual positioned their feet and the therapist asked if he/she was ready. The timing was stopped when individual moved their feet or grasped the therapist for support, or when 10 s had elapsed. Each participant began with semi-tandem stand. Those unable to hold the semi-tandem position for 10 s were evaluated with the feet in the side-by-side position. Those able to maintain the semi-tandem position for 10 s were further evaluated with the feet in full tandem position.
- For the test of the ability to rise from a chair, a straight-backed chair was placed next to the wall; participants were asked to fold their arms across their chest and to stand up from the chair five times using a normal speed.<sup>8</sup> The timing starts from the initial sitting position to the final standing position at the end of the fifth stand.

Sensor data were collected using the Balansera mobile application which collects acceleration and time data. The mobile phone with the Balansera application was placed in the lower back using a belt for maintaining its position, as it is shown in Fig. 8. Balansera was developed considering the SPPB test and it follows the same step-by-step assessment procedure previously described (Fig. 7).

<sup>8</sup> In the rise from a chair test introduced in [20] the participants are asked to stand up and sit down five times as quickly as possible.



**Fig. 8** Footage of the rising from a chair test using Balansera using a belt in the lower back

In parallel with Balansera capture, a therapist evaluated the performance of different SPPB tasks using qualitative scales. Therapist scored manually in a sheet the speed of each rising up and sitting down movement, e.g., slow rising up—fast sitting down and fast rising up—slow sitting down. The therapist also captured incorrect performance by choosing if any/all of the following observations holds during the tests: incorrect up rising (cheating, moving the feet, generating impulse with legs, etc.) or incorrect use of arms (arms moving away from the chest, generating impulse with arms, etc.). The therapist analyzed at the same time of the rise from chair test the sway movement of the individual, assessing the individual has: little sway or much sway.

The full data of the pilot study can be downloaded as a CSV file as well as a SPSS file from <https://github.com/esteban-g/Balansera-mob>.

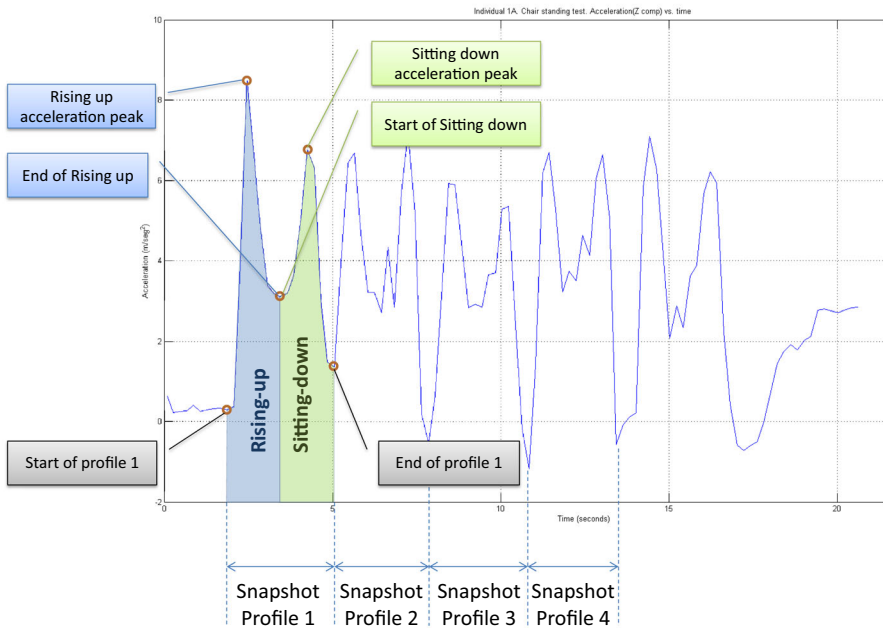
### 4.3 Data analysis

Interpretation of observations were performed in a bottom-up manner, from the capturing of sensor raw data to activity status definition and qualifier evaluation.

In SPPB tests, particularly in the sit-to-stand test we identify a “reference” signal from the raw data, identifying peaks of the filtered signal as is presented in Fig. 9.<sup>9</sup> We called these references as “snapshot profiles” which were obtained by running a peak detection algorithm detecting starting and ending points.<sup>10</sup>

<sup>9</sup> Side-by-side, semi-tandem and tandem data analysis was performed in the same manner to sit-to-stand test, we do not explain details of data analysis for paper length constraints.

<sup>10</sup> A parallel analysis was made using a Matlab script for obtaining reference snapshots. This Matlab script can be downloaded in <https://github.com/esteban-g/Balansera-mob>.



**Fig. 9** Snapshot profiles captured by Balansera for the chair sit-to-stand SPPB test (acceleration vs. time)

In Fig. 10, plots of the acceleration versus time during the test are shown. Data of individuals 1, 11 and 16 among others present disturbances in the first snapshot profile.<sup>11</sup> This lack of evidence was confirmed by the manual annotations of the therapist (see extra data in Balansera Web site <https://github.com/esteban-g/Balansera-mob>). We capture this lack of observable evidence directly using negation as failure *not*, e.g., the observation: *slow\_sitting* ← *not fast\_sitting*, in another words: “there is no evidence that individual has a fast sitting down, then is assumed that individual is sitting down slowly”.

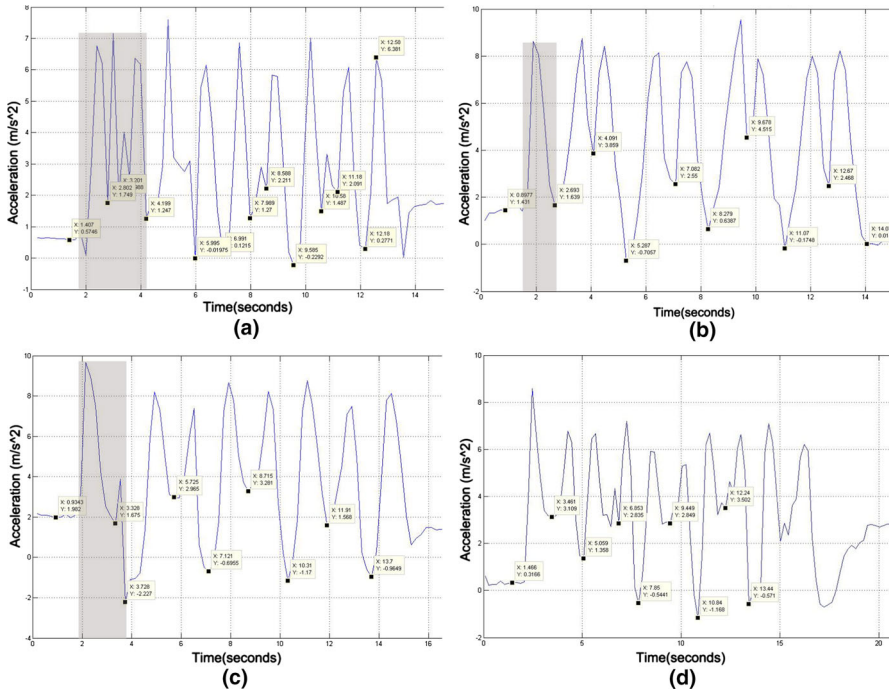
#### 4.4 Results of the demonstration

In this section, we explore how qualifiers can be used in an actual situation. We evaluate effects on qualifiers empirically using data from the demonstration study.

##### 4.4.1 Applicability and utility of qualifiers

In the health domain, an assessment protocols measurement is obtained by applying a standard scale to variables, thus translating direct observations to a numerical scoring system. Roughly speaking, there are four types of qualification measurements they are called: *nominal*, *ordinal*, *interval* and *ratio* scales (see [15] for a further discussion about scales). The SPPB is a ordinal four-battery test in which every subtest is evaluated by a scoring the exercise quality using a range from 0 to 4, being 0 a very slow or unable to execute the task, and a score of 4 if the individual executes well such task. The qualifier introduced in Definition 10 is based on

<sup>11</sup> The complete set of plots acceleration vs. time performed to 20 individuals using Balansera can be found in <https://github.com/esteban-g/Balansera-mob/tree/master/img>.



**Fig. 10** Acceleration versus time plots for the sit-to-stand test of four individuals using Balansera. Peak detection was not able to detect a pattern for individuals 1, 16 and 11. **a** Individual 1, **b** individual 16, **c** individual 11 and **d** individual 4

a similarity function  $\text{Sim}()$ , which can be implemented as an ordinal function for instance calculating the magnitude of the reference and current measurement values:

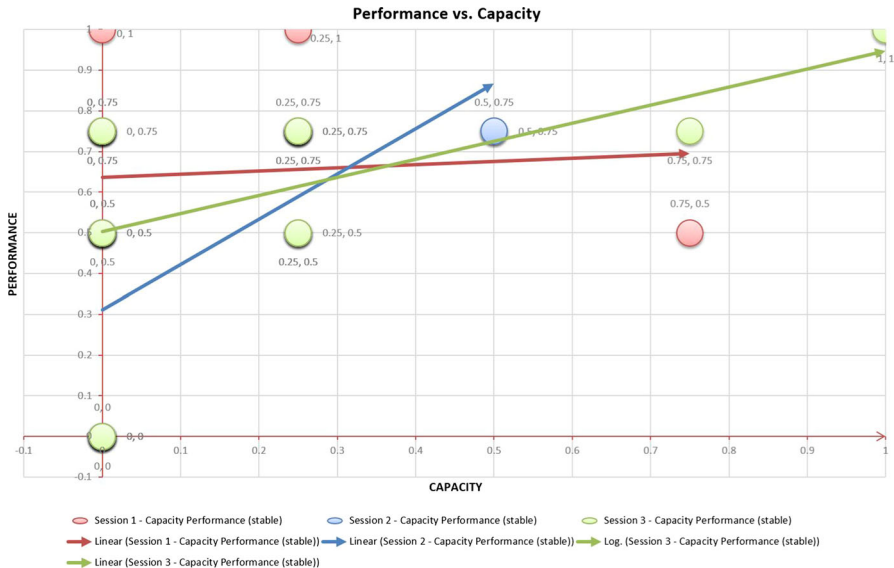
$$Q = K \frac{|\text{Output}_{\text{SEM}}|}{|\text{Concs}(\mathcal{H}\mathcal{F}_{\text{ref}})|} \tag{1}$$

In Eq. 1, the K factor can be defined to a specific range in order to normalize the result. A qualifier using magnitude of sets was proposed in [19] in terms of goals and observations.

Like in standardized pilot tests, qualifiers here presented require that the test administrator<sup>12</sup> has undertaken specific training on the scoring of the test. Qualifiers here presented show more usability when they are combined, in Fig. 11 Performance and Capacity scores for 20 persons in the three sessions evaluated with Balansera. In this figure, stable semantics was used to obtain both qualifiers. Results show that most of the individuals have a larger scores in Performance rather than in Capacity. Scores in Fig. 11 goes from 0 to 1, being 1 the maximum score using Equation 11 ( $K = 1$ ). Trend lines are highlighted showing tendency of the scores from three phases. Even when the test has a limited short period for following same individuals, the combined evaluation of qualifiers w.r.t. a time period, can provide a tool to evaluate the individual’s activity realization. Indeed, the ICF highlights the utility of the combined coding of performance and capacity, as a powerful technique to understand the final effect of the environment on a person, as well as allowing the user opportunities to effect changes to the environment to enhance function [33].

<sup>12</sup> Person who read the results of the test.



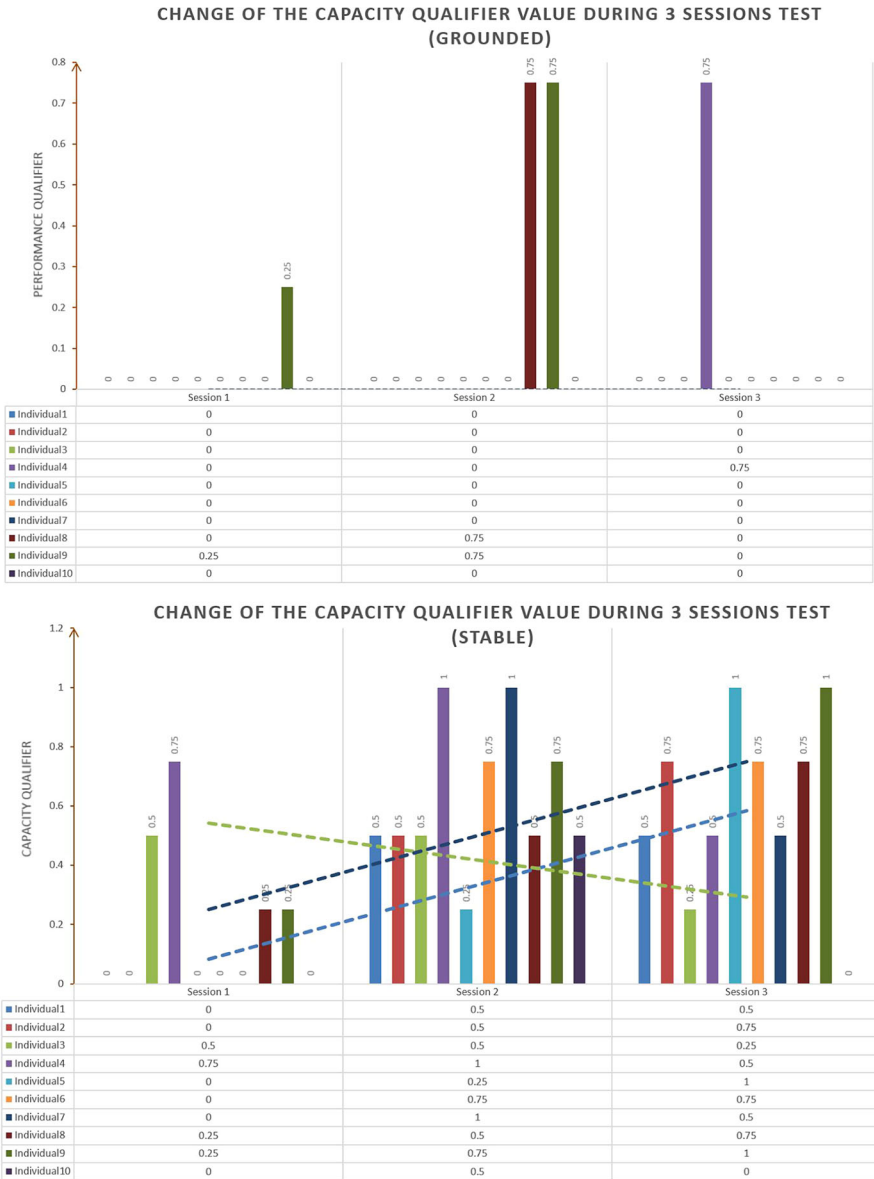


**Fig. 11** Plot of performance versus capacity qualifiers from 20 data measurements in the third SPPB evaluation phase

#### 4.4.2 Qualifiers using different argumentation semantics

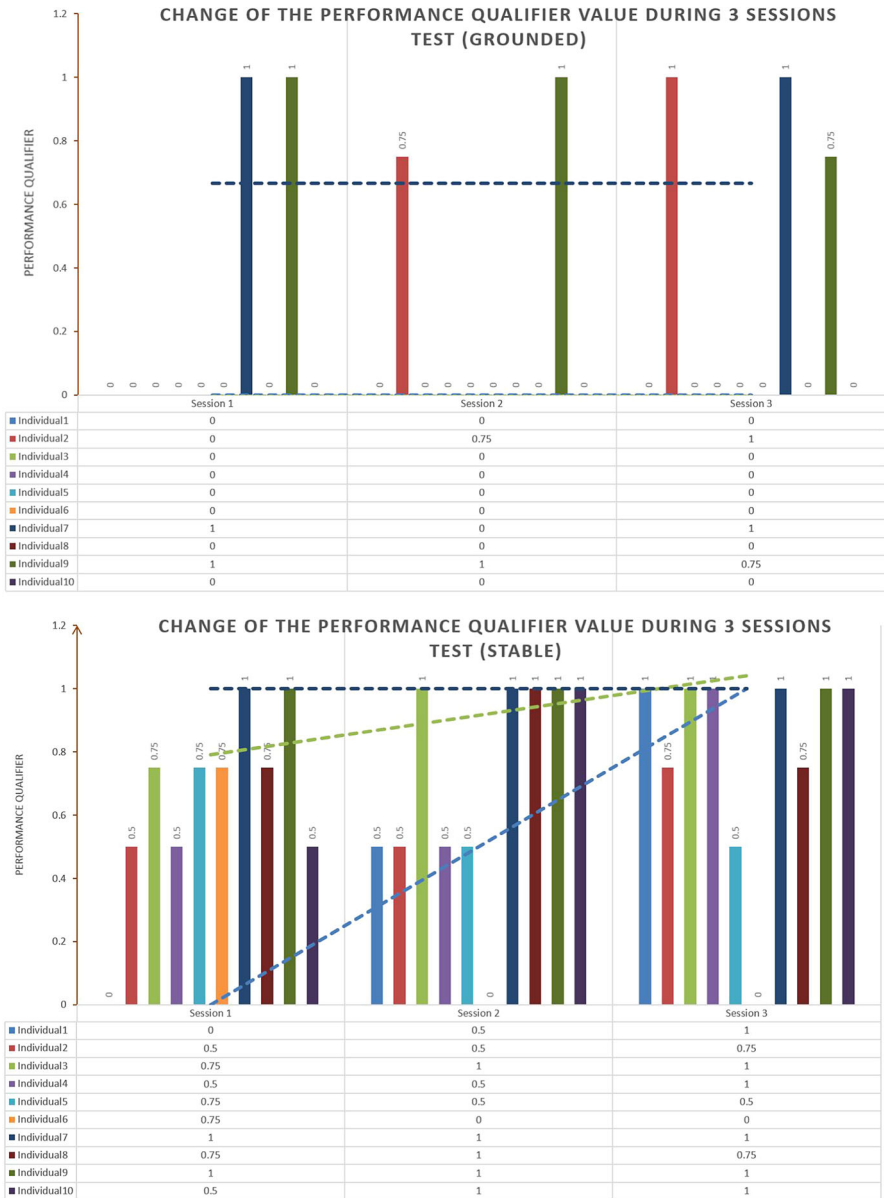
During the pilot experiment, we calculate Performance and Capacity using four argumentation semantics: grounded, complete, stable and preferred. We found exact correspondences between: complete and grounded, stable and preferred. These results show a clear difference between these two semantics groups, the so-called *skeptical* approaches: grounded, complete; and *credulous* semantics: stable and preferred. In Fig. 12 the Capacity qualifier calculated using Eq. 1 using stable and grounded semantics is shown. In this figure, the calculation of capacity using grounded presents results only from those individuals where the semantics suggests without any doubt the current condition, no matter was the level between 0 or 1. On the other hand, the stable version of capacity qualifier presents more information because the qualifier is calculated even under a level of uncertainty regarding the output. In both figures, dashed lines are plotted for individuals 1, 3 and 7 only for purposes of showing trend lines in the evolution of the qualifier during three sessions. These tendencies lines are 0 during all the three sessions in the three cases for grounded semantics. In case of stable semantics, the trend lines show a increasing level in Capacity for individual 1 and 7, but for individual 3 there is a tendency to decrease in the Capacity level.

In Fig. 13 the calculation of Performance qualifier w.r.t. stable and grounded semantics is shown. As in the Capacity qualifier calculation, the two semantics approaches are different regarding to the commitment for drawing a result. In the sit-to-stand test, Performance evaluates if an individual performed the full 4 up and down movements. According to the results of the first session (using a Stable semantics evaluation) the 15% of the individuals achieved the four sit-to-stand exercises (the so-called snapshot profiles) and the half of the group achieved at least 50% of the task. In the same test, Capacity evaluates the ability of executing such a task, in other words, our approach evaluates if the exercise was “correctly” executed or not. Considering the same data, none of the individuals execute the task as the



**Fig. 12** Capacity qualifier using stable and grounded semantics. A subset of 10 measurements from individuals is plotted. Trend dash lines represent the change of the qualifier during the three sessions

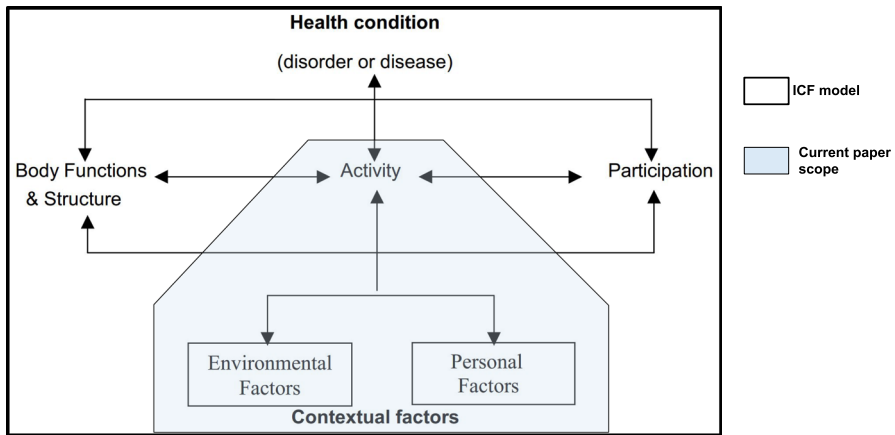
therapist referenced as correct at the same time that 35% of the persons did not succeed in any of the four movements as was expected. Roughly speaking, individuals in the sit-to-stand test are able to execute all the routine achieving the activity goals, but not all of them conducted such activity in an optimal manner.



**Fig. 13** Performance qualifier using stable and grounded semantics. A subset of 10 measurements from individuals is plotted. Trend dash lines represent the change of the qualifier during the three sessions

## 5 Discussion

In this section we compare our activity qualifiers and other related approaches. We discuss our findings considering the use case demonstration in the health domain.



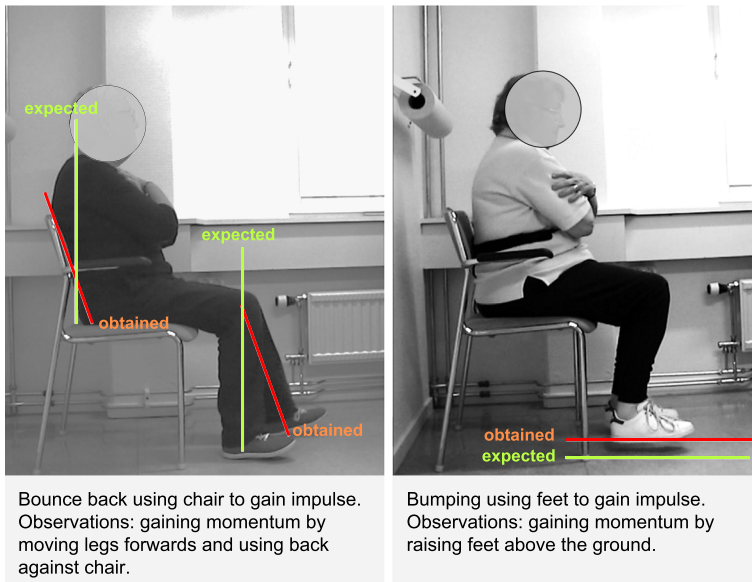
**Fig. 14** Current paper scope with respect to ICF model. Adapted from [34]

## 5.1 Qualifiers in health domain

The International Classification of Functioning, Disability and Health, known more commonly as ICF, provides a standard language and framework for the description of health and health-related states. Generally speaking, ICF assists in scientific research by providing a framework or structure for interdisciplinary research in disability and for making results of research comparable [34]. The approach presented in the current paper, is focused on providing tools for clinicians in the assessment of health condition. In this sense, our approach is centered in activity evaluation which is one of the main focus of ICF. In Fig. 14 is presented the ICF model of disability, providing a broad perspective of our current research and ICF framework.

Generally speaking, performance and capacity qualifiers for a particular health condition of an individual is normally evaluated by using different information sources and tools. Direct observations of the individual by a therapist [21] or asking directly to parents or relatives of an individual about her/his condition [29, 36], are examples of common assessment information sources. There is a main difference between our notion of Capacity and a number of health-oriented investigations, this lies in a restriction of the context where capacity is measured. This qualifier is linked to the ability of a person to execute a task; in order to assess the full ability of the individual, one would need to have a “standardized environment” to neutralize the varying impact of different environments on the ability of the individual [32]. This restriction is not captured in our notion of Capacity (Definition 12) which could eventually change results if our qualifier is measured, for instance, at home and in a controlled environment such as a hospital in the case of the SPPB. We believe that this lack of context restriction w.r.t. context, can be an opportunity to evaluate the ability of a person to execute tasks in real-scenarios (e.g., in a home environment), compared to controlled ones.

We introduce an actuation qualifier (Definition 13) being out of the scope of ICF qualifiers. We hypothesize that this new qualifier can be useful for assessing multiple task achievement or plan completion in an individual. We are aware about current assessment battery tests for evaluating attention, “impulsivity” among others (see [17, 27] for possible scenarios where actuation can be used). A future line of research could be the integration of time in actuation



**Fig. 15** Slight deflections (“cheating”) of the sit-to-stand task which were not detected by Balansera. This deflections of the exercise modify performance and capacity values

qualifier, in this sense, it can be possible to evaluate time planning execution, sustained attention, and other different cognitive/motor abilities of an individual.

Capacity and performance have been compared in a number of investigations [21,36]. If capacity is less than performance, then the person’s current environment has enabled him or her to perform better than what data about capacity would predict: the environment has facilitated performance. On the other hand, if capacity is greater than performance, then some aspect of the environment is a barrier to performance [32]. In our test evaluation, we obtained roughly higher values of perform in comparison with capacity (see Figs. 12 and 13). Regarding this phenomena, one expert physiotherapist stated the hypothesis that some individuals could have “cheated” in the task execution of the SPPB. For example, in the sit-to-stand from a chair task the SPPB clearly states that individuals should hold their arms crossed in the chest, avoiding bumping or bounce back when sitting. This was confirmed by re-checking videos that were recorded in the moment of the test (see Fig. 15). Balansera, using only one sensor was unable to detect such slight deflections of the exercise which can help an individual to perform the task. In a future work, we want to use external sensors of Balansera to obtain different aspects of the body movement.

## 5.2 Qualifiers based on structured fragments

Two main steps in the information processing can be highlighted for any argument-based system, first the process of building arguments and second, the selection of an output using a given argumentation semantics (Phase 2 in Fig 1). The first step is almost overlooked in abstract argumentation literature [7,11,13,23,31]. Our approach is based on the transference of a human activity hierarchical structure to the structure of an argument, the so-called *fragment* of an activity. Moreover, the process for building a fragment using Definition 3 is also novel, considering that most of the processes for building argument-based structures

are based on tree-like proof procedures [2,12,26,35] among others. In Definition 3, the step 3 provides a method for evaluating the consistency of the fragments support. Roughly speaking, the WFS() function in step 3 “filters” possible inconsistent rules for instance, “if there is not evidence that individual is walking, then is walking”: *walking*  $\leftarrow$  *not walking*. Some argument-based approaches can not handle these type of inconsistencies when structured arguments are built (see [10,14,18] for a discussion about argument-based systems consistency principles).

In *abstract argumentation* [7,11,23], each argument is regarded as atomic, i.e., there is no internal structure to an argument. In this sense, fragments following Definition 3 are *structured* arguments in which the set  $\{S, O', h\}$  represent the support (*premises*) for the conclusion  $g$ . Fragments built under Definition 3 entail a relationship between support and conclusion, for instance providing support evidence (observations and the execution of an action) for inferring that a goal  $g$  has been achieved.

An activity framework (Definition 2) encloses information about the human activity defined in AT and the behavior of an intelligent agent evaluating the activity. Notice that an activity framework defines the knowledge structure of an agent in terms of Activity Theory, using goals, actions and observations of operations. Describing the behavior of an intelligent agent by considering a human-like perspective is not new, other approaches such as the Belief-Desire-Intention (BDI) model [8] is a well-suited approach for describing intelligent agents state. In contrast to BDI, an activity framework addresses the problem of dealing with low-level operations considering them with an unconscious grade, being partially disregarded in the planning or recognition of high-level goals.

A fragment based on a semantic construction as is presented in Definition 3 and based on [18] provides an extra advantage to evaluate pieces of an activity, namely conclusions are drawn from related rules. This approach for building fragments explaining an activity endows to our approach, a method for drawing conclusions from the knowledge related to a particular goal. In other words, even when multiple goal-based activities are performed at the same time, fragments gathers information from that part of the knowledge base that is related.

### 5.3 The role of argumentation semantics in qualifier calculus

In the calculus of qualifiers, argumentation semantics have a major impact. Grounded semantics among the Dung’s semantics has a skeptical behavior [11], it belongs to the unique-status approach, i.e.,  $|\text{SEM}(\text{AAF})| = 1$  and it prescribes a maximal number of undefined fragments among Dung’s semantics. By considering a most *credulous* approach such as stable, we can obtain a contrary behavior than grounded, minimizing the number of fragments without a justified state. Stable belongs to the multiple-status approach, i.e.,  $|\text{SEM}(\text{AAF})| > 1$  being not universally defined, may in particular be the case that  $|\text{SEM}(\text{AAF})| = \emptyset$ . This last result presents drawbacks for the calculation of a qualifier if the  $\text{Sim}()$  function is implemented as a quantification of magnitude such as the case of Eq. 1. In this case, an inexistent value for the current evaluation of the activity will give an unexpected result. Note that the case of non-existence of extensions, namely  $\text{SEM}(\text{AAF}) = \emptyset$ , is significantly different from the case  $\text{SEM}(\text{AAF}) = \{\emptyset\}$ , where the semantics prescribes exactly one (actually empty) extension.

In the argumentation literature, various kinds of motivations have been used to support the introduction of new semantics with respect to Dung’s proposals [11]: stable, grounded, complete and preferred semantics. These motivations range from the desire to formalize some high-level intuition, not captured by other proposals, to the need to achieve the “correct” treatment of a particular example (or family of examples), regarded as particularly signif-

icant. For instance, CF2 semantics [3,6] has originally been conceived to deal with some problematic behaviors of preferred semantics when dealing with odd-length cycles. Semi-stable semantics [9] aims at avoiding the problem of non-existence of extensions affecting stable semantics in some cases, while preserving its behavior when stable extensions exist as well as the related property of minimizing undecided arguments [4]. We evaluated the notion of qualifier w.r.t. classical semantics, considering two groups skeptical and credulous ones, given that our pilot results suggest at least the following application scenarios:

- Credulous semantics usage: qualifiers using a credulous approach can be used to evaluate activity execution even under not rigorous achievement of goals or completion of actions or observations. A follow-up procedure over time, as is presented in Fig. 12 with stable semantics, was considered interesting by therapists.
- Skeptic semantics usage: our results using grounded semantics showed that the output of our system is sensitive to lack of observations information. Therapists consider this approach as a *screening* procedure, where individuals are a systematic assessment is done especially to detect anomalies with a minimal of ambiguity.

## 6 Conclusions

In this paper, we explore a novel approach based on argumentation theory for quantifying the evaluation of physical activities from the perspective of rational agents, by resembling the kind of assessment reasoning performed by clinicians: (1) gathering data through observations; (2) handling ambiguous and uncertain observation information; (3) generating current function status hypothesis; (4) deduce an explanatory outcome of explanation; and (5) retracting the explanation under new evidence.

We propose a general approach of activity qualifier notion using argumentation theory. Our contributions can be summarized as follows:

- We made a further extension of an argumentation-based activity framework [19,25] in order to reason about an activity model AT based on goals, actions and observations of a hierarchical activity.
- Qualifiers depend on the output of a non-monotonic reasoning (NMR) process. In this sense, when more information is added to the activity context, the output of the argumentation process will change. This NMR procedure follows a health-care perspective where a physician can change an assessment of an individual when more information about her/him is available.
- Three different instantiations of the general qualifier assessing three types of activity elements: goals, evaluated by Performance (Definition 11); actions, calculated by the Actuation qualifier (Definition 13); and observations, using the Capacity qualifier (Definition 12).
- We demonstrate our approach by using data from a pilot study. Our findings can be summarized as follows:
  - A combined evaluation of qualifiers, Capacity and Performance can be an useful tool for expert analysis of activity execution, particularly in a process of *follow-up* keeping track the individual's activity behavior and in *screening* methods for systematic detection of anomalies with a minimal of ambiguity.
  - There is a considerable difference between the values of qualifiers when different argumentation semantics are considered. We evaluate qualifiers considering “classic”

[11] semantics: stable, preferred, complete and grounded. We found that for the SPPB test, stable and preferred (credulous), complete and grounded (skeptical) semantics behave in the same manner.

- Our pilot experiment can not be considered as conclusive for assessing an individual's physical condition; however, it helped therapists in the awareness of different sensor-based advantages and restrictions.

As a future work, further differences between argumentation semantics for calculating qualifiers will be considered, for instance different criteria for semantics evaluation [4]. In our future plans, we consider establishing Balansera as an open platform for evaluating Physical Exercise in high performance athletes in North of Sweden.

**Acknowledgements** The authors are very grateful to anonymous referees for their useful comments. We thank to all the participants in the user studies and to Marianne Silfverskiöld who conducted the case study. Silfverskiöld's study was approved by the ethical committee (2014/113-31Ö).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Aggarwal JK, Ryoo MS (2011) Human activity analysis: a review. *ACM Comput Surv (CSUR)* 43(3):16
2. Apt KR, Blair HA, Walker A (1988) Towards a theory of declarative knowledge. In: Minker K (ed) *Foundations of deductive databases and logic programming*. Morgan Kaufmann, Los Altos, pp 89–148
3. Baroni P, Giacomin M (2003) Solving semantic problems with odd-length cycles in argumentation. In: *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, 7th European Conference, ECSQARU 2003, Aalborg, Denmark, 2–5 July 2003. *Proceedings*, pp 440–451
4. Baroni P, Giacomin M (2007) On principle-based evaluation of extension-based argumentation semantics. *Artif Intell* 171(10–15):157–168
5. Baroni P, Giacomin M (2009) Skepticism relations for comparing argumentation semantics. *Int J Approx Reason* 50(6):854–866
6. Baroni P, Giacomin M, Guida G (2005) Scc-recursiveness: a general schema for argumentation semantics. *Artif Intell* 168(1–2):162–210
7. Bondarenko A, Dung PM, Kowalski RA, Toni F (1997) An abstract, argumentation-theoretic approach to default reasoning. *Artif Intell* 93(1):63–101
8. Bratman M (1987) *Intention, plans, and practical reason*. Harvard University Press, Cambridge
9. Caminada M (2006) Semi-stable semantics. In: *1st International Conference on Computational Models of Argument (COMMA)*, vol 144. IOS Press, Amsterdam, pp 121–130
10. Caminada M, Amgoud L (2007) On the evaluation of argumentation formalisms. *Artif Intell* 171(5):286–310
11. Dung PM (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif Intell* 77(2):321–357
12. Dung PM, Kowalski RA, Toni F (2006) Dialectic proof procedures for assumption-based, admissible argumentation. *Artif Intell* 170(2):114–159
13. Dung PM, Kowalski RA, Toni F (2009) Assumption-based argumentation. In: Simari GR, Rahwan (eds) *Argumentation in artificial intelligence*. Springer, Berlin, pp 199–218
14. Dung PM, Thang PM (2014) Closure and consistency in logic-associated argumentation. *J Artif Intell Res* 49:79–109
15. Fawcett AL (2013) *Principles of assessment and outcome measurement for occupational therapists and physiotherapists: theory, skills and application*. Wiley, New York
16. Gelfond M, Lifschitz V (1991) Classical negation in logic programs and disjunctive databases. *New Gener Comput* 9(3–4):365–385
17. Gomez P, Ratcliff R, Perea M (2007) A model of the go/no-go task. *J Exp Psychol Gen* 136(3):389



18. Guerrero E, Nieves JC, Lindgren H (2015) Semantic-based construction of arguments: an answer set programming approach. *Int J Approx Reason* 64:54–74
19. Guerrero E, Nieves JC, Sandlund M, Lindgren H (2016) Activity qualifiers in an argumentation framework as instruments for agents when evaluating human activity. Springer, Cham, pp 133–144
20. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, Scherr PA, Wallace RB (1994) A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol* 49(2):M85–M94
21. Holsbeeke L, Ketelaar M, Schoemaker MM, Gorter JW (2009) Capacity, capability, and performance: Different constructs or three of a kind? *Arch Phys Med Rehabil* 90(5):849–855
22. Kaptelinin V, Nardi BA (2006) Activity theory in interaction design. Acting with technology, chapter I. The MIT Press, London, pp 1–117
23. Kowalski RA, Toni F (1996) Abstract argumentation. *Artif Intell Law* 4(3–4):275–296
24. Leont'ev AN (1978) Activity, consciousness, and personality. Prentice-Hall, Englewood Cliffs, NJ
25. Nieves JC, Guerrero E, Lindgren H (2013) Reasoning about human activities: an argumentative approach. In: Twelfth Scandinavian Conference on Artificial Intelligence, SCAI 2013, Aalborg, Denmark, 20–22 Nov 2013, pp 195–204
26. Prakken H (2010) An abstract framework for argumentation with structured arguments. *Argum Comput* 1(2):93–124
27. Rubia K, Smith A, Taylor E (2007) Performance of children with attention deficit hyperactivity disorder (adhd) on a test battery of impulsiveness. *Child Neuropsychol* 13(3):276–304
28. Simari GR, Rahwan I (eds) (2009) Argumentation in artificial intelligence. Springer, Berlin
29. Tieman BL, Palisano RJ, Gracely EJ, Rosenbaum PL (2004) Gross motor capability and performance of mobility in children with cerebral palsy: a comparison across home, school, and outdoors/community settings. *Phys Ther* 84(5):419–429
30. Van Gelder A, Ross KA, Schlipf JS (1991) The well-founded semantics for general logic programs. *J ACM* 38(3):619–649
31. Vreeswijk GA (1997) Abstract argumentation systems. *Artif Intell* 90(1):225–279
32. World Health Organization (2001) The international classification of functioning, disability and health (ICF)
33. World Health Organization (2013) How to use the ICF: a practical manual for using the International Classification of Functioning, Disability and Health (ICF). WHO, Geneva
34. World Health Organization et al (2002) Towards a common language for functioning, disability and health: ICF. World Health Organisation, Geneva
35. Wu Y, Caminada M, Gabbay DM (2009) Complete extensions in argumentation coincide with 3-valued stable models in logic programming. *Studia Logica* 93(2–3):383–403
36. Young NL, Williams JI, Yoshida KK, Bombardier C, Wright JG (1996) The context of measuring disability: Does it matter whether capability or performance is measured? *J Clin Epidemiol* 49(10):1097–1101



**Esteban Guerrero** His main interests are framed on artificial intelligence (AI) with a special focus on knowledge representation, and non-monotonic reasoning of rational software agents using formal argumentation theory and logic programming (answer-set programming). Formal contributions have been evaluated/tested in different application domains: sports science, health care, ambient intelligence projects (mobile sensor-based and agent-based systems) and e-learning. Currently, Dr. Guerrero leads research projects applying AI techniques (multiagents systems and machine learning) to improve performance of cross-country skiing elite athletes in Northern Sweden. He holds a B.Eng. in electronics and telecommunications from University of Cauca (Colombia), M.Sc. in computing science from Malmö University (Sweden), and he received his Ph.D. degree in computing science from Umeå University (Sweden).



**Juan Carlos Nieves** is a senior researcher at Department of Computing Science, Umeå University. He received the Diploma of Advanced Studies in artificial intelligence and Ph.D. degree in artificial intelligence from the Universitat Politècnica de Catalunya-BarcelonaTech, Spain, in 2006 and 2008, respectively. He has a wide experience in knowledge base systems with different application domains: Internet of Things, medical diagnosis, smart grid and wastewater management. He has taken part of different research projects granted by the Spanish council, Swedish Organizations and the European Union. His research domain is knowledge representation and reasoning, mainly using web semantics technology, argumentation theory agreement technologies) and answer set programming (declarative logic programming approach for capturing knowledge bases). He has authored and coauthored more than 100 peer review papers in workshops, conferences and journals, and chapter books.



**Dr. Marlene Sandlund** is an assistant lecturer (biträdande lektor) at the Department of Community Medicine and Rehabilitation, Physiotherapy division, Umeå University, Sweden. Main interests include using technology for training and improving ability.



**Helena Lindgren** is Associate Professor in Computer Science and has a professional background as occupational therapist. Lindgren leads research projects in the areas of human agent collaboration, personalization, persuasive and assistive technology, decision-support systems, and on co-design of such systems. Lindgren serves as member of the boards of Swedish AI Society and Swedish Association for Medical Informatics. She is also Sweden's representative in International Medical and Health Association.