Chapter 6
Action-Based Learning Assessment in Virtual Training Environments

Ali Fardinpour, Torsten Reiners and Lincoln C. Wood

Introduction

It is common to pass an induction and attend continuous training to receive and keep permission to work on a site with health and safety risks. Furthermore, it is generally not possible to perform the training on systems that are currently used for operation due to safety and business matters. A common alternative is the usage of replica in virtual training environments to simulate relevant processes and activities as well as the surroundings and scenarios. However, it requires a well-balanced orchestration of domain expert knowledge, (educational) technology and instructional designers with experience in developing virtual training units to recreate authentic, immersive and engaging learning experiences with later transferability to the real world. The suitability of virtual environments itself is justified by depicting the infeasibility of scenarios to be considered valuable in a real world setting, i.e. high costs (space simulator), high risks of injuries for learners and educators (handling of hazardous material) or near-impossibility but with a high degree of importance (natural disaster recovery); (see Hewitt et al. 2010). Advanced technology is used to detach the learner from real and potentially unsafe environments while maintaining authenticity, e.g. using aircraft mock-ups, realistic dashboards as user interfaces and simulations of movements and situations.

A. Fardinpour (✉)
Wise Realities, Curtin University, Perth, Australia
e-mail: Ali.fardinpour@curtin.edu.au

T. Reiners
Curtin University, Perth, Australia
e-mail: t.reiners@cbs.curtin.edu.au

L.C. Wood
University of Otago, Dunedin, New Zealand
e-mail: Lincoln.Wood@otago.ac.nz

S. Gregory and D. Wood (eds.), *Authentic Virtual World Education*,
https://doi.org/10.1007/978-981-10-6382-4_6

Virtual training environments are used in a manifold of situations, e.g. training for surgery, mechanical engineering or health and safety (Filigenzi et al. 2000; Gunn 2006; Hockenmeyer et al. 2009; Kizil 2003). A key aspect of these environments is to develop or advance previous inherited knowledge by immersing (presence) and engaging (motivation) for the learner to explore the scenario and to progress towards given objectives. Given the authenticity of activities, it is possible to map the acquired knowledge later to their correlated real situations counterparts (Bastiaens et al. 2014; Herrington et al. 2010). However, learning progress is related to the understanding of the nature of errors and reflection on the actions that lead to the outcome (Sadler 1989).

Assessment of learning outcomes can be classified as formative (identify the quality of the assessment and provide constructive explanations) or summative (ranking the quality of the assessment) (Scriven 1967). Summative feedback is easy to generate, yet the sole assessment of outcomes by scores and grades seems like a naïve excuse to ignore the richness of formative feedback. The limited dimension of scores ignores how the learner reacts to stimuli and applies learned knowledge to make decisions during the learning process. The assessment has to include the sequence of actions that lead to an outcome, as it is otherwise not possible to deduct the successful application of knowledge versus successfully achieving the objective by coincidence. Furthermore, the sequence of actions can be used to generate detailed formative feedback (Reiners et al. 2013, 2014).

The drawback of successful formative feedback is the high investment of resources such as the time it takes for marking assessment tasks, i.e. with intelligent assessment systems to support or even replace the human evaluator still in its early childhood (Fardinpour and Dreher 2012). Regarding activities in virtual training environments, we are facing even further unsolved challenges; among others the recognition of human performance (activities), interpretation of the human behaviour, and a commonly accepted standard to encode and communicate the actions. In this chapter, we are addressing the encoding by suggesting an adaptive taxonomy being designed to encompass the multitude of disciplines and requirements (Chodos et al. 2014; Robertson 1997, 2000; Verhulsdonck and Morie 2009). We further describe an approach to compare consecutive training sessions (i.e. the action sequences) to identify changes in the action and their impact on the overall outcome. It is further possible to assess the learners' actions against the ones from experts to generate formative feedback and provide guidance on how to improve in further training sessions. The following section covers a brief introduction to virtual environments and action-based learning. Following, we describe the used taxonomy for user actions and provide an introduction in our method to assess the action-based learning (ALAM) and how formative feedback is generated.

Learning and Assessment in Virtual Learning Environments

Virtual learning environments (VLE) are "computer-based environments that are relatively open systems, allowing interactions and encounters with other participants" (Wilson 1996, p. 8). Virtual training environments (VTE) represent a subset of VLE by setting the focus on skills training including some specialised methods like intelligent pedagogical agents (Rickel et al. 1998), game-based tutoring (Craighead 2008), gamification- or game-based structures (Wood and Reiners 2013) or educational simulation (Dede and Lewis 1995; Dede et al. 1999). VTE are used extensively in areas like surgery training (Gunn 2006), spinal anaesthesia (Hockenmeyer et al. 2009), dynamic hip screw surgery training in vitro (Ahmed et al. 2012) or oral implantology (Chen et al. 2012). Virtual environments "allow you to do things which would be difficult or impossible to do in the physical world—both literally and pragmatically" (Twinning 2009, p. 498). And despite the generally restricted functionality with focus on building scenarios and providing collaborative communication tools, virtual worlds such as second life found wide application in education, e.g. teacher education (Gregory et al. 2011; Masters et al. 2013), engineering (Bresciani et al. 2010), health sciences (Thompson and Hagstrom 2011), logistics and manufacturing (Wriedt et al. 2008) and simulation of hazardous situations for training purposes (Reiners et al. 2013; Reiners and Wood 2013).

Avatars are digital representations being used to project the learners' view in the virtual environment. The most common and the most immersive option is a positioning of the camera, the virtual eyes, at the position of the head to provide a first-person perspective. This allows a mirroring of head movements in virtual and real space. An alternative is the third-person perspective, where the camera follows the avatar. While the learner can follow gestures and interactions of the avatar from an observer perspective, it reduces the immersive perception. The control of the avatar includes various options; among others, traditional input devices (e.g. keyboard and mouse), advanced technology (e.g. Kinect or Razor Hydra) or replicas of real world control interfaces such as the dashboard of a truck to allow the real world haptic experience. The environment is often shared with other avatars; either controlled by humans or computers, so-called intelligent bots or agents (Wood and Reiners 2013).

Authentic learning has been used in different disciplines over time to increase the quality of training in education systems. Authentic learning is about engaging students in learning about, and solving, real-life problems by the means of simulation and (educational) technology (Herrington and Herrington 2006). By reviewing the research related to the use of simulations in the classroom, Smith (1987, p. 409) concluded that the "physical fidelity" of the simulation materials is not as important as the "realistic problem-solving processes" that simulation promotes, a process Smith (1987) describes as the "cognitive realism" of the task Barab et al. (2000, p. 38) also stated that authenticity occurs "not in the learner, the task, or the environment, but in the dynamic interactions among these various components

[…] authenticity is manifest in the flow itself, and is not an objective feature of any one component in isolation." In the same way, Herrington et al. (2003) argued that the cognitive authenticity is much more important than the physical authenticity in the design of authentic learning environments.

Learning by doing or action-based learning refers to orchestrate learning by the learner (Naidu and Bedgood 2012). Thus, legitimate learning actions may vary from an active participation (e.g. building, creating or drawing something) to passive observation that is later examined, reflected on or becomes a seed for a later decision-making process (Naidu and Bedgood 2012). The literature distinguishes different models of action-based learning (Fardinpour and Reiners 2014); including problem-based learning (Barrows and Tamblyn 1980), inquiry or goal-based learning (Schank 1997), scenario-based learning (Naidu 2010) and adventure learning (Doering 2006). Whilst each model has a distinguished focus or perspective, all start with a defined problem or objective (Naidu 2007). Action-based learning is characterised by a learner-centric model where the learner studies the learning material and afterwards applies the lesson learned. This learning by doing approach differentiates action-based learning from action learning, where the learning process is "using personal experience and reflection, group discussion, and analysis, trial-and-error discovery, and learning from one another" (Lasky and Tempone 2004, p. 87). For example, one group is sharing experiences in a discussion (action learning), while the other member of the other group learns by actively performing the tasks to solve a problem (action-based learning).

Summative assessment methods such as multiple-choice or closed-answer questions are useful to rank and grade the learning outcome, but lack supportive, probing or explanatory feedback. Learners require a detailed analysis to understand the reason for their actions and deduct a change in their learning behaviour (Rogers 1951; Sadler 1989). Action-based learning is about flexibility, complexity and creativity; attributes that are difficult to judge with a score (Naidu 2010). Therefore, additional flexibility is required to cope with the assessment patterns that reflect actions of the learners, particularly at more advanced levels of learning (Wood and Reiners 2013). Thus, in VTE the demonstrated actions and abilities of learners need formative feedback as an important component of the learning process.

Assessment of learners' mastery in VTE is conducted primarily by experts who observe and analyse the training. Shute, Ventura, Bauer and Zapata-Rivera (2009, p. 299) argue that the assessments should be "seamlessly woven into the fabric of the learning environment" so that it is practically indiscernible for the learner; therefore, causing no distraction. Their stealth assessment uses automated scoring and machine-based reasoning techniques to infer, for example, the "value of evidence-based competencies across a network of skills" (Shute et al. 2009, p. 299). Shute used stealth assessment formally for the first time in 2005 during an American Educational Research Association (AERA) symposium on diagnostic assessment, but it was designed and employed two decades earlier as part of a guided discovery world called Smithtown (Shute 2011; Shute and Glaser 1990). Al-Smadi et al. (2010) propose a framework using stealth assessment to assess action choices and sequences in serious games; creating formative feedback on the interpretive level of Rogers' feedback classification (Rogers 1951).

"Assessment is authentic when we directly examine student performance on worthy intellectual tasks. Traditional assessment, by contract, relies on indirect or proxy 'items', efficient, simplistic substitutes from which we think valid inferences can be made about the student's performance at those valued challenges" (Wiggins 1990, p. 2). He further explains authentic assessment by comparing it to the traditional testing of learning outcomes. See the following list for some distinction criteria:

- Instead of testing for recognising and recall, authentic assessment requires an effective application of acquired knowledge.
- Authentic assessment is holistic in the assessment tasks to represent priorities and challenges and not limited to paper-based only one specific valid answer questions.
- Authentic assessment is about testing the understanding and reasoning of answers, not plainly repeating the only solution from the textbook.
- The validity of answers in the context of authentic assessment depends on real world validity, not a match with the textbook or course material.
- Authentic assessment maps the "ill-structured" challenges and roles of real-word scenarios; not providing a clean, discrete and simplistic reflection of it.

Meyer (1992) believed that it is very important to consider "authentic to what?" and named "few facets of authenticity: stimuli, task complexity, locus of control, motivation, spontaneity, resources, conditions, criteria, standards, consequences" (p. 40). According to Meyer (1992), each assessment needs to address at least a few of these facets, if not all, to be considered as an authentic assessment.

The Taxonomy of Human Actions

One purpose of taxonomy is the ordered categorisation and unique specification of items; in our case human actions. Robertson (1997, 2000) created the taxonomy of embodied actions for the cooperative design in a distributed company "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" (2000, p. 130). Robertson embedded open and flexible categories, allowing others to adapt it to their own requirements. Embodied actions were split into different classes; relating to physical objects (among others movement or use), other person (among others monitoring or pretending) and workspace (among others moving or looking); (see also Verhulsdonck and Morie 2009).

The literature review on taxonomies of human's actions, embodied actions, actions in virtual worlds and behaviour modelling showed that researchers tend towards creating their own taxonomy of actions or behaviours based on their

Fig. 6.1 BEHAVE taxonomy classifying actions in three different classes

project's needs. Examples are Fleishman (1975) who identifies six categories of human performance: identification, discrimination, sequence learning, motor skill, scanning and problem-solving; or Goldman (1970) who identifies four categories: individuation, act-type, act-token, basic- and non-basic-actions. Goldman (1970, p. 6) disagrees with the assumption that two different actions can be both recognised as basic actions based on the same identity thesis. Goldman (1970, p. 6) reminds us that, "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."

Fardinpour and Reiners (2014) use the "Basic Exploratory Human Actions in Virtual Environments" (BEHAVE) taxonomy of human actions to classify learners' goal-oriented actions. This taxonomy further classifies learners' actions into three levels: The Goal Act, Constitutive Acts, and Functional Acts; see Fig. 6.1. To achieve the Goal Act (primary goal in the scenario), the learner must perform a sequence of Constitutive Acts, which are composed of Functional Acts. Constitutive Acts can be considered as milestones; an approach illustrated by Reiners et al. (2013) for their narratives in virtual environments. However, Functional Acts enhance the ability of an assessor to examine how these milestones are achieved.

Action-Based Learning Assessment Method

Although action-based learning scenarios have been used in virtual training environments before, there is still a lack of comprehensive assessment methods. In this section, we describe the action-based learning assessment method (ALAM), which is focusing on the formative assessment of the learners' performed, goal-oriented, actions. The core functionality of ALAM is the classification and codification of human actions according to the BEHAVE taxonomy previously described. ALAM is a part of an overall process (action-based learning assessment system, ALAS) to record actions, analyse the behaviour, assess against expert solutions and generate formative feedback for an overall compelling learning experience. ALAM uses Rogers' 5-stage feedback classification, which is still valid and commonly used in assessing the learning outcome (Al-Smadi et al. 2010; Dunwell et al. 2011). Human assessors are capable of providing feedback on all stages; however, it is more common to simplify the process by designing multiple-choice or short-answer assessments (Stage 1 and 2). This is particularly true as formative feedback at stage 4 or 5 requires experts to understand whether the student's answer 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. This assessment is of higher complexity as it requires understanding of the problem, the context, and often natural language as used by the student. Intelligent assessment algorithms are not yet capable of automatically assessing the learner without introducing sufficient constraints to reduce the problem and solution space (Shen et al. 2001) in a way that largely limits the value of these algorithms in practical settings.

Action Sequence

To consider a learning experience successful, learners need to achieve a predefined goal (i.e. the Goal Act) by performing actions in a given virtual training environment. Following the recording and recognition of actions, the first process comprises the identification of relevance, aggregation of atomic actions to higher level concepts as well as classification according to the BEHAVE taxonomy (Fardinpour and Reiners 2014). For example, smiling and stretching out the arm for a handshake implies a friendly welcome gesture. This sequence of actions is used to assess the learning by analysing its components against expected outcomes and the previously recorded sequences of experts. The creation of formative feedback based on the learning purposes and will support the learner in understanding his/her actions against the expected actions and will support in changing his/her behaviour. It is important to emphasise that the comparison is not requiring an exact match but integrates deviation from the expert into the feedback by visualising the differences. In case of non-mandatory steps, sequences can use alternative actions to achieve the same goal; however, it is up to the learner and experts to judge the validity against further metrics. For example, if the learner is taking extra, non-required steps, it will not impact the primary goal but divert from the expected solution in using more resources such as time or material. The feedback is showing this difference as it shows room for improvement such as working more efficient but also allows experts to gain more insight in the process itself.

Learners' action sequences are compared to the expected action sequences recorded by experts or instructional designers. A first approximation of the

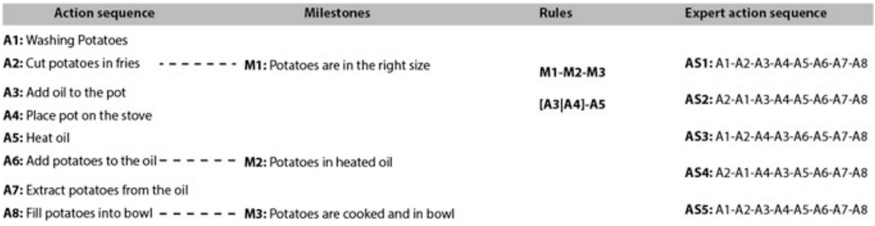

Fig. 6.2 Action sequence example for a making fries. Note that, we simplified the example for this purpose, e.g. more complex steps like “remove starch” and “adding spices” have been excluded

similarity is the match of specific milestones regarding of availability and order in both sequences and not considering the actions in-between (Reiners et al. 2013). This heuristic can provide an understanding of the general correctness of the solution, yet lacks the validation if milestones are achieved correctly. Note that, the learner is encouraged to explore the space unconstrained and at their preferred speed; thus, milestones are often the only source for (formative) feedback. This approach anticipates providing immediate feedback at milestones to compare the learning success to previous experiences.

Presented is a simple example of the action sequence and requirements with the goal to prepare a bowl of fries made from fresh potatoes. On the left side of Fig. 6.2 is a simplified presentation of the action sequence for the learner to boil potatoes with predefined rules (such as may be established by an instructional designer) but allowing for some freedom. On the right side of Fig. 6.2 are the sequences recorded from experts completing this task. Defined are three milestones, which have to be gained in the order M1–M2–M3. Milestones reflect the achievements of sub-sequences of actions, representing a defined state of the environment. In case of M1, it was expected to have the potatoes in the right size; while not being specific about the pot, type of oil or temperature of oil. M2 is reached when the condition “potatoes in heated oil” is achieved; in general, after completing the second set of actions (A3–A6). However, if M1 was not fulfilled or the learner forgot some actions, the learner will be deemed to have failed to achieve the goal.

Action Recognition

All activities in the virtual training environment were recorded as a raw stream of data, consisting of information such as coordinates of the avatar, viewing direction, relation to objects, position of arms. Regarding the understanding and comparison of action sequences, it is important to analyse the data stream and map parts of the data stream to certain actions in the taxonomy (codification process). This includes the action itself, but also the relevant attributes (e.g. adjective, preposition, location, quantity, unit, object and location). The time-based sorted actions form the action sequences that are evaluated by comparing them to the experts' action sequences. The validity of the sequence is further verified against a set of rules that are either manually specified by the experts or deducted automatically from expert action sequences. For example, the occurrence of a sequence of actions in a specific order (as part of expert action sequences) can be used to derive a rule about predecessor relations. Further, rules are stated for sequences that only appear in some approved action sequences, which would specify alternative solutions. Rules can be stated on all levels of the BEHAVE taxonomy.

Constitutive Acts must be defined by the expert and represent a specific state of the environment. The recognition can be undertaken by so-called triggers (i.e. events that happen in the VTE and are recorded in addition to the actions) or specific action sequences which define the end or start of a Constitutive Act or milestone. For example, in Fig. 6.2, the M2 milestone ("potatoes in heated oil") defines the end of the Constitutive Act of heating the oil and placing the potatoes in it. The trigger is further defined by the preceding sequence of specific action, e.g. the placing of the pot on the oven, pouring the oil in the pot and turning on the heat.

The rationale behind ALAM as an authentic assessment can be summarised by what Janesick (2006) stated about learning with authentic assessment, which students learn from experience, context, learning community and responsibility for improvement. ALAM provides a detailed formative feedback based on trainees' actions, especially their goal-oriented actions, in a certain context with a clear Goal Act. It does not limit trainees to a set of predefined questions and provides them with an opportunity to use their learned knowledge and not just the memorised knowledge. Using the generated feedback under the ALAM's feedback structure and standards, trainees can learn from their mistakes, experts' solutions and also correct their performance. These features make ALAM an authentic assessment method, used to evaluate trainees' learned knowledge in simulated environments.

Action Comparison

The evaluation of the action sequences is undertaken on the level of Constitutive and Functional Acts. The comparison analysis of the action sequence is based on the rules as well as the similarity to stored sets of expert action sequences. The comparison includes, among others, the following criteria:

- Non-compliance of rules: rules are either strict (i.e. all experts have the same sequence of actions) or loose (i.e. only some experts have the same sequence of actions). Strict rules must be followed; loose rules represent alternatives, such that an exact match is not required.
- Attributes that do not match; including a weighting of the relevance of an attribute.
- Timing of the attributes, i.e. length between two actions indicating problems in deciding what to do next.

- Sequence of Constitutional Acts in comparison to the experts.
- Achievement of the Goal Act.

In a summative scored based assessment, the non-achievement of the goal would be considered a failure, yet it can be the result of only of minor failures at one action towards the end. Therefore, the comparison is looking for partial sequences (i.e. between milestones), validity of states at milestones (objects and the environment have certain attribute settings) and how these sequences match the experts' behaviour in their action sequences. For example, the five experts have a 3:2 opinion if potatoes should be washed first then cut or vice versa (order of action A1 and A2 in Fig. 6.2). For milestone 1, the learner requires either one or both would be considered to be correct. However, the assessment frequencies (matching 2 out of 5 in case of A2–A1) can be used in the feedback generation to provide alternatives. The assessment outcome is also used in later partial action sequences to check for further correlations, i.e. the likelihood to follow the action sequence that the learner had the highest match with. This is used to provide feedback, but also to validate and weight expert solutions.

Feedback Generation

The comparison generates evaluations on the similarity of the action sequences conducted by the leaner and all experts. A straightforward approach to create an automated feedback is matching sequences with a binary answer of "yes" and "no". However, this would not reflect on the variety of possible solutions and undermines the assumed inerrability of experts. The feedback should relate the learner's outcome to the aggregated expert solutions provide a feedback that shows deviation and explains the impact of these. For example, not including action A4 (place pot on stove) is essential, thus causing an overall failure. The generated similarities (i.e. matching actions and sequences for partial sequences) can be used to generate feedback, e.g. visually comparing the chosen path and the one taken by the experts. The formative feedback must distinguish between Constitutive and Functional Acts, the first one generally being the milestones. If all experts have these milestones in their action sequence, it is required and should be achieved by the learner as well. Thus, the feedback must emphasise such mismatches. The same applies for actions that all experts have taken; yet, it is important for others that are not done by all experts (e.g. wiping the table after each step), might not be mandatory and therefore need not be done by the learner. The feedback should allow aggregating and disaggregating details; it is not relevant to focus on a single, irrelevant action if the Goal Act was not even achieved. Note that, feedback should consider case-relevant templates completed by the experts that are individualised with details from the learner (e.g. where the learner is repeating the same failure).

Figure 6.3 visualises an abstraction on the feedback generation, yet spares details with respect to the focus of the chapter. The learners performed actions (2nd

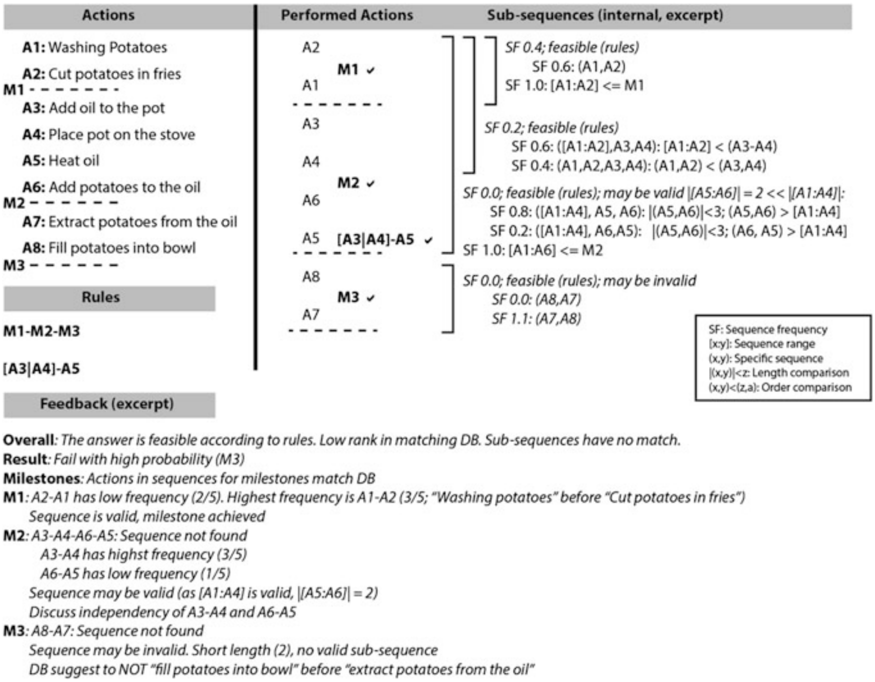

Fig. 6.3 Example of sequencing action sequences to generate automated formative feedback (excerpt)

column) are compared to the expert action sequences as shown in Fig. 6.2. The comparison is based on milestones (is a certain state achieved (environment) and corresponds with Constitutive Acts (experts)) and sub-sequences of different lengths. The feedback is based on the sequence frequency (how often is this specific sub-sequence found in the database), the feasibility according to the rules, and how likely it is to find a specific sub-sequence after previous actions. For example, the second sub-sequence covers the first four actions (A1, A2, A3, A4) and is feasible as there are no violations of the given rules. The sub-sequence is found in the database; however, has a low frequency of only 1 out of 5 cases. The low-frequency results from the order of the preceding actions (A1, A2), as a reversed order of these actions would increase the overall sub-sequence frequency to 0.4; implying a total validity of (A3, A4) with 0.6. The feedback to the learner emphasises the deviation from the expert solutions, yet shows the similarity according to frequencies that the solution is likely to be feasible.

Research Outlook

This chapter introduced the current research on evaluating performance within a virtual training environment with a focus on moving towards automated assessment using ALAM, human actions taxonomy and ALAS. The chapter is research in progress, however, preliminary experiments show the validity as well as enhanced opportunity to describe and evaluate training sessions in virtual training environments. The research is ongoing, with focus on the balancing of the similarity calculation and the improvement of automated generation of rules.

The system is intended to recognise and analyse action streams from different VTE, which are then mapped using the same taxonomy. Thus, the expert performance could be recorded in real world scenarios and later used to assess the learners' performance in a simulated environment. The described taxonomy and ALAM are part of a larger system called action-based learning assessment system (ALAS) shown in Fig. 6.4. Actions from experts (top layer) and trainee (lower layer) are identified, verified, mapped to the taxonomy and stored as action sequences. The expert sequences are further used to deduct rules; describing reoccurring patterns and dependencies that can be used during the comparison process. The reference sequences (experts) and performed sequences (trainee) are compared, and the evaluated outcome is used to generate the feedback.

Advancement in this area of automated assessment (focusing on providing formative feedback) is important to support wider adoption of the rapidly advancing use of virtual environments in education. At present, practical adoption demands formative assessment remain a small component of the system or relies on peer- or expert-provided feedback. Instead, high levels of formative feedback, as planned in the nDiVE project (ndive-project.com), require significant and effective use of formative feedback to be provided to learners to enable self-guided learning. In a nutshell, nDiVE is exploring the immersive space for health and safety training using head-mounted displays, i.e. addressing scenarios of high risk to have fatal injuries. The most prominent example in nDiVE is a container terminal simulation with tasks to solve while not risking your own or other lives.

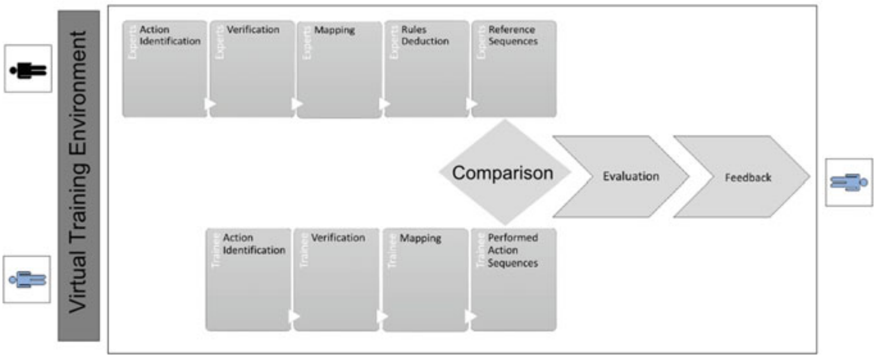

Fig. 6.4 Action-based learning assessment system; (see also Fardinpour et al. 2013)

Acknowledgements The Australian Government Office has provided support for the production of this publication for Learning and Teaching (Grant: Development of an authentic training environment to support skill acquisition in logistics and supply chain management, ID: ID12-2498). The views expressed in this publication do not necessarily reflect the views of the Australian Government Office for Learning and Teaching.

References

Ahmed, A., Maqsood, M., Saif, H., & Salman, A. (2012). Affordable interactive virtual reality system for the Dynamic Hip Screw surgery training in vitro. In *Proceedings of the IASTED International Conference on Computer Graphics and Imaging, CGIM 2012* (pp. 248–253). Crete, Greece: Acta Press.

Al-Smadi, M., Gütl, C., & Kannan, R. (2010). Modular assessment system for modern learning settings: Mass. *International Journal of Computer Applications, 1*(9), 37–43.

Barab, S. A., Squire, K. D., & Dueber, W. (2000). A co-evolutionary model for supporting the emergence of authenticity. *Educational Technology Research and Development, 48*(2), 37–62.

Barrows, H. S., & Tamblyn, R. (1980). *Problem-based learning: An approach to medical education*. Berlin: Springer.

Bastiaens, T., Wood, L. C., & Reiners, T. (2014). New landscapes and new eyes: the role of virtual world design for supply chain education. *Ubiquitous Learning: An International Journal, 6*(1), 37–49.

Bresciani, M. J., Morsi, K., Tucker, M., Siprut, M., Stewart, K., & Duncan, A. (2010). Designing and implementing an assessment plan for a virtual engineering lab. Eludamos. *Journal for Computer Game Culture, 4*(2), 277–285.

Chen, X., Lin, Y., Wang, C., Shen, G., & Wang, X. (2012). A virtual training system using a force feedback haptic device for oral implantology. In Z. Pan, A. Cheok, W. Müller, M. Chang, & M. Zhang (Eds.), *Transactions on Edutainment VIII* (Vols. 1–7220, pp. 232–240). Berlin, Heidelberg: Springer.

Chodos, D., Stroulia, E., King, S., & Carbonaro, M. (2014). A framework for monitoring instructional environments in a virtual world. *British Journal of Educational Technology, 45* (1), 24–35.

Craighead, J. (2008). Distributed, game-based, intelligent tutoring systems: The next step in computer based training? In *International Symposium on Collaborative Technologies and Systems* (pp. 247–256). Irvine, CA: IEEE.

Dede, C., & Lewis, M. (1995). *Assessment of emerging educational technologies that might assist and enhance school-to-work transitions*. Washington, DC, USA: Office of Technology Assessment, U.S. Congress.

Dede, C., Salzman, M., Loftin, R. B., & Sprague, D. (1999). Multisensory immersion as a modelling environment for learning complex scientific concepts. In W. Feurzeig & N. Roberts (Eds.), *Modelling and simulation in science and mathematics education* (pp. 282–319). New York: Springer.

Doering, A. (2006). Adventure learning: Transformative hybrid online education. *Distance Education, 27*(2), 197–215.

Dunwell, I., de Freitas, S., Jarvis, S., & de Freitas, S. (2011). Four-dimensional consideration of feedback in serious games. In S. de Freitas & P. Maharg (Eds.), *Digital games and learning* (pp. 42–62). London, UK: Bloomsbury Publishing.

Fardinpour, A., & Dreher, H. (2012). Towards developing Automated Assessment in Virtual Learning Environments. In P. Kommers, T. Issa & P. Isaías (Eds.), *Proceedings of the IADIS International Conference Internet Technologies and Society 2012* (pp. 211–216). Perth, Australia: IADIS Press.

Fardinpour, A., & Reiners, T. (2014). The taxonomy of goal-oriented actions in virtual training environments. *Procedia Technology, 1,* 38–46. doi:10.1016/j.protcy.2014.02.007.

Fardinpour, A., Reiners, T., & Dreher, H. (2013). Action-based learning assessment: Evaluating goal-oriented actions in virtual training environments. In H. Carter, M. Gosper and J. Hedberg (Eds.), *Electric Dreams. Proceedings Ascilite 2013 Sydney.* (pp. 267–277). Sydney, Australia: Macquarie University and ascilite. Retrieved from http://www.ascilite.org/conferences/sydney13/program/papers/Fardinpour.pdf.

Filigenzi, M. T., Orr, T. J., & Ruff, T. M. (2000). Virtual reality for mine safety training. *Applied Occupational and Environmental Hygiene, 15*(6), 465–469. doi:10.1080/104732200301232.

Fleishman, E. A. (1975). Toward a taxonomy of human performance. *American Psychologist, 30* (12), 1127–1149. doi:10.1037/0003-066X.30.12.1127.

Goldman, A. I. (1970). *A theory of human action*. Upper Saddle River, NJ: Prentice-Hall.

Gregory, S., Dalgarno, B., Campbell, M., Reiners, T., Knox, V., & Masters, M. (2011). Changing directions through VirtualPREX: Engaging pre-service teachers in virtual professional experience. In G. Williams, N. Brown, B. Pittard, & B. Cleland (Eds.), *Changing demands, changing directions*. *Proceedings Ascilite Hobart 2011* (pp. 491–501). Hobart, Australia: University of Tasmania and ascilite. Retrieved from http://www.ascilite.org/conferences/hobart11/downloads/VPREXAscilitePresentationpptx.pdf.

Gunn, C. (2006). Collaborative virtual sculpting with haptic feedback. *Virtual Reality, 10*(2), 73–83. doi:10.1007/s10055-006-0044-4.

Herrington, A., & Herrington, J. (2006). *Authentic learning environments in higher education*. Hershey, PA: IGI Global.

Herrington, J., Oliver, R., & Reeves, T. C. (2003). Patterns of engagement in authentic online learning environments. *Australian Journal of Educational Technology, 19*(1), 59–71.

Herrington, J., Reeves, T. C., & Oliver, R. (2010). *A guide to authentic e-learning*. New York, NY: Routledge.

Hewitt, A. M., Spencer, S. S., Mirliss, D., & Twal, R. (2010). Incident and disaster management training: Collaborative learning opportunities using virtual world scenarios. In E. Asimakopoulou & N. Bessis (Eds.), *Advanced ICTs for disaster management and threat detection: Collaborative and distributed frameworks* (pp. 179–200). Hershey, PA: Information Science Reference.

Hockenmeyer, C., Nussbaumer, A., Lövquist, E., Aboulafia, A., Breen, D., Shorten, G., & Albert, D. (2009). Applying a web and simulation-based system for adaptive competence assessment of spinal anaesthesia. In M. Spaniol, Q. Li, R. Klamma, & R. H. Lau (Eds.), *Advances in Web Based Learning: ICWL 2009* (Vol. 5686, pp. 182–191). Berlin: Springer. doi:10.1007/978-3-642-03426-8_23

Janesick, V. J. (2006). *Authentic assessment primer*. New York, NY, USA: Peter Lang.

Kizil, M. (2003). Virtual reality applications in the Australian minerals industry. In F. A. Camisani-Calzolari (Ed.), *Application of computers and operations research in the minerals industries* (pp. 569–574). Cape Town, South Africa: The South African Institute of Mining and Metallurgy.

Lasky, B., & Tempone, I. (2004). Practising what we teach: Vocational teachers learn to research through applying action learning techniques. *Journal of Further and Higher Education, 28*(1), 79–94.

Masters, Y., Gregory, S., Reiners, T., Knox, V., & Dalgarno, B. (2013). VirtualPREX: Developing teaching skills in second life. In C. DeCoursey & S. Garrett (Eds.), *Studies in virtual world learning and practice* (pp. 69–94). Oxford, UK: Inter-Disciplinary Press.

Meyer, C. A. (1992). What's the difference between "Authentic" and "Performance" assessment? *Educational Leadership, 49*(8), 39–40.

Naidu, S. (2007). Instructional designs for distance education. In M. G. Moore (Ed.), *Handbook of distance education* (pp. 247–258). Mahwah, NJ, USA: L. Erlbaum Associates.

Naidu, S. (2010). Using scenario-based learning to promote situated learning and develop professional knowledge. In E. P. Errington (Ed.), *Preparing graduates for the professions using scenario-based learning* (pp. 39–49). Brisbane, Australia: Post Pressed.

Naidu, S., & Bedgood, D., Jr. (2012). Action-based learning. In N. Seel (Ed.), *Encyclopaedia of the sciences of learning* (pp. 75–77). Boston, MA, USA: Springer, US.

Reiners, T., & Wood, L. C. (2013). Immersive virtual environments to facilitate authentic education in Logistics and supply chain management. In Y. Kats (Ed.), *Learning management systems and instructional design: Metrics, standards, and applications* (pp. 323–343). Hershey, PA: IGI Global.

Reiners, T., Wood, L. C., & Dron, J. (2014). From chaos towards sense: A learner-centric narrative virtual learning space. In J. Bishop (Ed.), *Gamification for human factors integration: Social, educational, and psychological issues* (pp. 242–258). Hershey, PA: IGI Global.

Reiners, T., Wood, L.C., Gregory, S., Petter, N., Teräs, H., Gütl, C., Chang, V., Herrington, J. (2013). nDiVE: The story of how Logistics and Supply Chain Management should be taught. In M. Gosper, J. Hedberg, & H. Carter (Eds.), *Electric Dreams. Proceedings ascilite 2013 Sydney* (pp. 734–744).

Rickel, J., Johnson, W. L., Way, A., & Rey, M. (1998). STEVE: A pedagogical agent for virtual reality. *The second international conference on autonomous agents* (pp. 332–333). New York, NY, USA: ACM.

Robertson, T. (1997). Cooperative work and lived cognition: a taxonomy of embodied actions. In *Proceedings of European Conferences on Computer-Supported Cooperative Work* (Vol. 97, pp. 205–220). Lancaster, UK: Springer Netherlands.

Robertson, T. (2000). Building bridges: negotiating the gap between work practice and technology design. *International Journal of Human-Computer Studies, 53*(1), 121–146. doi:10.1006/ijhc.2000.0373.

Rogers, C. R. (1951). *Client-centered therapy: its current practice, implications, and theory*. Oxford, UK: Houghton Mifflin.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119–144.

Schank, R. (1997). *Virtual learning: A revolutionary approach to building a highly skilled workforce*. New York: McGraw-Hill.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Chicago, IL: Rand-McNally.

Shen, R., Tang, Y., & Zhang, T. (2001). The intelligent assessment system in web-based distance learning education. In *31st Annual Frontiers in Education Conference,* (Vol. 1, p. TIF 7–11). Reno, NV, USA: IEEE.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte, NC: Information Age Publishers.

Shute, V. J., & Glaser, R. (1990). Large-scale evaluation of an intelligent tutoring system: Smithtown. *Interactive Learning Environments, 1*(1), 51–76.

Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Mahwah, NJ: Routledge, Taylor and Francis.

Smith, P. E. (1987). Simulating the classroom with media and computers past efforts. *Future Possibilities. Simulation & Gaming, 18*(3), 395–413.

Thompson, C. W. F., & Hagstrom, F. (2011). Modeling healthcare: Logistics in a virtual world. *IEEE Internet Computing, 12*(4), 100–104.

Twinning, P. (2009). Exploring the educational potential of virtual worlds: Some reflections from the SPP. *British Journal of Educational Technology, 40*(3), 496–514.

Verhulsdonck, G., & Morie, J. (2009). Virtual Chironomia: Developing Standards for Non-verbal Communication in Virtual Worlds. *Journal of Virtual Worlds Research, 2*(3). Retrieved from http://jvwr-ojs-utexas.tdl.org/jvwr/index.php/jvwr/article/view/657

Wiggins, G. (1990). *The case for authentic assessment* (pp. 1–6). Washington, DC, USA: ERIC Digest.

Wilson, B. G. (Ed.). (1996). *Constructivist learning environments: Case studies in instructional design*. Englewood Cliffs, NJ, USA: Educational Technology Publications.
Wood, L. C., & Reiners, T. (2013). Game-based elements to upgrade bots to non-player characters in support of educators. In A. Hebbel-Seeger, T. Reiners, & D. Schäfer (Eds.), *Synthetic worlds: Emerging technologies in education and economics* (pp. 257–277). Berlin, Germany: Springer.
Wriedt, S., Reiners, T., & Ebeling, M. (2008). How to teach and demonstrate topics of supply chain management in virtual worlds. In J. Luca & E. R. Weippl (Eds.), *Proceedings of ED-MEDIA 2008. World Conference on Educational Multimedia, Hypermedia & Telecommunications, Vienna, Austria* (pp. 5501–5508). Chesapeake, VA: AACE.

Author Biographies

Ali Fardinpour holds PhD (with hons) from Curtin University. His current role is the lead research scientist in Immersive Technologies in Learning and Assessment. His current research focus is action-based learning and assessment in virtual training environments. He also is an adjunct researcher at Curtin Learning and Teaching.

Torsten Reiners is Senior Lecturer in Logistics at the Curtin University, Australia. His research and teaching experiences are in the areas of operations research but include instructional design, development of adaptive learning environments, distant collaboration and mobile learning, which is also manifested in his PhD thesis (University of Hamburg, Germany) about adaptive learning material in the field of operations research. He has participated in multiple research projects using 3D spaces for learning support, i.e. to improve the authenticity of learning in classes in relation to production and simulation. His current research interests are in logistics, education technologies and emerging technologies.

Lincoln C. Wood is a Senior Lecturer (Operations and Supply Chain Management) at University of Auckland (New Zealand) and an Adjunct Research Fellow at Curtin Business School (Australia). Having received the 2009 Council of Supply Chain Management Professional's (CSCMP) Young Researcher Award (Chicago, USA) and the Outstanding Research Award at the 2010 International Higher Education Conference (Perth, Australia), Dr Wood is co-Leader on a 2012 OLT grant relating to gamified virtual environments for supply chain education. His research interests include operations and supply chain management; service operations; gamification; and educational, authentic, virtual environments.