

SLACTIONS 2013: Research conference on virtual worlds - Learning with simulations

## The Taxonomy of Goal-oriented Actions in Virtual Training Environments

Ali Fardinpour\*, Torsten Reiners

*School of Information Systems, Curtin University, Building 402, Level 7  
Bentley, Perth, 6152, Australia*

---

### Abstract

With the shift of training scenarios to virtual worlds and assessment being an inevitable part of any teaching and learning process, we require sophisticated assessment methods to analyze action-sequences of learners according to reference solutions defined by experts and provide automated formative feedback. We propose the ‘Action-based Learning Assessment Method’ (ALAM) using an action taxonomy to classify recognized actions performed by the user in the virtual world. Most of these taxonomies were developed to model the behavior and performance of users. Yet, current taxonomies of human actions were developed based on need in specific research, still lacking a general taxonomy. The taxonomy of goal-oriented actions in virtual training environments was developed to overcome this problem and will be discussed in this paper.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Selection and peer-review under responsibility of the University of Trás-os-Montes e Alto Douro (UTAD)

*Keywords: Taxonomy; Human actions; Action-based Learning Assessment*

---

### 1. Introduction

Human actions were the center of attention in different fields for long time. Modeling and recognizing human behavior, performance, tasks, and skills were subject of research in various areas such as psychology, sociology, education, and computer science. In the 50ies and 60ies early taxonomies of human actions were proposed by

---

\* Corresponding author. Tel.: +61-450-041-360; fax: +61-892-663-076.  
E-mail address: [AFardinpour@gmail.com](mailto:AFardinpour@gmail.com), [t.reiners@curtin.edu.au](mailto:t.reiners@curtin.edu.au)

researchers like G.E.M. Anscombe, D. Davidson, D.S. Shwayder, and E. D'Arcy, later being used as a foundation for extended taxonomies to classify human actions; e.g., Goldman [1], Fleishman [2], or Robertson [3] [4].

With the introduction of 3D virtual worlds, taxonomy of embodied actions started to appear in computer science literature. Computer-supported cooperative work (CSCW), behavior modeling in crowds, and avatars' non-verbal interactions used taxonomy of human actions in their research. Learning-by-Doing or Action-based Learning is a valuable methodology for educators and researchers in education, and refers to "all learning that is orchestrated by some activity on the part of learners" [5]. Adopting this learning method for use in virtual learning environments enables us to assess students' knowledge based on their performed actions. Even though avatars perform the actions, we still require the taxonomy of human actions as the avatars just reflect the human actions within the virtual training environment (VTE).

In this contribution, we propose taxonomy of actions to be used it with our proposed assessment method in virtual training environments. Two main characteristics of this taxonomy are openness and adaptability that enables other researchers to adopt the taxonomy to their own environments, such as games and virtual worlds, and add new classes and actions. We continue with a review of relevant literature and briefly introduce the Action-based Learning Assessment Method (ALAM). Following that, the taxonomy of actions in virtual training environments will be defined and demonstrated on an example scenario. We depict the output of a so-called Action Recognition Agent's to demonstrate the taxonomy and its application. Finally, we conclude the paper with a summary and highlight the importance of the taxonomy for future research.

## 2. Background

The review of literature of human actions and different proposed taxonomies shows the need for the taxonomy of actions in virtual training environments and in general, virtual worlds. The development of the new learning assessment method 'Action-based Learning Assessment method (ALAM)' indicates the need for the taxonomy of actions in VTE [6].

Before we continue, we have to create an understanding of the term "action". We adapted the definition of action for agents in artificial intelligence proposed by Allen [7]: An action is an occurrence caused in a 'certain' way by the 'Avatar'. We are using this definition for ALAM; yet require for the assessment a finer granularity to classify and compare actions.

The review of literature covering taxonomies of human's actions, embodied actions, actions in virtual worlds, and behavior modeling shows that most researchers in different disciplines create their own taxonomy of actions or behaviors based on their project's needs. We further discovered that human performance, task, behavior, and action in the literature sometimes were used interchangeable.

Fleishman [2] identifies seven categories of human performances in the literature: identification, discrimination, sequence learning, motor skill, scanning, and problem solving. Fleishman states that it is important to be clear about the motivation for task classification because those who create these classifications use them more as a tool in their research. He stated "We can elect to develop a system of classification having utility for a limited area (e.g., the classification of tasks with respect to which certain training methods are found most effective in promoting high levels of task performance), or we may look for a system from which a variety of applications may stem." (p.1128) [2].

Goldman [1] proposes a theory of human action. He discusses individuation, act-type, act-token, basic and non-basic actions. Goldman also discusses and defines the structure of actions and level-generation in his book. His theory of human actions further discusses intentional actions and wants. He criticizes the way of other researchers (G. E. M. Anscombe, Donald Davidson, D. S. Shwayder, Eric D'Arcy) [8, 9, 10, 11, 12] of using the term 'identity thesis' in the 50ies and 60ies. He disagree with the assumption that two different acts like 'moving right arm' and 'move queen to king-knight-seven' can be both recognize as basic actions based on the same identity thesis. He stated, "Moving my hand is a basic action, whereas checkmating my opponent and turning on the light are not basic actions. Rather, they are actions I perform by performing some basic actions".

Computer-supported cooperative work (CSCW) is one of the disciplines using virtual collaborative environments and following the embodied actions. Robertson [3, 4] creates the taxonomy of embodied actions for cooperative design in a distributed company. Like other researchers, he developed this taxonomy according to his need in his

research as he stated, "the taxonomy presented in this paper was developed as a possible bridging structure between the field study of cooperative work in practice and the design of technology that might support that work over distance". Robertson defined the categories open and flexible so people can adopt it to their own practice. The taxonomy divides the embodied actions into individual embodied actions and group activities constituted by individual embodied actions. Individual embodied actions are in relation to physical objects, other bodies, and the physical workspace. In relation to physical objects, Robertson classifies embodied actions into moving physical objects, producing a private physical representation, highlighting some aspects of an object, and the personal use of a physical object. In relation to other bodies, the classification includes emitting signs and monitoring signs, and pretending to be another body. Finally, in relation to the physical workspace, moving around, pointing at something, shifting direction of gaze, and moving in or out of the shared space are different classes of embodied actions. Group activities constituted by individual embodied actions are shaped by: conversing, looking at the same thing at the same time, organizing shared communication resources, creating a shared representation, shared physical use of an object, focusing group attention, breaking into smaller groups and reforming, seizing the moment, and doing something else. Cappella and Pelachaud [13] studied human behavior to build virtual interactions based on human interactions. Their archive includes 100 interactions, including "same sex and opposite sex pairs, dyads with longer histories (greater than six months as friends) and strangers, partners with similar and different attitudes, and expressive and reticent pairs". They code the interactions and use it in a system to analyze behavior. These studied and coded interactions include vocalic behaviors, eye gaze, smiles and laughter, head nods, back channels, posture, illustrator gestures, and adaptor gestures.

Verhulsdonck [14] believes that "a common framework for non-verbal behaviors in virtual worlds must include both rhetorical acts (actions of choice), as well as those that are procedurally driven by the utterances or the psychological state of the avatar. We argue that any developing standards should be open enough to allow for such evolution". (p. 8).

Mehrabain [15] classifies non-verbal interactions as follows:

- Oculistics: eye gaze, eye contact
- Deictics: Pointing
- Gesticulation: Hand and arm gesturing
- Proxemics: Body distance
- Chronemics: Time between interactions

In his literature review, Chodos et al. [16] reflects a classification of recorded action data in virtual worlds, which was proposed by Hurst in 2011. Hurst classified recorded data into (1) reflective data, (2) machinima and (3) virtual-environment data. The last category was identified as the most relevant, covering the recording of the avatars' in-world actions and their interactions with objects. Unfortunately, her paper was not published and the research was abandoned.

Chodos et al. [16] developed the Avatar Capabilities Model (ACM), which categorizes student's avatar actions in three pedagogically based themes: movement, experiencing the world, social interaction. This classification is shown in Table 1.

Table 1. Definition of Avatar Capabilities Model [16]

<b>Action</b> = < <b>Movement</b>   <b>Sensing</b>   <b>Object Manipulation</b>   <b>Communication</b> >
<b>Movement</b> = < <i>Move</i>   <i>Sit</i> >
<i>Move</i> = < Actor, Movement Type, Start Location, End Location >
<i>Sit/Stand</i> = < Actor, Sit/Stand Type, Sit/Stand Location >
<b>Sensing</b> = < Actor, Modality, Target >
<b>Object Manipulation</b> = < <i>Create</i>   <i>Hold</i>   <i>Transfer</i>   <i>Take</i>   <i>Interact</i> >
<i>Create</i> = < Actor, Created Entity >
<i>Hold</i> = < Actor, Held Entity >
<i>Transfer</i> = < Actor, Target, Transferred Entity >
<i>Take</i> = < Actor, Taken Entity >
<i>Interact</i> = < Actor, Entity, Message, Options, Choice, Response >
<b>Communication</b> = < <i>Speak</i>   <i>Write</i>   <i>Gesture</i> >

---

*Speak* = < Actor, Message >  
*Write* = < Actor, Message >  
*Gesture* = < Actor, Communication Type, Description >

---

The review of literature can be summarized in three main points:

1. Almost every researcher develops an individual taxonomy based on and limited to his or her project needs. The taxonomy is used as a tool.
2. Researchers develop their taxonomies to be open and flexible, such that these can be adapted by others.
3. All covered actions are in relation to other avatars, their environment and objects.

### 3. Action-based Learning Assessment Method (ALAM)

Action-based Learning Assessment Method (ALAM) is a new formative assessment method, developed to assess learners' goal-oriented actions and action-sequences, reflected by avatars in 3D virtual environments, and provide them with formative feedback to enable learners to learn from their assessment [6]. Assessment of action choices is used in educational games and virtual training systems for summative and formative assessment of memorized knowledge and application of the learnt knowledge. ALAM creates the opportunity of analyzing and assessing of how learners do things and not just what they do. The main difference of ALAM to other assessment methods that involve the activities of learners is that ALAM does not restrict learner with predefined action choices like educational games. Learners perform the full operation and monitor the consequences of their actions. Based on the performed actions and their sequences formative feedback is generated to report the correctness of learner's performance, learner's possible mistakes; always in relation to the best solution.

Alongside with experts, users interact with the VTE using different peripherals and goggles, performing a sequence of actions to achieve a predefined goal, namely the goal act. All logged actions are analyzed for actions, which are then checked for their relevance and belonging to action-sequences. Then, the compiled user's data will be compared to the experts' compiled data in terms of correctness and relevance of actions and action-sequences; based on this comparison and evaluation, a formative feedback and assessment score is generated and submitted to the user [6].

ALAM uses Rogers' 5-stage feedback classification [21], which is still valid and commonly used in assessing students learning outcome [17, 18]. Human markers are capable to provide feedback on all stages. Yet it is far more common to simplify (mainly concerning the workload) the process by designing multiple-choice or short answer assessment. Especially as formative feedback at Stage 4 or 5 requires experts, understanding if the answer of the student is valid with respect to the scope and body of knowledge, and if not, exploring the train of thoughts that lead to the given answer. Automating Stage 1 and 2 is relatively easy and often done. The other stages have a higher complexity as it requires understanding of the problem, the context, and often natural language, tasks that cannot yet be done automatically by intelligent assessment algorithms [19]. The complexity is reduced by specifying constraints to reduce the problem and solution space.

ALAM is about actions and action-sequences that lead from the initial state of the environment to a final state where the given problem is solved. For each change of state, we record the actions and action-sequences being executed by the learner; providing us with a complete protocol (sequence of actions representing the solution for a problem) of how the learning objectives were achieved. The learner action-sequences are compared to the expected action-sequences recorded by experts or instructional designers. It is not essential to have a complete match, as the solution to a problem might not be unique. Both sequences are compared and verified based on milestones that are needed to find a solution.

The learner has the opportunity to proceed from one milestone to the next without being constrained in between; yet milestones and their order might be crucial. The restricted scope allows us to implement an immediate formative feedback, being triggered when the learner hits a milestone.

#### 4. The Taxonomy of Actions for Action-based Learning Assessment in Virtual Training Environments

Virtual worlds include a variety of actions avatars use, to express themselves and interact with other avatars, objects and the environment itself. Not all these actions are used in VTEs, especially when the assessment and training systems are Unidirectional, avatars may not interact with other avatars or humans in real-time and these interactions are part of the assessment. Although, if the assessment and training environment is collaborative, these interactions can be adapted and that is why the current defined classes of actions in the taxonomy are totally open and flexible in how people combine them in practice. Relations in taxonomy can be hierarchical, associative, and/or equivalent.

##### 4.1. Classes of Actions in Virtual Training Environments

Every goal-oriented action in a VTE belongs to one of the following classes:

The Goal Act

- Constitutive Act
- Functional Act
  - Gestural Actions
  - Responsive Actions
  - Decisional Actions
  - Constructional Actions
  - Operative Actions
  - Locomotive Actions

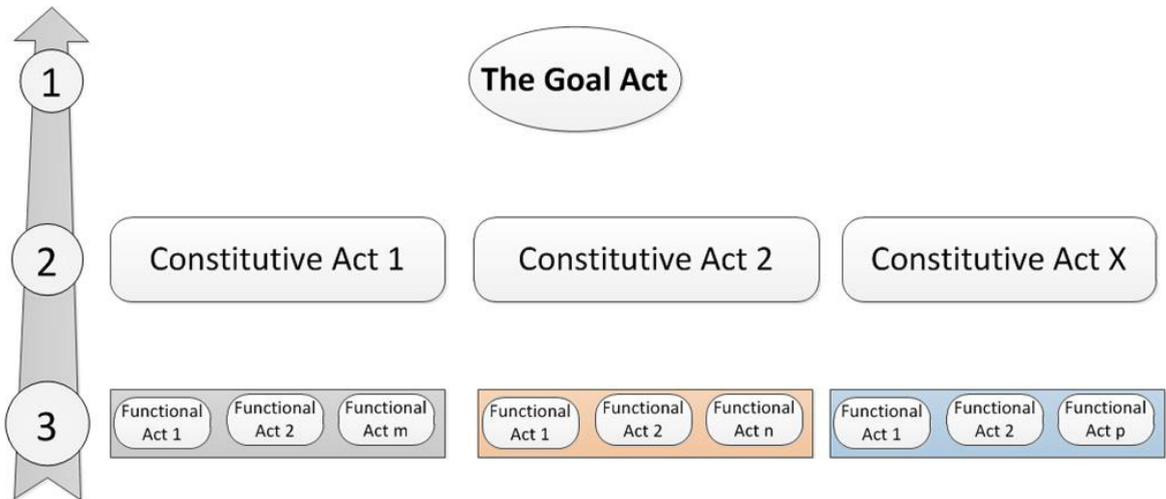


Fig 1. The Taxonomy of Actions for Action-based Learning Assessment in Virtual Training Environments

To achieve the Goal Act (specifically defined goal in the examination scenario), the learner needs to perform a sequence of Constitutive Acts, which are composed of Functional Acts. Constitutive Acts resemble the milestones used by Reiners et al. [20] for their narratives in virtual worlds; yet the how is not as important as the achievement of the milestone itself. In contrary, ALAM focuses on actions performed correctly to achieve these milestones or Constitutive Acts and as a result, the Goal Act.

## 4.2. Definitions and Example scenario

In first section, we introduced different levels and classes of actions in the taxonomy for Action-based Learning Assessment in virtual training environments. Each definition follows by some example actions. Next section demonstrates an example scenario, supporting copper rebar in lathe machine's chuck, and highlights the actions and their classes and briefly introduces the syntax ALAM uses to encode these actions and their sequences.

### 4.2.1. Definitions

To assess users' knowledge, experts create different problems and scenarios to be solved by the user. To solve this problem and achieve this goal, users need to fulfill some milestones correctly, and these milestones are formed by different actions. Following lines will define these goals, milestones, and actions.

**The Goal Act:** The Goal Act is the highest level of action in VTE, can be complex and/or composite, which is a specific goal to achieve by the user. The Goal Act is formed of one or more Constitutive Acts. Like, fixing the rotating shaft or doing a heart surgery.

**Constitutive Act:** to achieve the Goal Act in VTE, user needs to perform a sequence of high-level actions called Constitutive Acts; these high-level actions are composed of other low-level actions named Functional Acts. The objective of Constitutive Acts is to achieve the Goal Act.

**Functional Act:** They are the lowest level of actions in VTE, which enables avatars to act within VTE. Objective of Functional Acts is to form Constitutive Acts.

Functional Acts are classified in six action classes: Gestural, Responsive, Decisional, Operative, Constructional, and Locomotive, defined as follows:

- *Gestural actions* are movements in the avatar's body and/or face expressing different meanings, and communicating particular messages, a variety of feelings and thoughts, from contempt and hostility to approval and affection.
- *Responsive actions* are responses triggered by changes in the environment or objects; like pushing the button when the green light comes on or taking your hand back by touching the hot metal.
- *Decisional actions* enable avatars to reflect their decisions by choosing between different options; like choosing between left or right, up or down, yes or no, quantity, numbers, etc.
- *Operative actions* are simple basic acts enabling avatars to function in VTE by executing different non-constructive actions, like push, collect, grab, etc.
- *Constructional actions* are simple fundamental manipulative acts allowing avatars to impact on their environment as well as its objects, like cutting, screwing, etc.
- *Locomotive actions* are empowering avatars to move around or teleport to different parts of the virtual environment to execute their tasks; such as walk, run, fly, teleport, etc.

### 4.2.2. Example Scenario (Supporting Copper rebar in lathe machine's chuck)

As an example scenario, we used 'Supporting Copper rebar in lathe machine chuck' (The Goal Act). To support the rebar in lathe machine user has to:

1. Enter the shop and collect safety equipment and clothing (Constitutive Act).
2. Choosing and sizing the cooper rebar (Constitutive Act).
3. Put the rebar in chuck and fix the tailstock (Constitutive Act).

These three Constitutive Acts or milestones are composed of different Functional Acts in a certain sequence of actions. User enters (Locomotive) the virtual shop, goes to safety room (Locomotive) and collects safety equipment and clothing (Operative); then moves to machine shop (Locomotive), chooses a cooper rebar with diameter of 0.4 inch (Decisional) and cut two pieces of rebar (Constructional) in sizes 1.5 and 0.5 inch (Decisional). While cutting the rebar, user places his hand too close to the saw blade so by feeling the blade near his fingers he takes his hand away (Responsive) very fast to avoid hurting himself. User opens the chuck (Operative) puts the rebar (Operative) in it and supports the rebar by turning the chuck wrench (Operative) to the right (Decisional). User then pushes the tailstock (Operative), puts the barrel's center to the end of rebar (Operative), and tightens it (Operative). User checks

the rebar between the chuck and the center by shaking it (Operative) and makes sure it is tight enough (Decisional) and nodes (Gestural) to the operator to turn on the lathe. Fig. 2 shows these three Constitutive Acts and the sequence of Functional Actions needed to be performed to achieve the Goal Act.

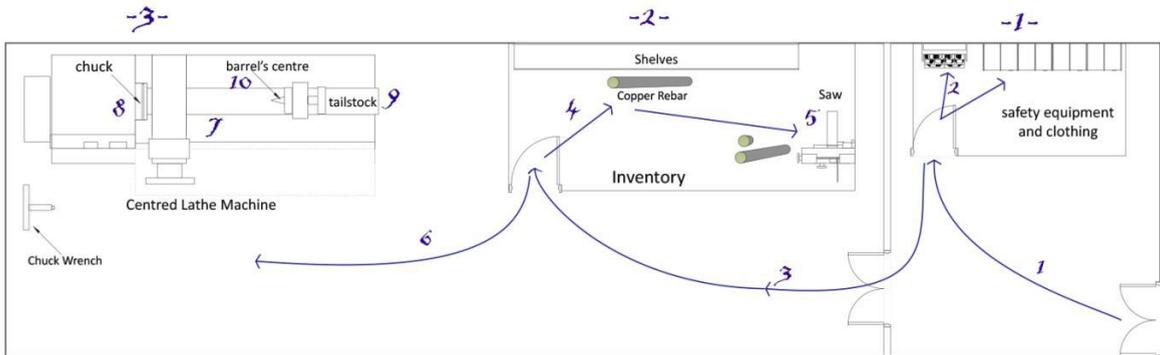


Fig 2. The Taxonomy of Actions for Action-based Learning Assessment in Virtual Training Environments: example scenario- Supporting copper rebar in the lathe machine chuck

In ALAM, action-sequences are encoded in form of a list of single actions using the following syntax: <id><Action.Class><Action.Type>[<Action.Attribute>][<Action.Relation>]. <id> identifies the position in the sequence, <Action.Class> is a Functional Act, <Action.Type> is the instantiation of different actions (specific representative of the class), [<Action.Attribute>] is a list of possible attributes such as quality, quantity, and locations, and [<Action.Relation>] is a list of possible relations to other actions. Note, that ALAM also recognizes Irrelevant Actions to allow fault-free assessment and comprehensive feedback [6]. All the performed actions by avatars in VTE will be encoded by the mentioned syntax and the output will be in form of a list sent to the Evaluation Agent to be compared with experts' given solution and be evaluated. Action-based Learning Assessment System (ALAS) is not in the scope of this paper and will be discussed in other publications.

Table 2: Action Recognition Agent's output for constitutive act 1 (enters the shop and collect safety equipment and clothing)

User's Actions and Action-sequences	Expert's Actions and Action-sequences
<1><Locomotive><teleport>[<Enter>]	<1><Locomotive><Walk><Enter>
<2><Operative><read>[<manual>]	<2><Locomotive><Walk><in>[<in><SafetyRoom>
<3><Locomotive><Walk>[<in><SafetyRoom>]	>]
<4><Decisional><choose>[<equipment>]	<3><Decisional><choose>[<equipment>]
<5><Operative><collect>	<4><Operative><collect>
<6><Decisional><choose>[<cloths>]	<5><Decisional><choose>[<cloths>]
<7><Operative><collect>	<6><Operative><wear>
<8><Locomotive><run>[<out><SafetyRoom>]	<7><Locomotive><Walk>[<out><SafetyRoom>]

### 4.3. Interactions within Virtual Training Environments

When avatars reciprocate with other avatars or their environment they use different acts to show different forms of reciprocation. These interactions can be in form of nonverbal, speech acts, or gestures. The major interactions in this taxonomy are:

- Avatar-Avatar
- Avatar-Environment (and vice versa)
- Avatar-Objects (and vice versa)

Virtual Training Environments may be used for collaborative assessment or assessment within a group. This can add different kinds of interactions to current three mentioned ones in this taxonomy.

#### 4.3.1. Functional Act towards Avatar-Avatar interactions

Avatars use verbal and nonverbal actions to interact with other avatars. They express themselves and communicate with other avatars using speaking, chat, and gestures. Avatars also express themselves using different Functional Acts including Responsive actions. Using facial gestures and head and hand movements avatars communicate with each other or in the case of assessment the respond to the questions.

#### 4.3.2. Functional Act towards Avatar-Environment interactions

Avatars use Functional Acts such as Locomotive, Constructional, Decisional, and Responsive to interact with the environment and make changes in it. Moving around, building or destroying, and modifying are some of these interactions; avatars also use Decisional actions to interact with their environment like turning on or off a switch. An avatar can also respond to the environment, for example touching a hot surface and reacting to that. Avatars' actions may change the environment, which leads to new interactions.

#### 4.3.3. Functional Act towards Avatar-Objects interactions

Avatars use Constructional actions to create or manipulate objects in their environment and Operative actions to use or move objects. They also use Responsive actions to show their reaction towards the effect of objects on them. As part of the environment objects affect avatars and require correct response to show the understanding and knowledge of the user, reflected by the avatar.

## Summary

To sum up briefly, assessing knowledge of students based on their goal-oriented actions reflected by their avatars in virtual training environments needs a clear classification of actions in these environments. Different taxonomies of human actions, performance, and behavior were developed in recent decades, but all these taxonomies are specifically developed for the developer's research and there is not any general taxonomy of actions. Similarly, the taxonomy of actions in virtual training environments is developed for Action-based Learning Assessment Method (ALAM), but it is an open taxonomy and it can be easily adapted and extended for other fields of research. Interactions in VTE are limited to avatars with avatars, environment, and objects, but in virtual worlds, it may include intelligent bots as well.

## References

- [1] Goldman A. I. *A Theory of Human Action*, Prentice-Hall; 1970.
- [2] Fleishman EA. Toward a taxonomy of human performance. *American Psychologist*; 1975, 30(12), 1127–1149.
- [3] Robertson T. Cooperative work and lived cognition: a taxonomy of embodied actions. In: *Proceedings of European conferences on Computer-Supported Cooperative Work*, 1997, 97, 205–220.
- [4] Robertson T. Building bridges: negotiating the gap between work practice and technology design. *International Journal of Human-Computer Studies*, 2000, 53(1), 121 – 146.
- [5] Naidu S, Bedgood Jr D. Action-Based Learning. In: Seel N, editor, *Encyclopaedia of the Sciences of Learning*. Springer US; 2012, p. 75-77.
- [6] Fardinpour A, Rainers T, Dreher H. Action-based Learning Assessment Method (ALAM) in Virtual Training Environments, In: *ASCILITE 2013: 30th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Electric Dreams*; 2013, 267-277
- [7] Allen JF. Towards a general theory of action and time. *Artificial Intelligence* 1984; 23(2): 123 – 154.
- [8] Anscombe GEM. *Intention*. Harvard University Press; 1957.
- [9] Davidson D. Actions, reasons, and causes. *The Journal of Philosophy*; 1963, 60(23), p. 685-700.
- [10] Davidson D. The logical form of action sentences. *Essays on actions and events*; 1967, 5, p. 105-148.
- [11] Shwayder DS. *The stratification of behaviour*. Routledge & Kegan Paul; 1965.
- [12] D'Arcy E. *Human acts*. Clarendon Press; 1963.

- [13] Cappella J, Pelachaud C. Rules for responsive robots: using human interaction to build virtual interaction. In: Vangelisti AL, Reis HT, Fitzpatrick MA, editors. *Stability and Change in Relationships*, Cambridge University Press, New York; 2001, p. 325-353
- [14] Verhulsdonck G, Morie J. Virtual Chironomia: Developing Standards for Non-verbal Communication in Virtual Worlds. *Journal of Virtual Worlds Research*, 2009, 2(3).
- [15] Mehrabian, A. *Nonverbal communication*. Transaction Publishers; 1972
- [16] Chodos D, Stroulia E, King S, Carbonaro M. A framework for monitoring instructional environments in a virtual world. *British Journal of Educational Technology*; 2012.
- [17] Al-Smadi M., Gütl C., & Kannan R. Modular Assessment System for modern learning settings: Mass. *International Journal of Computer Applications*; 2010, 1(9), 37-43.
- [18] Dunwell I, De Freitas S, Jarvis S. Four-dimensional consideration of feedback in serious games. *Digital games and learning*; 2011, p. 42–62.
- [19] Shen R, Tang Y, Zhang T. The intelligent assessment system in Web-based distance learning education. In: *31st Annual Frontiers in Education Conference*, 2001, 1, TIF 7–11.
- [20] Reiners T, Wood LC, Dron J. From chaos towards sense: A Learner-centric Narrative Virtual Learning Space. In: J. Bishop, editor, *Gamification for Human Factors Integration: Social, Educational, and Psychological Issues*. Hershey, PA: IGI; forthcoming.
- [21] Rogers CR. *Client-centered therapy: its current practice, implications, and theory*. Oxford, England: Houghton Mifflin; 1951.