

Josef Wiemeyer, Michael Kickmeier-Rust
and Christina M. Steiner

Abstract

In every digital game, players both act in and interact with the game. They use the options of game mechanics to achieve goals. For example, players move a controller to steer the motions of an avatar, or they press buttons to trigger certain actions. These actions and interactions lead to certain results like a successful finish of a quest, solving a problem, or increasing a score. Both the quality and results of actions and interactions are subsumed under the term *performance*. Assessment of player performance is required for several purposes, for example for in-game or online adaptation and for offline evaluation. This chapter addresses the issue of performance assessment in serious games. Performance is a complex concept comprising results and processes of actions and interactions of the players in and with the game. First, generic and domain-specific models of performance are introduced to illustrate the variety of approaches. Based on this knowledge, online and offline assessments of performance are discussed. Finally, the integration of online and offline performance assessment into the process of game adaptation is described.

10.1 Introduction

Imagine playing an educational serious game in physics where you try to solve tasks according to various quests. You take action and realize that you are unable to solve the tasks. Even worse, you do not get any hints or instructions to guide you to

J. Wiemeyer (✉)
Technische Universität Darmstadt, Darmstadt, Germany
e-mail: wiemeyer@sport.tu-darmstadt.de

M. Kickmeier-Rust · C.M. Steiner
Graz University of Technology, Graz, Austria

the solution. As a consequence, you feel angry or even disappointed. Further, imagine making progress in the game (after spending a long, hard time gaming). With increasing expertise, the tasks offered by the game may become very easy to solve. Therefore, you get increasingly bored. Obviously, the game does not offer tasks that are balanced enough—in the sense of establishing optimal difficulty according to your current competence and mood state—to challenge you. The game could have done much better if your current performance had been assessed continuously in order to use this information for appropriate adaptation and personalization. Perhaps the game would have realized your specific problem with a missing competence component, e.g., lacking knowledge about refraction. Furthermore, the game could have also determined your personal learning styles or preferences, and adapted the offered information accordingly.

Performance assessment in serious games is important for several reasons (see also Bellotti et al. 2013). First, as illustrated by the previous example, to maintain player experience and keep the players within the corridor of game flow the game has to be adaptive considering the current performance of the player. Second, to deliver feedback in the form of instructions, hints, score or awards the current performance of the player needs to be assessed. Third, to improve the game, information concerning the performance of all players is required, either for formative or summative evaluation. Fourth, summative evaluation is also required to deliver evidence for the effectivity and efficiency of the final version of the game.

Evaluation of serious games can be formative or summative. *Formative evaluation* is used during development and aims at testing and improving the serious game or parts of the game iteratively to eliminate all the weaknesses before finishing and releasing. *Summative evaluation* is the evaluation of the final serious game and aims at testing the end product according to certain guidelines, principles or standards.

To serve the above-mentioned goals, performance assessment has to be performed either online, i.e., during gameplay (“in-game” or “stealth”), or offline, i.e., after having finished playing the game (Ifenthaler et al. 2012a, b).

In this chapter, the issue of performance assessment in games is addressed. First, concepts and measures of game performance are introduced. Based on these conceptual fundamentals, advantages and challenges of online and offline assessment will be addressed. Finally, the relation of game assessment and game adaptation is discussed.

10.2 Performance in Games—Concepts and Measures

Performance in games is a complex concept. First, performance is always a process that evolves over time. For example, solving a problem may start with facing a problem, followed by thinking of possible solutions, and finally choosing and realizing one solution deemed appropriate. Second, performance brings forth observable or measurable results. Performance results or outcomes are specific to the particular application domain. Therefore, performance can mean solved tasks, change of health-related behavior or knowledge, awarded scores, knowledge gains, improved perceptual or motor skills or abilities, changed attitudes, etc. Whereas performance outcomes can be measured or observed easily, the processes lying behind these outcomes are often not as easily measurable or observable. For example, the correct or incorrect solution of a game task can be easily measured. However, the reasons for success or failure are not that obvious. In case of failure, the player may not have been motivated, or she may lack knowledge required to solve the task. Another possible reason may be that knowledge existed, but appropriate hints were missing to activate this knowledge. A further reason could be that the player was inattentive. Therefore, there are several possible reasons for a particular performance outcome. As a consequence, serious games require an explicit definition of performance to ensure appropriate assessment of both results and process of performance.

Performance in games comprises both the *processes and the results of actions and interactions* of the player(s) in the game. Equal performance results can be established by different processes. Different levels, components, and stages of performance can be distinguished. Performance measures are specific to the domain of application.

In general, human performance is organized on different levels. First, performance is observable or measurable at the social, behavioral and physiological level. Second, psychological processes like perception and cognition, as well as motivation, emotion, and volition, play an important role in the control of human performance; unfortunately, psychological processes are not available for direct measurements in the strict sense. However, they can be either inferred from behavioral and physiological data or assessed by asking the players to report (see also Chap. 9).

There are numerous models that attempt to identify the building blocks of human performance. One type of models addresses the structure of *movements* as spatio-temporal or neurophysiological phenomena, whereas other models analyze the structure of *actions* as goal-directed, and intentionally organized human-environment interactions including cognitive, motivational, emotional, and volitional processes (see Fig. 10.1). The structure of movement is usually analyzed by means of methods

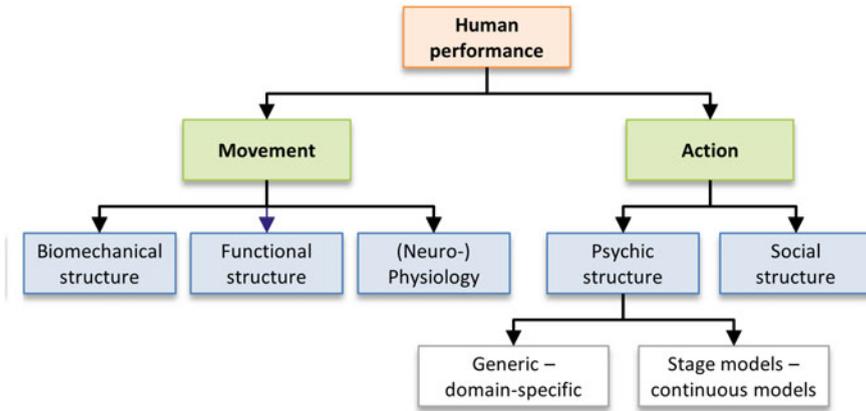


Fig. 10.1 Approaches to the structure of human performance

used in natural sciences like mathematics, physics, and biology. The structure of action is usually analyzed by psychology.

Human performance can be considered as movement or action. The term *movement* denotes the spatio-temporal and neurophysiological aspects of human-environment interactions, whereas *action* denotes the socio-psychological aspect of goal-directed and intentionally organized human-environment interactions, including cognition, motivation, emotion, and volition.

According to models of movement structure, game movements can be divided into different components, e.g., preparatory, main and end phase, supportive or main functional phases, or action-effect relations (e.g., Wiemeyer 2003). These concepts are very important for exergames, where complex movements have to be learned. For example, for an evaluation and correction of errors it is important to know which error is most important and how errors are associated and interrelated. An error in movement preparation may lead to a subsequent error in the main phase that can easily be corrected when addressing the preparation error.

The neurophysiological aspects of movements comprise neural functions of the central nervous systems, particularly of the brain and spinal cord. Specific parts of the brain are responsible for controlling movement and action, e.g., the primary sensory and motor cortex, the supplementary and premotor cortex, the basal ganglia, and the cerebellum. In the spinal cord, several sensory-motor reflexes are organized (e.g., Houk and Wise 1995; Buschman and Miller 2014).

Psychological models of performance structure can be divided into general and domain-specific models. Another distinction is made between stage or process models and continuous models (e.g., Schwarzer 2008). Whereas stage models distinguish different phases of human performance regarding long-term or

short-term control, continuous models distinguish different components of human performance, irrespective of stages. An example of a long-term stage model popular in health sciences is the transtheoretic model proposed by Prochaska et al. (2008). This model distinguishes six stages in the process of behavioral change: precontemplation, contemplation, preparation, action, maintenance, and termination. An example of a short-term stage model is the Rubicon model by Heckhausen (1991). This model tries to integrate motivational and volitional components in the preparation, initiation, maintenance, and evaluation of actions. Thus, stage models contribute to the understanding of the whole health process. They explain what happens in the respective phases and how phase transitions are established.

Examples of continuous models are the theory of planned behavior (TPB; Ajzen 1991), the self-determination theory (SDT; Ryan and Deci 2000), the social-cognitive theory (SCT; Bandura 1991), and behavioristic models, e.g., operant conditioning (OC; Abraham and Michie 2008; Lieberman 2001; Michie et al. 2011). Whereas the TPB emphasizes the influence of attitudes, subjective norms, self-efficacy, and perceived behavioral control on human intentions and behavior, SDT explains that human intrinsic motivation is substantially influenced by the human need for autonomy, competence, and social relatedness. Social relatedness is also addressed by SCT, which emphasizes the significance of role models for human behavior. Finally, OC stresses the importance of reinforcements and rewards for human behavior.

There are also attempts at integrating stages and components. A well-known example in the health domain is the health-action process approach (HAPA) by Schwarzer (2008). In this model, direct and indirect effects of outcome expectancies, risk perception and different forms of self-efficacy (i.e., concerning the initiation, maintenance, and recovery of actions) on intending, planning, and performing actions are described.

Another important type of general psychological models of human performance is proposed by action theory (e.g., Schack and Hackfort 2007). On the one hand, these models claim to be interdisciplinary in the sense that they try to integrate the above-mentioned monodisciplinary perspectives on human performance. For example, these models address social, physical, and psychological systems (respective levels) of human performance. On the other hand, human performance is analyzed based on both stage and continuous models. The basic stage model distinguishes three main stages of human performance: preparation or anticipation (including planning and calculation), realization (including processing and tuning), and interpretation (including evaluation and controlling). The cyclic nature of human action is illustrated in Fig. 10.2. The continuous model identifies three main constraints of action: Person, task and environment. This means: When an action is performed, there is a complex interaction of person, task, and environment. For example, if a person has to solve a task, the relations of the properties of the task (e.g., complexity), the person (e.g., competencies and motivation) and the environment (e.g., supportive vs. hostile) determine whether the attempt is successful or not. In case of failure, a person's competencies may not have been sufficient, the task was too difficult for the person, or the environment did not support (or even

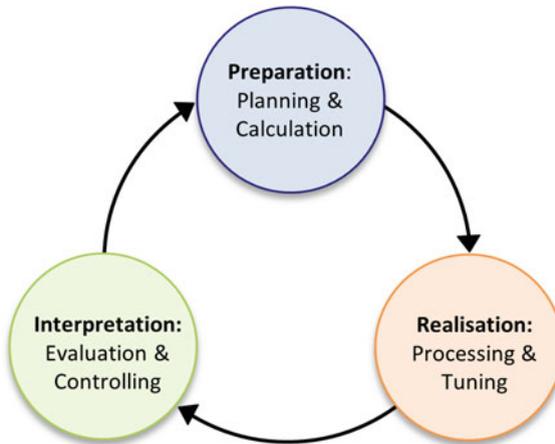


Fig. 10.2 Cyclic organization of human actions

impeded) a successful solution. The correct evaluation of these complex interactions is an important prerequisite of adequate adaptation in serious games.

Human performance can be generally analyzed from different perspectives (Fig. 10.1). On the one hand, natural sciences primarily analyze the *structure of movements*, whereas psychology analyzes the *structure of actions*. On the other hand, psychological models distinguish between *general* and *domain-specific* models as well as *stage* and *continuous* models of action.

Beyond generic aspects of human performance, every application field of serious games, every characterizing goal has its specific characteristics. Therefore, either the generic models mentioned above are adapted to the respective field, or new domain-specific performance models are developed. Taking the example of serious games for health: When the objective of the serious game is to change health-related behavior—for example, to stop or reduce smoking, to reorganize nutrition, to perform safer sexual practices or to increase habitual physical activity—the above-mentioned behavioral models are applied to derive game interventions (e.g., Lieberman 2001). Furthermore, in the area of health, numerous models have been developed addressing human behavior in particular. Examples are the Health Belief Model (HBM; Becker et al. 1977), the Health Action Process Approach (HAPA; Schwarzer 2008), and models explaining sustainable adherence to regular exercising (SARE; Wagner 2000; Fuchs et al. 2011; Williams and French 2011). Whereas the HBM explains the influences of individual cognition and motivation and modifying factors (i.e., demographic and sociopsychological variables) on the probability to initiate and maintain health-related actions, the HAPA model described above also considers intentions and planning activities. SARE

models differentiate cognitive, motivational, and volitional variables influencing human intentions and behavior in early and late(r) stages of health activities. From these models, specific concepts can be derived for online (ingame) and offline assessment, for example subjective perceptions of benefits, risks, and barriers regarding health and health-related behavior, self-efficacy, or social support by friends, other players, or relatives.

For the domain of learning, education, and training: Generic learning approaches like behavioristic, cognitive, or constructivistic models are relevant for assessment (for a review, see Kearsley 1993). An interesting approach that has been successfully integrated into educational games is the theory of micro-adaptability (Kickmeier-Rust and Albert 2010, 2012a, b). This approach is based on competence-based knowledge space theory (CbKST); this approach separates observable behavior from non-observable constructs, i.e., skills or competencies. This approach will be addressed later in this chapter.

An important model for assessment in exergames and exertion games is the structural model of sport performance. This model is illustrated in Fig. 10.3.

Figure 10.3 shows that sport performance has several components. The components that can be addressed by exergames are mainly coordination (specific skills and general abilities), condition (endurance, strength, speed, and flexibility), as well as psychological and tactical competencies. Endurance regards cardiovascular fitness. Examples of coordinative abilities are balance, sensory differentiation, and spatial orientation.

Assessment in serious games has to consider the particular genre or field of application. Therefore, either generic models are adapted to the respective field or domain-specific models are used. From these models, measures can be derived for direct or indirect online and offline assessment.

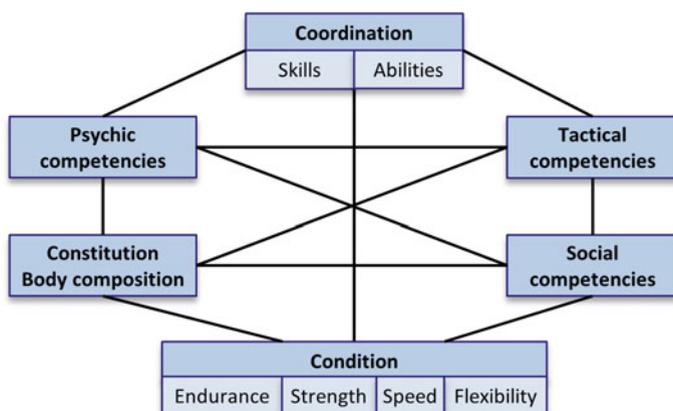


Fig. 10.3 Structure of sport performance

On the one hand, performance measures are very specific to the domain of the respective serious game. On the other hand, there are also generic performance measures.

At the spatio-temporal level the following measures can be assessed by means of biomechanical measurements (Fig. 10.4; e.g., Hay 1985; Bartlett 2007):

- *Kinematic measures*: trajectory or displacement, velocity, acceleration, joint angles, angular velocity, angular acceleration
- *Kinetic measures*: torque, angular momentum, force, work, power, and energy
- *Electrophysiological measures*: Electromyogram (EMG)

At the (neuro-)physiological level there are numerous measures located at the central or peripheral level (Fig. 10.5; Kivikangas et al. 2011; Bellotti et al. 2013):

- *Brain activation*: Electroencephalogram (EEG), Electrocorticogram (ECoG), Magnetoencephalogram (MEG), functional magnetic resonance imaging (fMRI)
- *Cardiovascular system*: electro-cardiogram (ECG), heart rate (HR), heart rate variability (HRV), blood pressure (BP)
- *Respiratory system*: breathing rate, inspiratory volume, oxygen uptake, ventilatory thresholds (VT1, VT2), respiratory exchange ratio (RER)
- *Skin*: electrodermal activity (EDA), skin conductance response (SCR), skin conductance level (SCL)
- *Visual system*: pupil diameter, pupil response

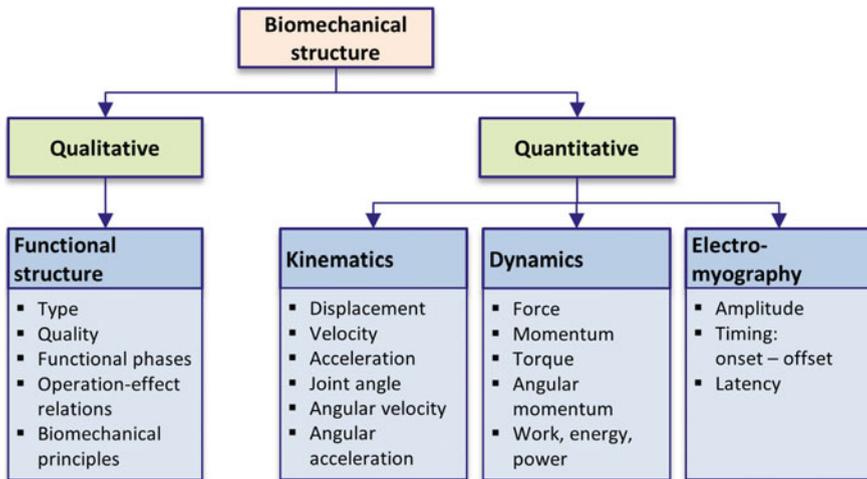


Fig. 10.4 Biomechanical analysis of human movements

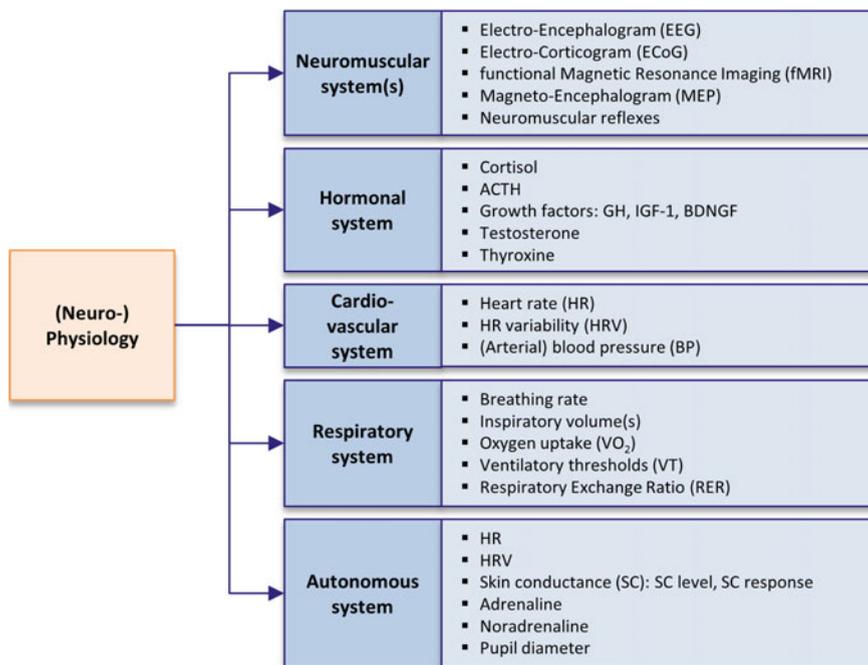


Fig. 10.5 (Neuro-)physiological measure of human performance

At the socio-psychological level, observable relations and interactions between people can be categorized, like making contact, withdrawing from contacts, cooperation, or conflicts, as well as communication behavior.

Specific performance measures pertain to the respective domain of the serious game. In Table 10.1, selected performance measures of selected domains of serious games are illustrated.

Table 10.1 Examples of domain-specific performance measures

Domain	Measures (examples)
Educational games	knowledge, attitude, skill level
Games for Health	knowledge, health-enhancing physical activity (HEPA), health-related behavior, health outcomes, attitude, fitness level
Reha(b)games	activities of daily living (ADL), various clinical scales and scores
Sport games	skill level, ability level, knowledge, sport performance
Advergames	knowledge, attitude, purchase
Persuasive games	persuasion, knowledge, attitude
Simulation and training games	“real world” behavior (transfer), strategic knowledge

In *educational games*, the educational impact is most important. Depending on the particular educational goal, performance can be measured by the application of knowledge or skill tests. Another commonly used method is attitude questionnaires.

In *games for health*, the most important performance measure is the specific impact on the respective health indicator. For example, in cardiovascular diseases (CVD), the relevant measures are heart rate, oxygen uptake, blood pressure, respiratory parameters, blood parameters, and energy expenditure. In addition, specific risk and protection factors are also important performance indicators. For CVD, these factors comprise adequate level of regular physical activity, smoking, and nutrition. In the area of health, often a distinction is made between primary and secondary outcomes.

The field of *rehabilitation games* shows a considerable variety. According to the International Classification of Diseases (ICD-10; WHO 1992), numerous diseases can affect specific anatomical regions or functions of the human body. Therefore, the primary performance measures are specific to the particular disease. As already mentioned with games for health, primary and secondary risk and protection factors can be measured, as well as the components of health, according to the definition of the WHO.

Due to the variety of sports, *sport games* also comprise numerous relevant measures. Considering the performance model illustrated in Fig. 10.3, numerous measures of conditioning and coordination can be distinguished.

Advergames aim at advertising a product or a company. Therefore, all aspects relevant to product or company marketing are candidates for measures. Of course, the most important measure is purchases of the product. Secondary performance measures are publicity, knowledge of, and attitude towards the product or company.

Persuasive games aim primarily at changing attitudes and subsequently behavior. Persuasive games can address many application fields like politics, history, social sciences, and life sciences. Beyond attitude and behavior, knowledge measures are also often applied.

Simulation and training games aim at managing a particular situation or task under time pressure. Therefore, the most important performance measure is the successful transfer to real-world behavior. Often, specific declarative and strategic knowledge is required for and acquired in simulation and training games, e.g., if-then rules for decisions.

For the adequate choice of a relevant performance measure, there is no “one size fits all” solution. Rather, either specific measures exist for particular groups—e.g., children versus adults, or healthy versus ill people—or generic measures are used, e.g., mood or motivation questionnaires. In addition, in many application domains there is no common agreement on a “gold standard.” Therefore, choosing the correct measure is not at all trivial.

10.3 Online Assessment

Online assessment refers to the idea to ground assessment on the data that can be pulled out of games, primarily the log file data. The analysis of such data, however, is not trivial. Basically, there are two challenges that must be addressed. On the one hand, there is the problem that a large amount of information coming from the game—such as movements, actions, interactions with the game world and perhaps with other players—must be analyzed in real time (cf. Koidl et al. 2010). On the other hand, there is the perhaps even more difficult challenge to interpret the data in a valid and relevant way. There have been various attempts to draw specific conclusions about motivation, or sentiment, however, the results of research are unclear, and concrete applications are still sparse (Mattheiss et al. 2010).

The main focus of online assessment in the educational community is on learning performance. This naturally leads to the concepts of *Learning Analytics* (LA) (e.g., Ferguson 2012; Siemens 2012) and *Educational Data Mining* (EDM; EDM 2015). The concepts are paired with a real-time analysis of data coming from games. The main idea is to interpret the assessment results, which often have limited utility in an educational sense, in a formatively-inspired way. This means that the aim is to make the step from diagnosing to finding the right treatment. All this is closely related to the notion of intelligent tutorial technologies that, in turn, rely on robust assessment.

EDM has been defined as “an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students and the settings which they learn in” (International Society of EDM, EDM 2015).

Baker and Yacef (2009) provided an extensive overview of EDM applications, developments, and definitions. EDM originates from many research areas, such as statistics, data mining, machine learning, visualization, and computational modeling, and aims to automatically discover patterns and models in huge and growing datasets. While in the beginnings of EDM, most data were retrieved from experimental learning sets that did not last longer than a few weeks. Today, such data are often tracked over the duration of an entire course and can last up to one year of studying. Collected data are further analyzed to gain valuable insights into learning processes. With these enormous amounts of data, new challenges arise, especially with regard to visualizing and modelling the information to make it readable and interpretable for human stakeholders.

While EDM aims to discover patterns and models in scaled data, learning analytics take into account the needs of different educational stakeholders and the strength of their judgement, in addition to computational measurements. Although EDM and LA focus on slightly different areas, they have similar goals that relate to improving educational technology and evaluating pedagogically sound instructional designs (Ferguson and Shum 2012). In particular, LA emphasizes supporting pedagogical approaches by providing assistance to teachers with practical issues (e.g., the quality of the learning material or the engagement of students

in specific exercises). Data gained by LA tools can be used to evaluate pedagogically-sound instructional designs in classroom settings. In most cases, this mainly involves monitoring learner actions, and interactions with learning tools and learning peers (Lockyer and Dawson 2011). Many attempts have been made of visualizing such learning traces either to make significant relationships explicit or to allow stakeholders to discover such relationships independently. Research on using dashboards in LAs was performed for example by Duval (2011). These dashboards are graphical representations of activities and performance data of learners. As known in other domains (e.g., sports), visualization of collected interaction training data and their comparison with data of like-minded peers may not only provide insights into poor and good practices, but also increase motivation due to the playful introduction of competitiveness.

Chatti et al. (2012) presented a reference model for LA in which they distinguished four main dimensions:

- *Who?—Stakeholders*: This dimension refers to the people targeted by the analysis.
- *Why?—Objectives*: This dimension refers to the motivation for or goals of doing the analysis.
- *What? - Data and environment*: This dimension refers to the kind of data that is gathered, managed and used for analysis.
- *How?—Methods*: This dimension relates to the techniques and tools used for performing the analysis of collected data.

In addition to these main dimensions for the domain and application of LA, Greller and Drachlser (2012) identified two additional dimensions in their approach of defining a generic framework for LA:

- *External limitations*: This dimension refers to conventions (ethics, personal privacy, socially motivated limitations) and norms (legal and organisational constraints).
- *Internal limitations*: This dimension refers to relevant human factors, like competence (e.g., interpretation, critical thinking) and acceptance that may conflict with or complicate LA.

In principle, the objectives for using LA are in line with the different views of its stakeholder groups. Chatti et al. (2012) identified the following main objectives; these certainly have overlaps, and usually a specific application of LA will serve several of them (see Fig. 10.6).

- *Monitoring and analysis*: tracking and checking the learning process, which is then used by teachers or educational institutions as a basis for taking decisions, e.g., on future steps, the design of new learning activities, improving the learning environment.

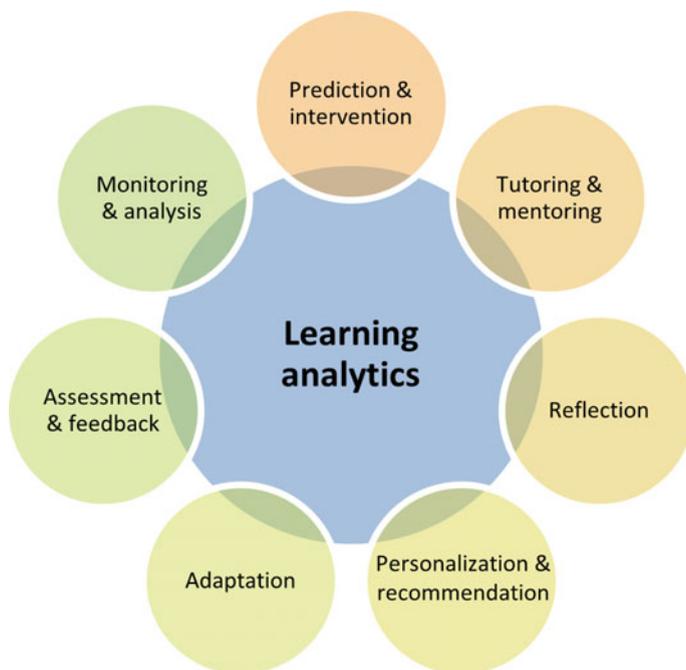


Fig. 10.6 Objectives of learning analytics

- *Prediction and intervention*: estimating learners' future knowledge or performance in terms of finding early indicators for learning success, failure, and potential dropouts, to be able to offer proactive interventions and support for learners in need of assistance.
- *Tutoring and mentoring*: helping learners with and in their whole learning process, or in the context of specific learning tasks or a course, providing guidance and advice.
- *Assessment and feedback*: supporting formative and summative (self) assessment of the learning process, examining efficiency and effectiveness of learning, and providing meaningful feedback of results to teachers and learners.
- *Adaptation*: Finding out what a learner should do or learn next and tailoring learning content, activities, or sequences to the individual. This idea of carefully calculated adjustments corresponds to the central aim and component of adaptive learning environments and intelligent tutoring systems.
- *Personalization and recommendation*: Helping learners to decide their own learning and learning environment, and what to do next by providing recommendations while leaving the control to the learner.
- *Reflection*: Prompting and increasing reflection or self-reflection on the teaching and learning process, learning progress and achievements made; providing

comparison with past experiences or achievements between learners, across classes, etc.

The methods applied in LA consider methods from data mining and analytics in general, as well as psychometrics and educational measurement as the main sources of inspiration, which fall into five main classes (see Fig. 10.7): prediction methods, structure discovery, relationship mining, discovery with models, and distillation of data for human judgment (Baker and Inventado 2014; Adomavicius and Tuzhilin 2005).

- *Prediction methods*: These are the most popular methods in EDM. They essentially aim at developing a model to predict or infer a certain variable (e.g., mark, performance score) from a combination of other indicators of the educational data set. Common prediction methods are classification (for prediction of binary or categorical variables), regression (for prediction of continuous variables), and latent knowledge estimation (assessing learner knowledge or skills).
- *Structure discovery*: Algorithms of structure discovery aim at detecting structure in educational data without an a priori assumption of what should be found (in contrast to prediction methods, where the predicted/dependent variable is known). Methods of this type are clustering (splitting data sets into clusters), factor analysis (finding dimensions of variables grouped together), and domain structure discovery (deriving the structure of knowledge in an educational domain).
- *Social network analysis (SNA)* is another method from this class, which is quite popular in LA (Siemens 2012). It allows one to analyze relationships and interactions between learners in terms of collaboration and communication activities, information exchange, etc. SNA uncovers the patterns and structure of interaction and connectivity, which can then be visually illustrated and provide the possibility of quantification (e.g., via centrality measures), to identify learners that are very important, represent “hubs,” or are in isolation (Romero 2010).
- *Relationship mining*: The aim of this group of methods is to find out relationships between variables, and how strong those relationships are. Concrete

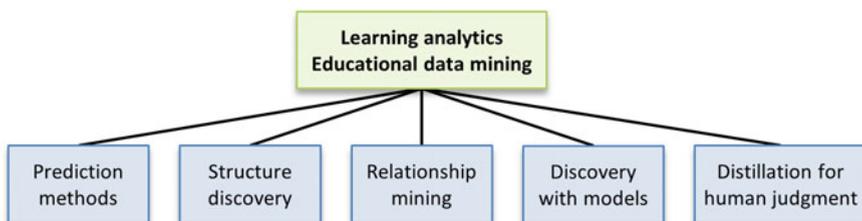


Fig. 10.7 Methods of learning analytics and educational data mining

methods are association rule mining (finding *if-then* rules), correlation mining (finding positive or negative linear correlations), sequential pattern mining (finding temporal associations between events), and causal data mining (finding out whether one observation is the cause of another).

- *Discovery with models*: This class does not denote a specific group of techniques but refers to the general approach of using the results of one analytics method within another analysis. A popular way of doing this is, for instance, to use a prediction model within another prediction model; however, there are a variety of other ways for conducting discovery with models.
- *Distillation of data for human judgment*: This is an approach quite common in LA, in a narrower sense, but not considered as a method of EDM, since it consists in providing teachers immediate access to reports and visualisations of the learner data for their interpretation, judgement and to support decision making and pedagogical action. Examples are learning curves or heat maps (Homer 2013; Baker et al. 2007).

Typically, LA is used to support teachers and instructors with deeper insights into learning processes. Games may serve as an ideal data source for LA. A crucial question is how to harness and make sense of this data in an effective and efficient manner. LA is currently in the process of initiating the elaboration of analytics that can be used for serious games. By using and combining ideas from gaming analytics, web analytics, and learning analytics, it is possible to establish meaningful analytics on data from games for educational purposes.

A great challenge with learning analytics in educational games is the wide variety of different games available, which complicates the development of analytics tools that are applicable to all games. To overcome this, Serrano-Laguna et al. (2012) propose a two-step generic approach to apply learning analytics in educational games, which is applicable to any kind of game. First, generic traces are gathered from gameplay, including game traces (start, end, and quit), phase changes (game chapters), input traces like mouse movements or clicks, and other meaningful variables like attempts or scores (depending on the game). This data gives rise to reports with general and game-agnostic information, like the number of students who played the game, the average playing time, game phases in which users stopped playing, etc. This information can be visually reported and may provide initial useful information on how learners interacted with the game. In a second step, additional information may be extracted by letting teachers define game-specific assessment rules based on and combining the generic game trace variables to obtain new information (e.g., setting maximum time thresholds, comparing actual and expected/required values of variables). These rules clearly need to be closely defined in line with each game to match the educational objectives; however, since the building blocks of these kinds of rules are elements from the basic set of traces, the creation and provision of template rules to support teachers in defining their own is conceivable.

To make use of learning analytics in educational games, a game platform needs to be used that allows the collection of the relevant data, and that holds a representation of game variables. The data for learning analytics will likely need to be stored and processed separately and remotely. To technically implement such analytics in an educational game, the definition of a learning analytics model and implementation of a learning analytics engine—which is separate from the game engine but communicates with it—has been proposed (Homer 2013). The learning analytics engine is conceived as comprising a set of modules enabling the different steps of the learning analytics process, from capturing data via aggregating and reporting, to evaluating in terms of transforming information into educational knowledge.

Assessment in a learning game may have two main purposes. First, just measuring the success of the student will serve to provide teachers and students with the derived information as a basis for action, like selecting new educational resources, deciding on additional support or learning tasks, etc. Second, the derived information may be used for realizing dynamic adaptation during game time through an adaptation model, and adapter (part of the learning analytics framework) communicating with the game engine.

An example of using learning analytics in a serious game has been presented by Baker et al. (2007); see also Miller et al. (2014). The authors also realized skill assessment in an educational action game by using game events (e.g., attacking and fleeing from an enemy) as evidence for users' mathematical skills. The authors deployed exponential empirical learning curves to determine player improvement in accuracy and speed. This approach proved useful for formative assessment in educational games, and may also be used to inform the redesign and improvement of intelligent tutoring systems. Another very recent LA attempt has been made towards elaborating an automated detector of engaged behavior in a simulation game (Stephenson et al. 2014). Their goal was to identify and model learner actions that give evidence of user engagement and, in the end, are predictive for success in the game. An integration of the engagement detector in the game will enable to report the results back to learners and teachers for reflection.

An approach to consider the structure of competencies was introduced by Kickmeier-Rust and Albert (2010); it is based on the notions of so-called competence-based knowledge space theory (CbKST) (cf., Kickmeier-Rust and Albert 2012a, b). The principal idea is to monitor each activity of a learner or a group of learner's exhibits, and to interpret the behavior in terms of available or lacking competencies or cognitive states such as motivation. Originally, this concept was developed in the European ELEKTRA project (ELEKTRA 2014) (and advanced in the following 80Days project (80Days n.d.)). In the following, generic Web services have been developed around the micro-adaptation framework. The service-oriented architecture (as described by Carvalho et al. 2015) is based on a set of recommendations, policies, and practices for the design of software architectures which implements business processes, and it is using loosely coupled components that are arranged to deliver a certain level of service or set of functionalities (Hurwitz et al. 2007). The services are (partly) available and accessible

through the service catalog platform of the Serious Games Society (Serious Games Society n.d.).

The service approach has been implemented and evaluated, for example, in primary level maths games. One example is the *Sonic Divider* (Kickmeier-Rust and Albert 2013a), a tool to practice the formal sequence of solving divisions at the level of third grade. In this case, the micro-adaptation framework builds upon a domain model that includes about 100 atomic competencies (including number dimensions, knowledge about sequences, rounding of numbers, etc.). The system identifies correct and incorrect actions of the learners and updates an underlying probability model of available competency states (in the sense of CbKST). This kind of believe model is then used to trigger highly targeted interventions (such as guidance or feedback), matching the competency levels of the learners well. The approach was also successfully applied to a multiplication game, named the *1 × 1 Ninja* which is based on a domain model for second grade multiplication skills including the number dimensions for multiplicand and multipliers. The tool can give tailored feedback, and it automatically adapts the difficulty level of the multiplication tasks to the performance of the learners. In school studies, we could show that suitable and individualized interventions are superior to no, non-individualized, or simple right/wrong statements.

A highly interesting application of the micro-adaptation framework and LA was realized in the context of the European Next-Tell project (Next Tell 2015). In this example a full teacher control suite that allows realizing educational sessions in virtual worlds (such as Second Life or OpenSim) has been developed. The tool analyzes the log files from the virtual world in real-time and—in greater detail—post hoc, and provides teachers with activity statistics, chat summaries, probabilities over competencies and competence states (based on heuristic-based analyses of activities), and real time messages (e.g., in case of unwanted activities such as using inappropriate language). In example studies with Norwegian and Austrian children who met and learned English together in an OpenSim environment, it was demonstrated that an appropriate feedback based on LA resulted in clear benefits for the teacher, who had the opportunity to monitor and document activities and, more importantly, to review language competencies (Kickmeier-Rust and Albert 2013b).

Learning Analytics (LA) and Educational Data Mining (EDM) aim at using activity and performance data from a game to draw educationally relevant conclusions. Their purposes are primarily providing learners with feedback, providing teachers and instructors with insight and overview, and allowing an adaptation of the game.

10.4 Offline Assessment

Whereas online assessment has to face challenges like real-time diagnostics and non-intrusiveness, offline assessment is much less challenging in this regard. Playing activities can be paused or finished before methods of offline assessment are administered. However, offline assessment also has to meet specific requirements. The most important requirement is to fulfill quality standards in the respective field or discipline. Unfortunately, in many application fields there is no consensus on one assessment method that represents the gold standard. Rather, many different methods are in use. Therefore, in this section the options in selected application domains of serious games are discussed without recommending particular methods. Depending on the structure of the respective field performance, knowledge, attitude and other psychological variables may be assessed on a qualitative or quantitative scale. The following fields will be addressed: health and rehabilitation, learning and education, sports and exercise, and training and simulation.

In principle, offline assessment can be qualitative or quantitative (see Fig. 10.8). Available methods comprise measurements, tests, observations, questionnaires, and (written or verbal) self-reports. Whereas measurements, tests and observations can be considered more or less direct methods of assessment, questionnaires and self-reports assess performance indirectly.

Offline Assessment in Games for Health

Due to the complexity of the health domain, numerous assessment methods are used. These methods are either based on specific health models or on a general understanding of health as a state of physical, psychological, and social well-being. Health and health-related behavior addressed in games for health comprise (regular) physical activity, nutritional behavior, perceptual-motor skills (e.g., falls, spatial vision) and abilities (e.g., balance and reaction), stress management, smoking, drug (mis-)use, asthma prevention, and safer sexual behavior (Baranowski et al. 2008; Lager and Bremberg 2005). In this section, we focus on assessment in prevention. Therapy and rehabilitation will be addressed in the next section.

In general, in the health domain the current health status has to be assessed. This can be done either by laboratory diagnostics, field tests or surveys. Well-known laboratory diagnostics are measurements of arterial blood pressure, biochemical

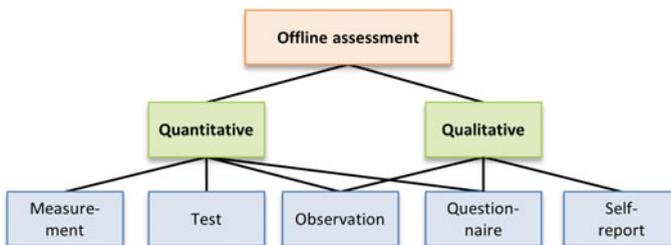


Fig. 10.8 Overview of methods for offline assessment

blood parameters like hemoglobin, liver enzymes, and other biochemical markers of organic functions. Concerning cardiorespiratory fitness, a broadly accepted indicator is the maximum oxygen uptake (VO_{2max} ; for an overview of physiological methods, see Maud and Foster 2006). This measure indicates the aerobic capacity of the cardiorespiratory system. Normally, VO_{2max} is determined by a stepwise ergometer protocol (either cycle ergometer or treadmill). Starting from a low initial load (e.g., 25 or 50 W), the load is increased stepwise until certain criteria for exhaustion are reached. For example, in the WHO protocol for children, the initial load of 25 W is stepwise increased by 25 W every two minutes (Finger et al. 2013; see also Andersen et al. 1971). Furthermore, ramp tests exist that require less time for assessment (e.g., Poole et al. 2008).

Another indicator of physical activity (PA) that has often been assessed in videogames is energy expenditure (EE; e.g., Biddiss and Irwin 2010; Peng et al. 2011, 2013; Sween et al. 2014; Deutsch et al. 2015). EE can be measured as oxygen uptake (l/min), burned calories (kcal/h) or METs (metabolic equivalents). A well accepted finding is that active videogames increase EE at a low to moderate level, both in healthy and diseased populations (Fig. 10.9).

As another instrument to assess physical activity accelerometers are often used (Reilly et al. 2008). This instrument allows for objective assessment of PA over a longer period. Furthermore, activity logs or diaries (e.g., Garcia et al. 1997) as well as specific questionnaires exist for assessing regular PA (e.g., Baecke et al. 1982; for a critical review see Shephard 2003; Prince et al. 2008). For example, the Baecke questionnaire consists of 16 items addressing PA at work, in sport, and

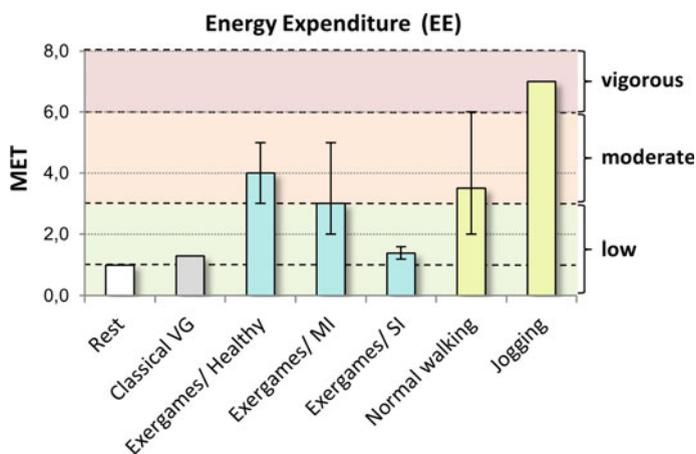


Fig. 10.9 Energy Expenditure (EE, in METs) while playing different kinds of video games versus performing locomotion activities (mean \pm minimum/maximum). *Legend* VG—Video games; *MI*—Mild neurological impairment; *SI*—Severe neurological impairment

during leisure. Based on these items three scores are calculated: work index, sport index, and leisure time index.

Concerning individual health, an often-applied instrument is the health status questionnaire for medical outcomes studies (MOS SF-36; McHorney et al. 1993; Ware and Sherbourne 1992; Ware 2000). The MOS SF-36 consists of 36 items and eight scales representing mental and physical health. It has been developed for self-administration by persons at the age of 14 years and older and for administration by trained interviewers. There also exists a short version (SF-12) with 12 items.

Nutritional knowledge and/or behavior can also be assessed by questionnaires for adults (Parmenter and Wardle 1999) and children (Wilson et al. 2008). The nutrition questionnaire for adults comprises 50 items addressing knowledge about food and nutrition, whereas the questionnaire for children consists of 14 items addressing actual nutrition behavior. A shorter 20-item questionnaire has been validated by Dickson-Spillmann et al. (2011). There also exists a computer-supported interview tool for assessing nutrition (Bakker et al. 2003). Furthermore, nutrition can be documented over a certain time period in a nutrition log or diary.

Health-related indicators of physical fitness like motor skills and abilities can be measured by specific field tests like balance, jump, or run tests. Bös (2001) gives an overview of about 700 single tests. According to Fig. 10.3 these tests can be used to assess elementary and complex motor skills, motor abilities, or conditioning abilities like strength, power, endurance, speed, and flexibility—as well as complex combinations of coordination and condition. As in the other areas, questionnaires exist asking the players to self-estimate their physical fitness level (e.g., Strøyer et al. 2007; Bös et al. 2002; Knapik et al. 1992).

Offline Assessment in Therapy and Rehabilitation

Offline assessment in therapy and rehabilitation depends on the specific disease. Beyond the primary outcomes targeted by the therapy further effects can be assessed. For example, in neurorehabilitation, the primary outcome is the improvement of mental and sensorimotor functions. Therefore, instruments like Fugl-Meyer assessment (FMA), Postural Assessment Scale (PASS), the Assessment for Motor Ability (AMA), Wolf Motor Function Test (WMFT), Stroke Impact Scale (SIS), and the Functional Independence Measure (FIM) are applied. For example, the FMA is a 226-point scale comprising five domains (motor function, sensory function, balance, joint range of motion, and joint pain; Gladstone et al. 2002). There are also domain-specific instruments like the Berg Balance Scale (for a review, see Blum and Korner-Bitensky 2008). The Berg Balance Scale has been widely used in stroke rehabilitation. The scale consists of 14 items rated from 0 to 4. The above-mentioned instruments have also been applied in the evaluation of serious games in neurorehabilitation (reviews: Staiano and Flynn 2014, Wiemeyer 2014).

Another important therapy field is cancer. In order to establish a good primary outcome, patients have to adhere to long-lasting and strenuous therapy, often including periods of self-medication and chemotherapy. Compliance to therapy can be assessed by subjective methods like self-reports and questionnaires, or by

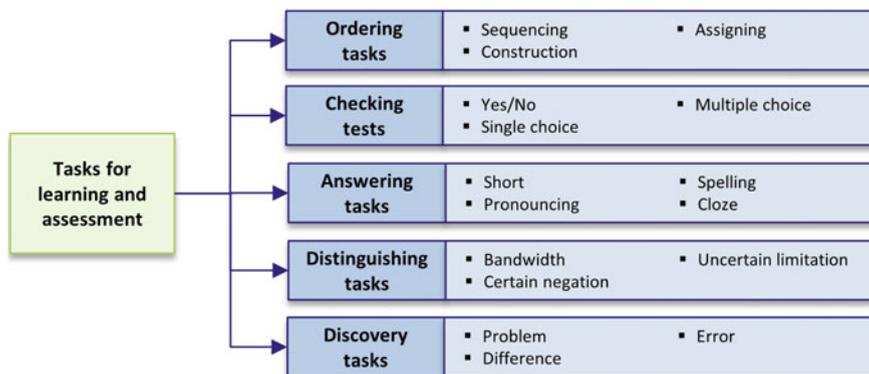


Fig. 10.10 Taxonomy of tasks for assessing learning effects (according to Meder 2006)

objective measures like blood assays and clinic visit attendance. An example of a successful application of these assessment methods is given by Kato et al. (2008), who could prove that playing a cancer-related videogame (*Re-mission*; Hope Labs 2015) has positive effects both on subjective and objective assessment scores.

Offline Assessment in Learning and Education

The primary outcome of serious games in learning and education is increased knowledge and skills. Knowledge can be declarative, e.g., the successful recall of facts, or procedural, e.g., drawing a circle or constructing a triangle. Therefore knowledge and skill tests are appropriate assessment methods. There are many approaches for assessment of learning and education in pedagogy and psychology. Three mainstreams in learning theory have already been mentioned: Behaviorism, cognitivism, and constructivism (Kearsley 1993; Egenfeldt-Nielsen 2006). Whereas *behaviorism* focusses on learning by stimulus-response connections supported by repetitions and reinforcement, *cognitivism* analyzes the information processing during learning supported by instruction and feedback. *Constructivism* states the importance of authentic learning environments and social communication. As an example of a cognitive approach, the Component Display theory by Merrill (1983) distinguishes between four types of learning content (fact, concept, procedure, principle) and three types of use (remember, use, find). From these 10 meaningful content combinations,¹ a use can be derived that can be assessed, e.g., remembering facts, using concepts or principles, or finding new procedures or principles. A similar approach comprising four types of learning content (fact, concept, procedure, meta-cognition) and six type of use (remember, understand, apply, analyze, evaluate, create) has been proposed by Krathwohl (2002). Another approach is to distinguish different kinds of tasks. In his “Web-Didaktik,” Meder (2006) develops a hierarchical taxonomy of tasks (see Fig. 10.10).

¹Note that two combinations have been removed because they do not make sense, i.e., finding and using facts.

Prominent application fields of educational games are learning mathematics, physics, geography, history, health-related behavior and languages (for a review, see Egenfeldt-Nielsen 2006). Concerning the assessment of serious games for learning and education it is not sufficient to test the primary outcomes, i.e., what has been learned by using the games. Rather, the secondary effects of educational games, i.e., how a player has learned, deserve scientific attention.

Offline Assessment in Sport and Exercise

In the fields of sport and exercise, the primary outcome is performance in specific sports or exercises. For example, in basketball the scored baskets and in soccer the scored goals can be assessed. Furthermore, according to the model of sport performance illustrated in Fig. 10.3, various subareas of competencies can be assessed: conditioning, coordination, tactics, psychological, and social competencies. For example, numerous tests for strength, endurance, flexibility, and speed as well as for tactical behavior, coordinative abilities, and sport skills have been developed. For example, commonly used performance measures of endurance are the time needed to run a certain distance (e.g., 3,000 m), or the distance covered within a prescribed time (e.g., 12 min in the Cooper test). An example of a power test is the jump-and-reach test, where the task is to jump as high as possible from a standing position.

Furthermore, sport psychology has validated many generic or sport-specific tests for cognition, perception, emotions, motivation and volition (e.g., Tenenbaum and Eklund 2007).

Particularly sensorimotor skill and knowledge tests have been assessed in studies on serious games in sport and exercise (for a review, see Wiemeyer and Hardy 2013).

Offline Assessment in Training and Simulation

In a way, training and simulation is a specific form of learning. Training denotes all measures aiming at the systematic, purposeful and sustainable change of human competencies and behavior. Simulations are manipulations applied to a physical or computational model. Simulations are used when it is too expensive, unethical or simply impossible to perform experiments with the original. Therefore, the most important outcome of learning and training with simulations is the transfer to the “real world” situation. This transfer can be direct, i.e., showing the behavior acquired in the simulation immediately in the real world situation, or indirect, i.e., gaining knowledge about principles or strategies to facilitate the transfer. As a consequence, methods for direct and indirect assessment of training and simulation can be applied. Unfortunately, existing reviews and meta-analyses do not distinguish between these two kinds of assessment (Lee 1999; Vogel et al. 2006). However, they prove that appropriately instructed, engaged, and playful use of simulations may enhance performance.

Offline assessment of human performance is specific to the respective application field of serious games. In this regard, many assessment methods exist ranging from measurements via tests and observations to reports and surveys. In many

fields, no “gold standard” has been agreed upon. Therefore, the great challenge is to select the method(s) that meet best the requirements for the particular offline assessment.

10.5 Performance Assessment and Game Adaptation

As pointed out in the chapter introduction, performance assessment serves numerous functions in serious games. One important function is the adaptation of the game to changes in the current performance of the player (see also Chap. 7). First, adaptation of a game has to consider more or less static characteristics of the player like age, gender, experience level, level of expertise, etc. In addition, rehabilitation games have to adapt to the degree of impairment. This type of adaptation is called *macro-adaptation* (e.g., Kickmeier-Rust and Albert 2012a, b). A second type of adaptation is much more dynamic and depends on the current state of the game. This type of adaptation, called *micro adaptation*, depends very much on online performance assessment. In Fig. 10.11, the relation of performance assessment and micro adaptation is illustrated.

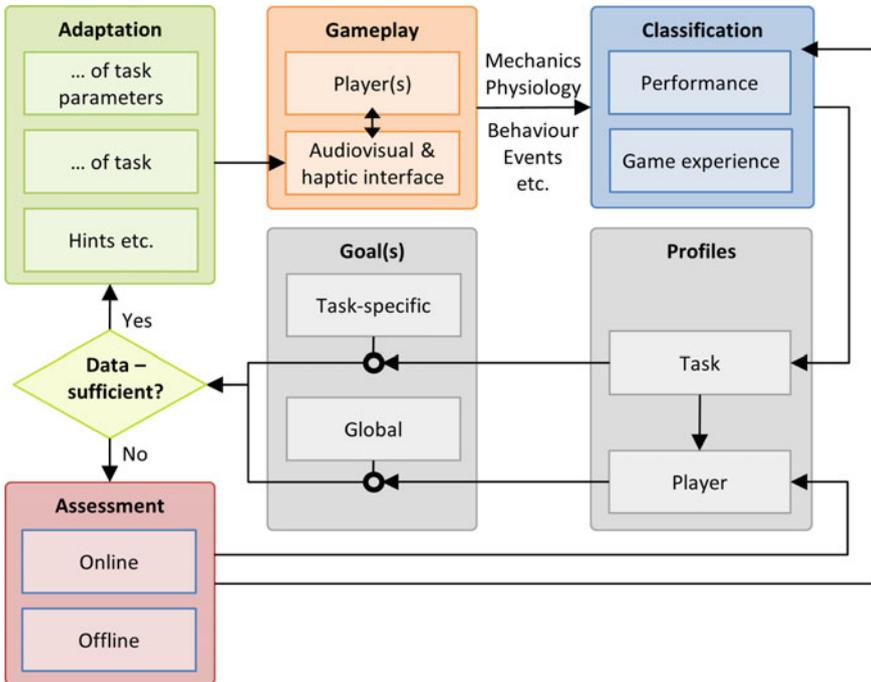


Fig. 10.11 The relation of performance assessment and adaptation in serious games

During gameplay the player interacts with the audiovisual and haptic interface of the game. Depending on the sensors used, different sensory information can be recorded as a result of the player-game interaction. For example, the kinematics and dynamics of the player's movements like forces or acceleration can be measured, as well as the physiological reactions like heart rate or energy expenditure. Furthermore, events like pressed buttons or keys can be registered. This sensor data can be used both for performance classification and player experience classification (see Chap. 9). For example, the biomechanical data may signify a particular movement error that has to be corrected. Or the physiological data may indicate that the emotional arousal of the player is decreasing. Based on the results of the classification, the task profile—and consequently the player profile—can be updated. These profiles are then compared to the task-specific and global goals, respectively. If the data suggests an adaptation of the game, decisions have to be taken whether to change the task as such, adjust task parameters, or to use other means of adaptation, e.g., hints or encouragements. If the data is not sufficient for an adaptation, a decision has to be made whether to initiate further online, i.e., in-game assessment, or off-line assessment. In order to not disrupt the game, online assessment is preferred. However, the options for online assessment may be not sufficient to get valid information concerning the current performance state. For example, this may happen in serious games for rehabilitation when the performance of the patient does not improve. To identify potential causes of the stagnation, a thorough clinical examination may be required.

10.6 Summary and Questions

Performance assessment in serious games is important for several reasons, for example to adapt the game dynamically to progress of the player(s) and to evaluate game quality. Performance denotes the process and result of actions and interactions within a game. The numerous approaches to the structure of performance can be distinguished into movement-based versus action-based approaches. Furthermore, different levels, stages, and components of performance regulation can be distinguished. Assessment of different aspects of human performance can be based on generic or domain-specific models. Assessment can be performed online, i.e., during gameplay, or offline, i.e., at the end of playing the game. Due to the different levels and components of performance, biomechanical, (neuro-)physiological, observational, and psychological methods are available for assessment.

Online assessment poses major challenges because assessment has to be done in realtime and without disturbing the ongoing game. Using the example of games for education and learning, the approaches of learning analytics (LA) and educational data mining (EDM) have been introduced to illustrate the demands on online assessment.

Offline assessment has to consider the methodological standards in the respective application field. In the health domain, specific options of measuring, observing, testing, or self-reporting health-related activities and knowledge are available.

In education and learning domains, as well as in simulation and training domains, the most important requirement is to assess learning and transfer, respectively. In sport and exercise domains, assessment focuses on the whole or parts of the complex structure of sport and exercise performance.

Check your understanding of this chapter by answering the following questions:

- What does the concept of “human performance” mean?
- What is the difference between movement-oriented and action-oriented approaches to human performance?
- Which stages are distinguished by the action theory of human performance?
- Which components are distinguished in the generic model of sport performance?
- Which metrics are assessed by biomechanical analysis of performance?
- Which metrics are assessed by (neuro-)physiological analysis of performance?
- What does the term “learning analytics” mean?
- What is meant by “educational data mining?”
- What are the objectives of learning analytics?
- What are the methods of learning analytics and educational data mining?
- What are the specific challenges in online assessment?
- What are the options for offline assessment in the field of health?
- What are the options for offline assessment in the field of rehabilitation?
- What are the options for offline assessment in the field of learning and education?
- What are the options for offline assessment in the field of training and simulation?
- What are the options for offline assessment in the field of sport and exercise?
- How can online and offline assessment be integrated in the adaptation of serious games?

Recommended Literature

- GALA (ed) (2014) Learning analytics for and in serious games. http://css-kmi.tugraz.at/mkrwww/leas-box/downloads/ectel14_booklet.pdf. Accessed 18 Feb 2016—*This booklet gives an overview of current issues and methods in the field of learning analytics in serious games*
- Ifenthaler D, Eseryel D, Ge X (eds) (2012) Assessment in game-based learning: Foundations, Innovations, and Perspectives. Springer, New York—*This book provides an overview of assessment in games. Different assessment methods and methodologies are introduced and discussed concerning their specific strengths and weaknesses. Technological and methodological approaches are addressed, as well as issues of practical application of assessment in games and game-based learning*
- Larsson JA & White B (eds) (2014) Learning Analytics—From Research to Practice. Springer, New York—*This book is an introduction to learning analytics: Methods, theory, and practical applications are addressed*

References

- 80Days (n.d.) 80Days—Around an inspiring virtual learning world in eighty days. www.eightydays.eu. Accessed 18 Feb 2016
- Abraham C, Michie S (2008) A taxonomy of behavior change techniques used in interventions. *Health Psychol* 27(3):379–387
- Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans Knowl Data Eng* 17:734–749
- Ajzen I (1991) The theory of planned behaviour. *Organ Behav Hum Dec* 50:179–211
- Andersen KL, Shephard RJ, Denolin H, Varnauskas E, Masironi R (1971) Fundamentals of exercise testing. WHO, Geneva
- Baecke JA, Burema J, Frijters JE (1982) A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr* 36(5):936–942
- Baker R, Inventado PS (2014) Educational data mining and learning analytics. In: Larusson JA, White B (eds) *Learning analytics: from research to practice*. Springer, New York, Germany, pp 61–75
- Baker R, Yacef K (2009) The state of educational data mining in 2009: a review and future visions. *JEDM* 1(1):3–17
- Baker RS, Habgood MJ, Ainsworth SE, Corbett AT (2007) Modeling the acquisition of fluent skill in educational action games. *Proceedings of User Modeling 2007*. Springer, Berlin, pp 17–26
- Bakker I, Twisk JWR, Van Mechelen W, Mensink GBM, Kemper HCG (2003) Computerization of a dietary history interview in a running cohort: evaluation within the Amsterdam growth and health longitudinal study. *Eur J Clin Nutr* 57(3):394–404
- Bandura A (1991) Social cognitive theory of self-regulation. *Organ Behav Hum Dec* 50 (2): 248–287
- Baranowski T, Buday R, Thompson DI, Baranowski J (2008) Playing for real: video games and stories for health-related behavior change. *Am J Prev Med* 34(1):74–82
- Bartlett R (2007) *Introduction to sports biomechanics: analysing human movement patterns*, 2nd edn. Routledge, London
- Becker MH, Haefner DP, Kasl SV, Kirscht JP, Maiman LA, Rosenstock IM (1977) Selected psychosocial models and correlates of individual health-related behaviors. *Med Care* 15(5): 27–46
- Bellotti F, Kapralos B, Lee K, Moreno-Ger P, Berta R (2013) Assessment in and of serious games: an overview. *Adv Hum-Comput Interact* 2013:136864. doi:10.1155/2013/136864
- Biddiss E, Irwin J (2010) Active video games to promote physical activity in children and youth: a systematic review. *Arch Pediat Adol Med* 164(7):664–672
- Blum L, Korner-Bitensky N (2008) Usefulness of the berg balance scale in stroke rehabilitation: a systematic review. *Phys Ther* 88(5):559–566
- Bös K (ed) (2001) *Handbuch motorische tests*. [Handbook of perceptual-motor tests] Hogrefe, Göttingen
- Bös K, Abel T, Woll A, Niemann S, Tittlbach S, Schott N (2002) Der Fragebogen zur Erfassung des motorischen Funktionsstatus (FFB-Mot). [Questionnaire for assessing the state of motor functions]. *Diagnostica* 48(2):101–111
- Buschman TJ, Miller EK (2014) Goal-direction and top-down control. *Philos Trans R Soc B* 369 (1655):0130471
- Carvalho MB, Bellotti F, Berta R, De Gloria A, Gazzarata G, Hu J, Kickmeier-Rust M (2015) The journey: a service-based adaptive Serious Game on probability. *Entertainment Comput* 6:1–10
- Chatti MA, Dyckhoff AL, Schroeder U, Thüs H (2012) A reference model for learning analytics. *IJTEL* 5:318–331
- Deutsch JE, Guarrera-Bowly P, Myslinski MJ, Kafri M (2015) Is there evidence that active videogames increase energy expenditure and exercise intensity for people poststroke and with cerebral palsy? *Games Health J* 4(1):31–36

- Dickson-Spillmann M, Siegrist M, Keller C (2011) Development and validation of a short, consumer-oriented nutrition knowledge questionnaire. *Appetite* 56(3):617–620
- Duval E (2011) Attention please! Learning analytics for visualization and recommendation. In: Proceedings of 1st international conference on learning analytics and knowledge. ACM, New York, pp 9–17
- Egenfeldt-Nielsen S (2006) Overview of research on the educational use of video games. *Digital kompetanse* 1(3):184–213
- EDM (2015) International data mining society. <http://www.educationaldatamining.org/>. Accessed 18 Feb 2016
- ELEKTRA (2014) ELECTRA—Enhancing learning in ENPI countries through clean technologies and research related activities. <http://kti.tugraz.at/css/projects/elektra/>. Accessed 18 Feb 2016
- Ferguson R (2012) Learning analytics: drivers, developments and challenges. *IJTEL* 4(5/6): 304–317
- Ferguson R, Shum SB (2012) Social learning analytics: five approaches. In: Proceedings of 2nd international conference on learning analytics and knowledge. ACM, New York, pp 23–33
- Finger J, Gößwald A, Härtel S, Müters S, Krug S, Hölling H, Kuhnert R, Bös K (2013) Measurement of cardiorespiratory fitness in the German health interview and examination survey for adults (DEGS1). *Bundesgesundheitsbl* 56:885–893
- Fuchs R, Goehner W, Seelig H (2011) Long-term effects of a psychological group intervention on physical exercise and health: The MoVo concept. *J Phys Act Health* 8(6):794–803
- Garcia AW, George TR, Coviak C, Antonakos C, Pender NJ (1997) Development of the child/adolescent activity log: a comprehensive and feasible measure of leisure-time physical activity. *Int J Behav Med* 4(4):323–338
- Gladstone DJ, Danells CJ, Black SE (2002) The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehab Neural Re* 16(3):232–240
- Greller W, Drachler H (2012) Translating learning into numbers: a generic framework for learning analytics. *JETS* 15:42–57
- Hay JG (1985) *The biomechanics of sports techniques*. Prentice Hall, New Jersey
- Heckhausen H (1991) Volition: Implementation of intentions. In: Heckhausen H (ed) *Motivation and Action*. Springer, Berlin, pp 163–188
- Homer BD (2013) Overview. Introductory Talk to the Learning Analytics and Educational Data Mining Workshop,” CREATE Lab, New York University, April 2013 http://create.nyu.edu/wordpress/wp-content/uploads/2013/04/01.HOMER_LEARNING_ANALYTICS.WORKSHOP.04.04.13.pdf. Accessed 18 Feb 2016
- Hope Labs (2015) Re-Mission. www.re-mission.net. Accessed 18 Feb 2016
- Houk JC, Wise SP (1995) Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. *Cereb Cortex* 5(2):95–110
- Hurwitz J, Bloor R, Baroudi C, Kaufman M (2007) *Service oriented architecture for dummies*. Wiley, Indianapolis
- Ifenthaler D, Eseryel D, Ge X (eds) (2012b) *Assessment in game-based learning: foundations, innovations, and perspectives*. Springer, New York
- Kato PM, Cole SW, Bradlyn AS, Pollock BH (2008) A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. *Pediatrics* 122:e305–e317
- Kearsley G (1993) Hypertext as a tool for the metatheoretical analysis of learning theories: the TIP database. *JCHE* 4(2):43–56
- Kickmeier-Rust MD, Albert D (2010) Micro adaptivity: protecting immersion in didactically adaptive digital educational games. *J Comput Assist Lear* 26:95–105
- Kickmeier-Rust M, Albert D (2012a) Educationally adaptive: balancing Serious games. *Int J Comput Sci Sport* 11(1):15–28
- Kickmeier-Rust MD, Albert D (eds) (2012b) *An Alien’s guide to multi-adaptive educational games*. Informing Science Press, Santa Rosa, CA

- Kickmeier-Rust MD, Albert D (2013a) Gamification and intelligent feedback mechanisms for a division learning tool. In: Escudeiro P, de Carvalho CV (eds) Proceedings of 7th European conference on games based learning, Oct 2–3, 2013, Porto, Portugal, pp 290–296
- Kickmeier-Rust MD, Albert D (2013b) Learning analytics to support the use of virtual worlds in the classroom. In: Holzinger A, Pasi G (Eds) Human-computer interaction and knowledge discovery in complex, unstructured, big data. lecture notes in computer science, 7947, Springer, Berlin, pp 358–365
- Kivikangas JM, Chanel G, Cowley B, Ekman I, Salminen M, Järvelä S, Ravaja N (2011) A review of the use of psychophysiological methods in game research. *JGVW* 3(3):181–199
- Knapik JJ, Jones BH, Reynolds KL, Staab JS (1992) Validity of self-assessed physical fitness. *Am J Prev Med* 8(6):367–372
- Koidl K, Mehm F, Hampson C, Conlan O, Göbel S (2010) Dynamically adjusting digital educational games towards learning objectives. In: Proceedings of 3rd European Conference on Game Based Learning (ECGBL), October 2010
- Krathwohl DR (2002) A revision of Bloom's taxonomy: an overview. *Theor Pract* 41(4):212–218
- Lager A, Bremberg S (2005) Health effects of video and computer game playing. A systematic review. Swedish National Institute of Public Health, Stockholm
- Lee J (1999) Effectiveness of computer-based instructional simulation: a meta analysis. *Int J Instr Media* 26(1):71–85
- Lieberman DA (2001) Management of chronic pediatric diseases with interactive health games: theory and research findings. *J Ambul Care Manage* 24(1):26–38
- Lockyer L, Dawson S (2011) Learning designs and learning analytics. In: Proceedings of 1st international conference on learning analytics and knowledge. ACM, New York, pp 153–156
- Mattheiss E, Kickmeier-Rust MD, Steiner CM, Albert D (2010) Approaches to detect discouraged learners: assessment of motivation in educational computer games. In: Proceedings of eLearning Baltics (eLBa) 2010, July 1–2, 2010, Rostock, Germany
- Maud PJ, Foster C (2006) Physiological assessment of human fitness. Human Kinetics, Champaign, Illinois
- McHorney CA, Ware JE Jr, Raczek AE (1993) The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 31(3):247–263
- Meder N (2006) Web-Didaktik—Eine neue Didaktik webbasierten, vernetzten Lernens. [Web didactics—a new didactics of web-based networked learning] Bertelsmann, Bielefeld
- Merrill MD (1983) Component display theory. In: Reigeluth CM (ed) Instructional-design theories and models: an overview of their current status. Erlbaum, Hillsdale, NJ, pp 279–333
- Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP (2011) A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. *Psychol Health* 26(11):1479–1498
- Miller WL, Baker RS, Rossi LM (2014) Unifying computer-based assessment across conceptual instruction, problem-solving, and digital games. *Technol Knowl Learn* 19:165–181
- Next Tell (2015) Next-Tell. Next generation teaching, education and learning for life. <http://next-tell.de/>. Accessed 18 Feb 2016
- Parmenter K, Wardle J (1999) Development of a general nutrition knowledge questionnaire for adults. *Eur J Clin Nutr* 53(4):298–308
- Peng W, Lin J-H, Crouse JC (2011) Is playing exergames really exercising? A meta-analysis of energy expenditure in active video games. *Cyberpsychol Behav Soc Netw* 14(11):681–688
- Peng W, Crouse JC, Lin JH (2013) Using active video games for physical activity promotion: a systematic review of the current state of research. *Health Educ Behav* 40(2):171–192
- Poole DC, Wilkerson DP, Jones AM (2008) Validity of criteria for establishing maximal O₂ uptake during ramp exercise tests. *Eur J Appl Physiol* 102(4):403–410
- Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M (2008) A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act* 5(1):56

- Prochaska JO, Redding CA, Evers K (2008) The transtheoretical model and stages of change. In: Glanz K, Lewis FM, Rimer BK (eds) *Health behavior and health education*, 4th edn. Jossey-Bass, San Francisco, pp 97–121
- Reilly JJ, Penpraze V, Hislop J, Davies G, Grant S, Paton JY (2008) Objective measurement of physical activity and sedentary behaviour: review with new data. *Arch Dis Child* 93(7): 614–619
- Romero C (2010) Educational data mining: a review of the state of the art. *IEEE Trans Syst Man Cybern C* 40:601–618
- Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol* 55(1):68–78
- Schack T, Hackfort D (2007) Action theory approach to applied sport psychology. In: Tenenbaum G, Eklund RC (eds) *Handbook of sport psychology*, 3rd edn. John Wiley and Sons, New York, pp 332–351
- Schwarzer R (2008) Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. *Appl Psychol* 57(1):1–29
- Serious Games Society (n.d.). Serious games web services catalog. <http://services.seriousgamessociety.org/>. Accessed 18 Feb 2016
- Serrano-Laguna A, Torrente J, Moreno-Ger P, Fernández-Manjón B (2012) Tracing a little for big improvements: application of learning analytics and videogames for student assessment. *Procedia Comput Sci* 15:203–209
- Shephard RJ (2003) Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med* 37(3):197–206
- Siemens G (2012) Learning analytics: envisioning a research discipline and a domain of practice. In: *Proceedings of international conference on learning analytics and knowledge*, ACM, New York, pp 4–8
- Staiano AE, Flynn R (2014) Therapeutic uses of active videogames: a systematic review. *Games Health J* 3(6):351–365
- Stephenson S, Baker R, Corrigan S (2014) Towards building an automated detector of engaged and disengaged behavior in game-based assessments. In: *Poster, annual conference on games + learning + society*
- Strøyer J, Essendrop M, Jensen LD, Warming S, Avlund K, Schibye B (2007) Validity and reliability of self-assessed physical fitness using visual analogue scales. *Percept Mot Skills* 104(2):519–533
- Sween J, Wallington SF, Sheppard V, Taylor T, Llanos AA, Adams-Campbell LL (2014) The role of exergaming in improving physical activity: a review. *J Phys Act Health* 11(4):864–870
- Tenenbaum G, Eklund RC (eds) (2007) *Handbook of sport psychology*. Wiley, New York, NY
- Vogel JJ, Vogel DS, Cannon-Bowers J, Bowers JC, Muse K, Wright M (2006) Computer gaming and interactive simulations for learning: a meta-analysis. *J Educ Comput Res* 34(3):229–243
- Wagner P (2000) Aussteigen oder Dabeibleiben? [Drop out or continue?] *Wissenschaftliche Buchgesellschaft*, Darmstadt
- Ware JE Jr (2000) SF-36 health survey update. *Spine* 25(24):3130–3139
- Ware JE Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 30(6):473–483
- Wiemeyer J (2003) Function as constitutive feature of movements in sport. *Int J Comput Sci Sport* 2(2):113–115
- Wiemeyer J (2014) Serious games in neurorehabilitation—a review of recent evidence. *SeriousGames'14, Proc. ACM international workshop on serious games*. ACM, New York, pp 33–38
- Wiemeyer J, Hardy S (2013) Serious games and motor learning—concepts, evidence, technology. In: Bredl K, Bösche W (eds) *Serious games and virtual worlds in education, professional development and healthcare*. IGI Global, Hershey, PA, pp 197–220

- Williams SL, French DP (2011) What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour—and are they the same? *Health Educ Res* 26(2):308–322
- Wilson AM, Magarey AM, Mastersson N (2008) Reliability and relative validity of a child nutrition questionnaire to simultaneously assess dietary patterns associated with positive energy balance and food behaviours, attitudes, knowledge and environments associated with healthy eating. *Int J Behav Nutr Phys Act* 5(1):5
- World Health Organization [WHO] (1992) The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. World Health Organization, Geneva