Designing CbKST-based Adaptive Educational Game Levels: A Walk Through by the Example of *Physics Playground*

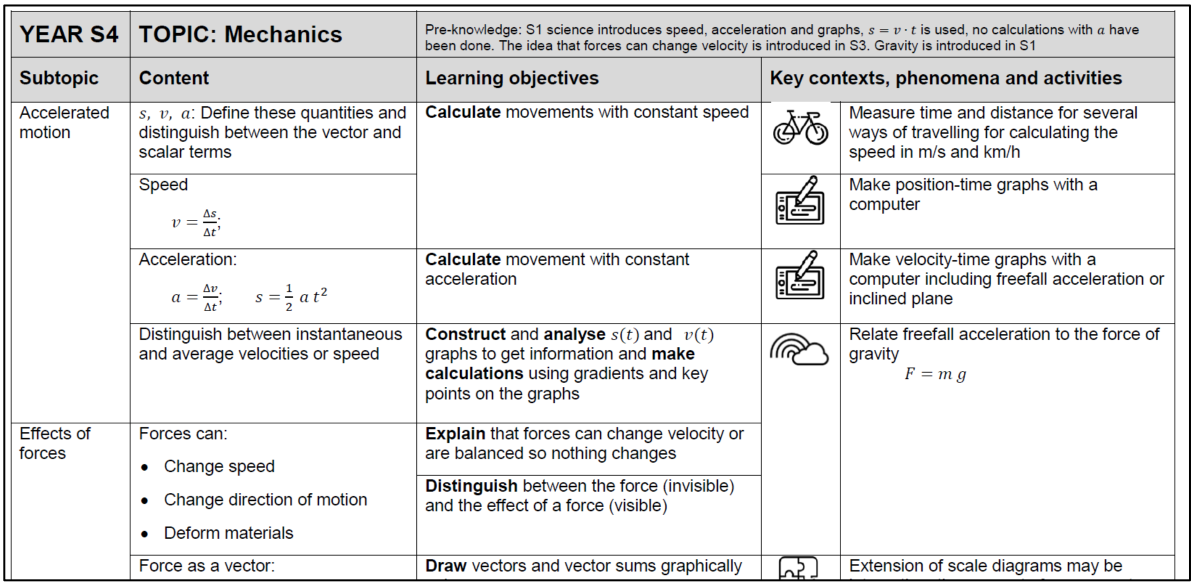
Michael D. Kickmeier-Rust, April 2020

This walk through exemplifies the design process for digital educational games, featuring *Stealth Assessment* and adaptation based on *Competence-based Knowledge Space Theory* (CbKST) and its advancements *Micro Adaptivity* and *Micro Learning Spaces*. This example includes the process of generating underlying domain, task, adaptation, assembly, and learner models and illustrates how concrete design decisions will be made alongside these models. The walk through is based on the game *Physics Playground* (Shute et al., 2020). Conceptually, it combines the original design and modeling processes of the game, that is, *Evidence Centered Design* (ECD; Mislevy, Steinberg, & Almond, 2003) and Stealth Assessment (Shute & Ventura, 2013), with the procedures and models typical for the CbKST approach.

Adaptive tutoring systems usually build upon formal, computable models that hold information about the domain, tasks (e.g., game levels), and the learner as well as models that hold information about assessment and adaptation processes. These models are developed by experts and/or informed by the reasoning over data (e.g., performance data of students). While the models are, in the first instance, hypotheses, these hypotheses need to be evaluated and validated based on performance data, and iteratively revised. The models serve as a scaffolding to guide the concrete design of the elements of a tutoring system (e.g., learning objects, assessment objects, tasks, learning supports). It shall be noted that Physics Playground is a level-based game that does not feature a specific narrative or that requires sophisticated artwork.

Step 1: The Domain Model

The first step in the design process is the development of a domain model. This process starts with the specification of learning goals of the game and thus the concrete domain. In this example the physics sub-domain “force and motion”. For this (sub-)domain the existing curricula and domain models are reviewed (Figures 1a-c) and possibly learning media and typical teaching scenes (see Box 1). Specifically interactive environments such as the MIT’s RELATE program may provide the modelling and design process with great insights into technology-enhanced physics pedagogy and the methods of an accurate assessment of physics competencies. Interactive learning environments for physics such as computer simulations and games are a promising educational genre because they can support high-fidelity modelling of physical concepts and processes. Virtual physics laboratories (e.g., the PHET of the University of Colorado at Boulder) provide inspiration and represent the state of the art.



YEAR S4	**TOPIC: Mechanics**	Pre-knowledge: S1 science introduces speed, acceleration and graphs, $s = v \cdot t$ is used, no calculations with a have been done. The idea that forces can change velocity is introduced in S3. Gravity is introduced in S1	
Subtopic	**Content**	**Learning objectives**	**Key contexts, phenomena and activities**
Accelerated motion	$s,\ v,\ a$: Define these quantities and distinguish between the vector and scalar terms	**Calculate** movements with constant speed	Measure time and distance for several ways of travelling for calculating the speed in m/s and km/h
	Speed $v = \frac{\Delta s}{\Delta t}$;		Make position-time graphs with a computer
	Acceleration: $a = \frac{\Delta v}{\Delta t}$; $s = \frac{1}{2}\, a\, t^2$	**Calculate** movement with constant acceleration	Make velocity-time graphs with a computer including freefall acceleration or inclined plane
	Distinguish between instantaneous and average velocities or speed	**Construct** and **analyse** $s(t)$ and $v(t)$ graphs to get information and **make calculations** using gradients and key points on the graphs	Relate freefall acceleration to the force of gravity $F = m\, g$
Effects of forces	Forces can: • Change speed • Change direction of motion • Deform materials	**Explain** that forces can change velocity or are balanced so nothing changes	
		Distinguish between the force (invisible) and the effect of a force (visible)	
	Force as a vector:	**Draw** vectors and vector sums graphically	Extension of scale diagrams may be

Figure 1a. European Schools (2020)[1] physics curriculum.

1 Die Schülerinnen und Schüler können Bewegungen und Wirkungen von Kräften analysieren.

Querverweise

Physik: Bewegungen und Kräfte

NT.5.1

Die Schülerinnen und Schüler ...

3

a » können gleichförmige Bewegungen von Körpern in Diagrammen erkennen und darstellen.

b » können Wirkungen von Kräften untersuchen und beschreiben (z.B. verformte Plastilinkugel nach dem Herunterfallen, Bedeutung der Gurte beim Autofahren, Veränderung der Flugbahn eines Balls durch Krafteinwirkung). ☰ Angriffspunkt, Richtung und Betrag einer Kraft; Verformung, Bewegungs- und Lageänderungen durch Krafteinwirkung

c » können experimentell zeigen und in Diagrammen darstellen, dass die Gewichtskraft proportional zur Masse ist. ☰ Umgang mit einem Kraftmesser

d » können Kräfte einordnen und darstellen. ☰ Kräftediagramm

» können experimentell zeigen, dass bei einfachen Maschinen die benötigten Kräfte verringert werden können (z.B. Hebel, schiefe Ebene, Flaschenzug, Ketten-/Zahnradgetriebe).

» können begründen, dass bei einfachen Maschinen die benötigten

Figure 1b. Snap shot of the Swiss Lehrplan 21.

---

[1] https://www.eursc.eu/Syllabuses/Forms/Syllabuses.aspx

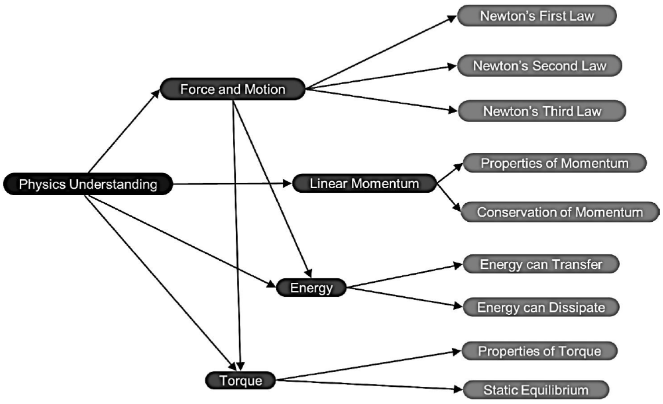


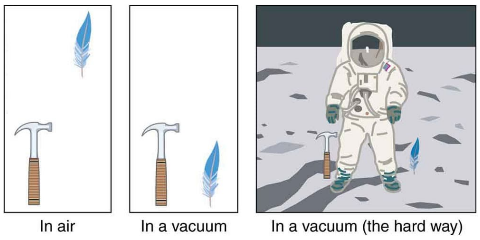
Figure 1c. Competence model of Physics Playground (Shute et al., 2020).

In addition, possible misconception the domain are identified in the research literature (e.g., Kuczmann, 2017). In this walk-through, I want to refer to the common misconception that heavier objects fall faster than lighter ones (Box 1). This misconception is already described by Galileo when he dropped two objects with different masses on the Leaning Tower of Pisa. He has shown on that experiment that objects with different masses fall with the same acceleration. The problem lies in the identification of another force that is involved, which is air resistance. All objects moving through air, and hence, all falling objects, experience air resistance. This force is proportional to the area of the object in the direction of motion. Usually, this force is negligible, but for light objects, it has an effect. This physical law has been illustrated by NASA's famous hammer and feather drop experiment on the moon in 1971[2].

This process leads to a classification/categorization of involved concepts and a detailed list of fine-grained competencies (Table 1)

Box 1: Topic gravity – teaching scenes

Gravity does not act as an equal force to every object, Newton's gravitational law tells us that the more massive a body is, the more force will experiment due to gravity. It turns out that the two effects (more mass implies more force, but also less acceleration) cancel out, so the total acceleration due to gravity is exactly the same for all the bodies. This principle can be demonstrated by dropping objects with different mass and observe acceleration. On Earth, however, another force acts on an object, that is air resistance. This easily leads to a common misconceptions, namely the assumption that heavier objects are more affected by gravity than lighter ones. Teaching scenes for this phenomenon are, for example, Galileo's Leaning Tower of Pisa experiment or NASA's famous ball and hammer drop experiment on the moon[3] .



---

[2] https://www.youtube.com/watch?v=MqJoobBvKBk
[3] https://nssdc.gsfc.nasa.gov/planetary/lunar/apollo_15_feather_drop.html

Table 1. Example competencies involved in “force and motion”,
concretely the famous hammer drop experiment on the moon (Box 1).

Topic	Description	ID
Accelerated motion	s,v,a: Define these quantities and distinguish between the vector and scalar term	a
	Speed v=$\Delta s\Delta t$;	b
	Acceleration: a=$\Delta v\Delta t$; s=12 a $t2$	c
	Distinguish between instantaneous and average velocities or speed	d
Effects of forces	Forces can: - Change speed - Change direction of motion - Deform materials	e, f, g
	Force as a vector: - Summing forces in 1 dimension - Extension of the concept of a sum of forces (resultant) to 2 dimensions	h, i
	Examples of common forces: - Gravitational force - Tension - Normal force - Friction force	j
Forces in action	Newton’s 1st law: Every body persists in its state of being at rest or of moving uniformly straight forward, except insofar as it is compelled to change its state by force impressed	k
	Newton’s 2nd law: The alteration of motion is proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed	l
	Consider mass as a measure of how easy or hard it is for a given force to change the motion of an object	m
	Newton’s 3rd law: To every action there is always opposed an equal reaction	n
Air resistance	Consider air resistance as a force impacting the acceleration of an object	o
	- Increased speeds result in an increased amount of air resistance. - Increased cross-sectional areas result in an increased amount of air resistance.	p, q
	Objects in free fall in air reach a terminal velocity because the force of air resistance becomes large enough to balances the force of gravity.	r
Additional	On moon, no air resistance exists.	s

Based on the list of atomic competencies, the CbKST related prerequisite relation can be established (Figure 2, left panel) and subsequently the competence structure can be derived (Figure 2, right panel).

Prerequisite relations, in essence, determine the order in which competencies need to be taught/learned. There are other notions of relations in CbKST such as “surmise relation”, which means from solving a particular item one can surmise (assume) that other, easier items can be solved as well. Specially in applied settings, the term “prerequisite” is often used less strict with the meaning of “is usually learned before”. In the context of Cognitive Diagnostic models, these relationships are capture using the Q-matrix (the conceptual foundations are described by Chen et al., 2015)[4]. The process of identifying the prerequisite relation (or Q-matrix) is an iterative, expert driven process. Experts support the identification of involved competencies (Table 1) and establishing the prerequisite relation. In a

---

[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4539161/

first instance, the prerequisite relations are hypotheses of the nature of the domain and are subject to empirical evaluations (see below). Figure 2 (left panel) shows an exemplary prerequisite relation as a *Hasse diagram* (which is a common form of presenting prerequisite relations in CbKST). The Hasse diagram reads from bottom to top and states, for example, that competencies *b* and *c* are prerequisites for competency *d* (or should taught before / are usually learned before competency *d*).

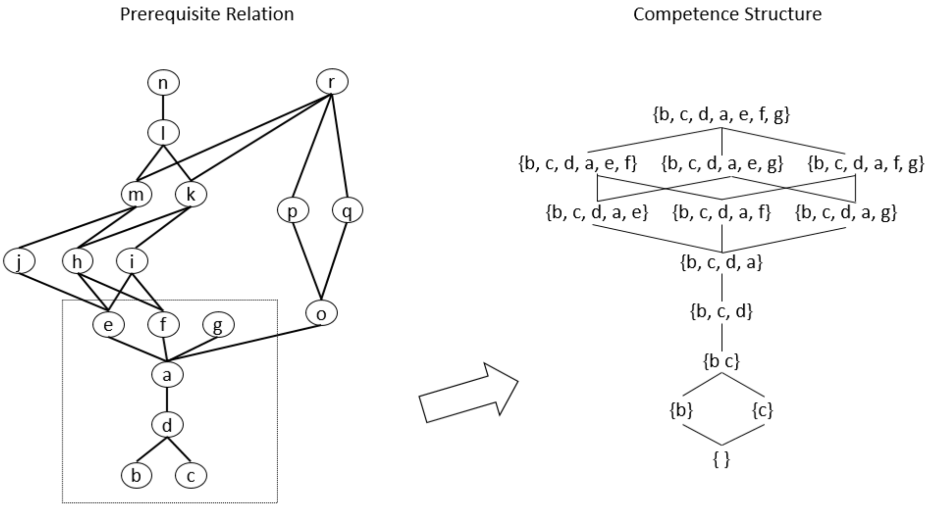


Figure 2. The figure illustrates the prerequisite relation for the competencies in Table 1 (left panel) and the derived competence structure.

Based on the prerequisite relation the competence structure (or *competence space*, depending on the mathematical properties) can be derived. Figure 2 (right panel) illustrates the principle. Given that the prerequisite relation is correct, learners can be in certain admissible states. For example, a student can have none of the competencies (which is the empty set { }). A student can have the competencies {b, c, d} but it is not admissible that a student has only competency *d*, because competencies *b* and *c* are prerequisites for *d*. The lines in the competency structure illustrate possible learning paths. For example, a student can proceed from {b, c, d, a} to either {b, c, d, a, e}, {b, c, d, a, f, or {b, c, d, a, g} by acquiring either competency *e*, *f*, or *g*. The assumption of acquiring only one competency at a time is the property of so-called well-graded knowledge structures/spaces. This property is in applied settings not necessarily given. It is possible, though, to consider multiple hidden learning steps, as long as the perquisite relationships are not compromised (this is outlined in more detail in the context of Micro Adaptivity and Micro Learning Spaces; Kickmeier-Rust, in press; Kickmeier-Rust & Albert, 2010).

The competence structure, in the end, is the actual domain model underlying the game. The entire process is driven by domain experts. The derived domain model is subject to an evaluation on the basis of response data from students.

2. Task Model

The task model is developed by generating prototypical tasks or challenges based on the domain model and under the consideration of additional aspects such as the target school level, prior knowledge, age, gender, etc. Again, this is done along the recommendations of curricula (that usually specify prototypical tasks and learning objectives; Box 2) and general item development guidelines (e.g., that of ETS[5]).

Box 2: Example tasks

(t1) Calculate the sum of forces in one dimension and determine force and acceleration from Newton's 1st and 2nd law.

(t2) Use the second law to calculate velocity at given time during a uniform acceleration

(t3) Identify the forces exerting on an object in a particular situation.

(t4) Can force change the speed of an object?

(t5) What is the difference between speed and acceleration?

3. Evidence Model

In CbKST, so-called *interpretation* and *representation functions* are applied to associate tasks and competencies. The interpretation function states which competencies can be surmised from solving a task and, vice versa, the representation function states which tasks represents which competencies. Figure 3a shows the representation function for the prototypical tasks in Box 2. It shall be noted that non one-to-one correspondence is required between tasks and competencies. In terms of ECD, interpretation and representation functions establish the evidence model.

Based on the interpretation and representation function, a task model (*knowledge structure/space* in the terminology of CbKST) can be derived (Figure 3b). The task model determines the logical order of tasks (game levels) for the game.

(a) Representation function

Task	Competencies
t1	a, b, c, d, e, f, g, h, i, k, l
t2	a, b, c, d, e, f, g, h, i, k, l
t3	a, b, c, d, e, f, g, j, o
t4	a, b, c, d, e
t5	a, b, c

(b) Knowledge structure

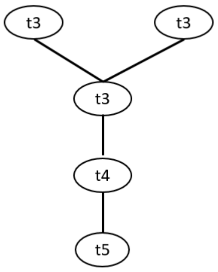


Figure 3. (a) CbKST-type representation function and (b) the resulting knowledge (or task) structure.

---

[5] https://www.ets.org/k12/capabilities/item-development/

4. Assembly Model

An assembly model specifies how domain, task, and evidence models interact to generate sufficient evidence to form a valid assessment (Almond et al., 2015). The assembly model includes general rules and weightings for the process of concluding the probabilities of certain competencies from the observed task performance. Specifically in game-based contexts such models assures comparability of different sources of evidence (game levels) and a reasonable influence of particular actions on the assessment, within a task/level. Basically, this processes is reflected in the concept of Micro Adaptivity (Kickmeier-Rust & Albert, 2010), where all actions and behaviors a student exhibits in a game environment are interpreted – with a certain weighting – in terms of holding or not holding specific competencies.

Technically, this is done by establishing a probability distribution over the states of the competence structure, which sums up to a probability of 1 to be in one of the states. Based on the assembly model, after each action in the game, the probability distribution is updated according to the result of an action. As an example, when a task is mastered, the probabilities of states including the competencies necessary to master this task (as of the evidence model) will be increased and the probabilities of other states decreased accordingly. The foundations of this procedure are described, for example, by Augustin et al. (2013, 2015).

5. Learner Model

The learner model is an individual instance of competence and performance structures. Based on the performance of a student I the game the probability distribution over the model is updated. This leads to a believe model stating the most likely state a student is in and therefore which competencies she holds or not (Figure 4).

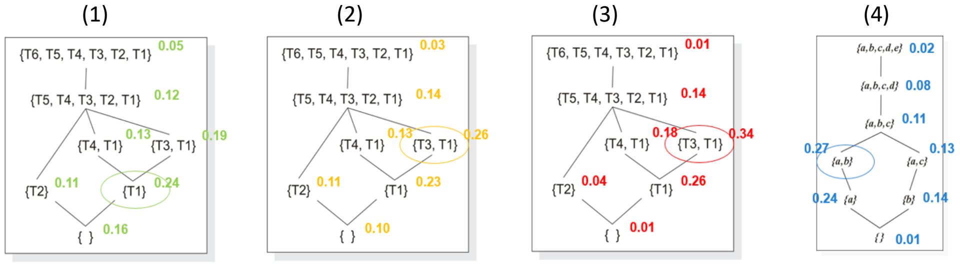


Figure 4. Iterative assessment procedure. With each action of the learner, the probability distribution over the knowledge structure is updated (1-3). The probability distribution over the latent competence structure can be derived and, in turn, the probabilities of individual competencies (4).

6. Game Design

In a next step, the prototypical tasks, defined in the task model, must be translated in concrete game levels. Since we intend to extend the Physics Playground game, the general design of the game and the set of game mechanics and learning supports are already given.

Physics Playground is a 2-dimensional computer game on Newtonian physics for students from 7th grade into adulthood. The goal across all the hundreds of existing game levels is to move a ball so that it hits a balloon. Students solve a level by either drawing (sketching levels) or manipulating variables (manipulation levels). That is, sketching levels are solved by drawing ramps, levers, pendulums, and/or springboards on the screen using a mouse or stylus. Manipulation levels are solved by moving sliders of various physics parameters, such as changing the ball's mass, the level's air resistance, gravity, and/or bounciness of the ball. Students may also change the amount of force exerted on the ball by using a puffer and/or adjusting a blower in certain manipulation levels. (Shute et al., 2020). Figure 5 shows some game levels, Table 2 lists the existing game mechanics and level types.

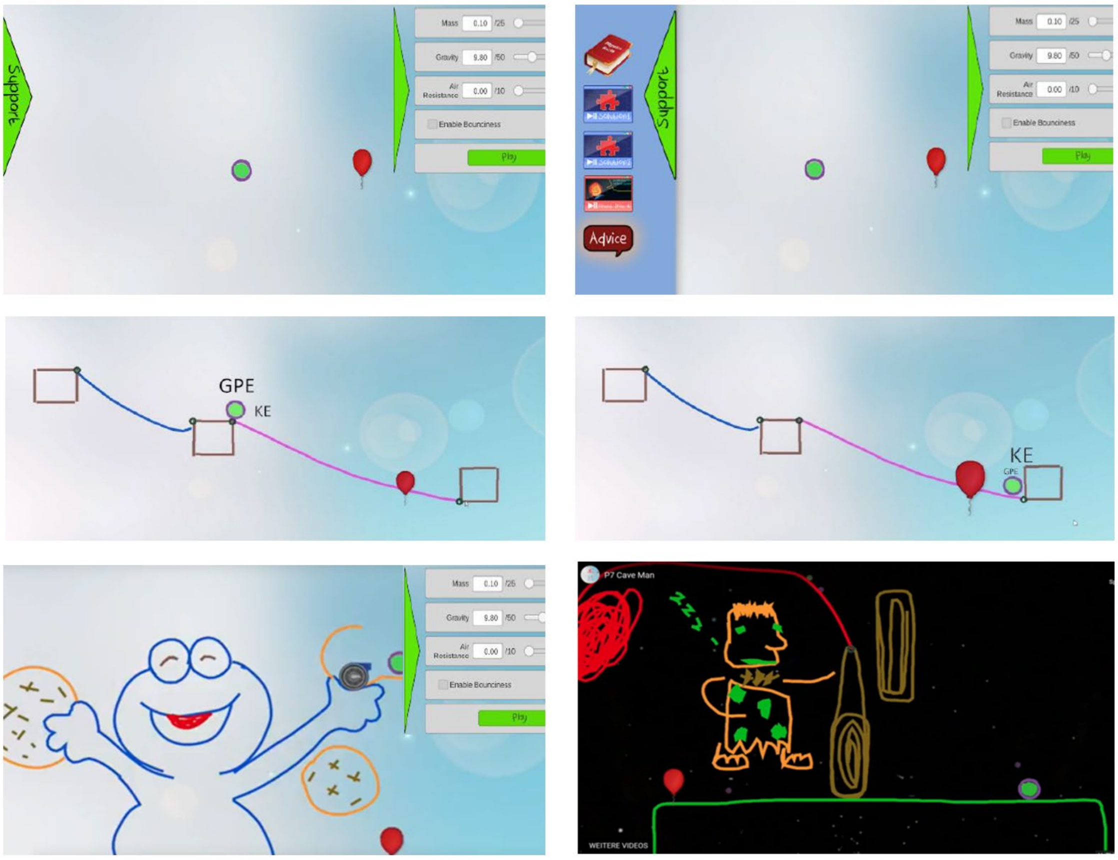


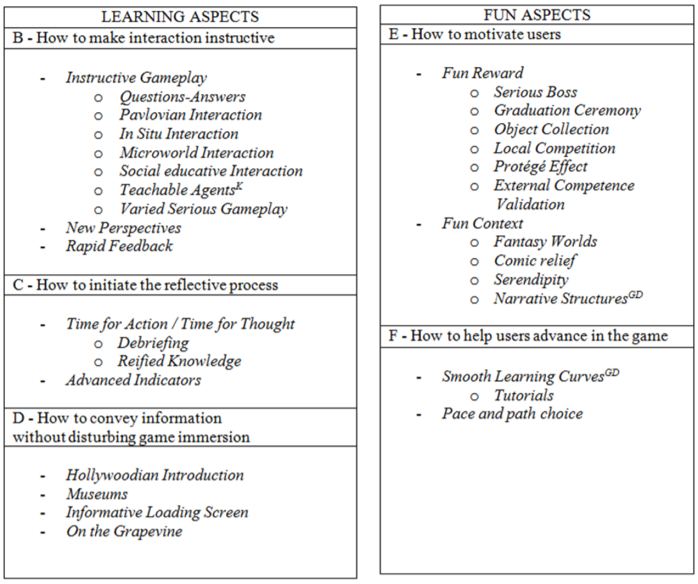
Figure 5. Example levels of Physics Playground.

Table 2. Game mechanics of Physics Playground.

Level types	Mechanics	Usage
Sketching levels	Ramps Levers Pendulums Springs, springboards Weights Boosters	Drawing on the screen with mouse, stylus or fingers
Manipulation levels	Wind Mass Air resistance Gravity Bounciness Force Friction	Adjusting parameters with sliders

In the course of future project, the existing level types and mechanics will be extended. Examples for further level types are Tetris type levels, crafting levels (e.g., building objects such as towers, caves, etc.), and race levels. Examples for further objects and mechanics are vehicles, tools (hammer, drill, jackhammer), elements (e.g., ramps) of different materials (to alter friction), weather conditions, and location (mountaintop, sea level, other stellar objects such as moon).

The process of designing new levels is based on the specifications of the task model and translate the task requirements into the available game mechanics. This process is guided, as far as applicable, by serious game design patterns, for example, Huynh-Kim-Bang, Wisdom, & Labat (2010) (Figure 6).



LEARNING ASPECTS

B - How to make interaction instructive

- *Instructive Gameplay*
    - *Questions-Answers*
    - *Pavlovian Interaction*
    - *In Situ Interaction*
    - *Microworld Interaction*
    - *Social educative Interaction*
    - *Teachable Agents*^{K}
    - *Varied Serious Gameplay*
- *New Perspectives*
- *Rapid Feedback*

C - How to initiate the reflective process

- *Time for Action / Time for Thought*
    - *Debriefing*
    - *Reified Knowledge*
- *Advanced Indicators*

D - How to convey information without disturbing game immersion

- *Hollywoodian Introduction*
- *Museums*
- *Informative Loading Screen*
- *On the Grapevine*

FUN ASPECTS

E - How to motivate users

- *Fun Reward*
    - *Serious Boss*
    - *Graduation Ceremony*
    - *Object Collection*
    - *Local Competition*
    - *Protégé Effect*
    - *External Competence Validation*
- *Fun Context*
    - *Fantasy Worlds*
    - *Comic relief*
    - *Serendipity*
    - *Narrative Structures*^{GD}

F - How to help users advance in the game

- *Smooth Learning Curves*^{GD}
    - *Tutorials*
- *Pace and path choice*

Figure 6. Game design patterns according to Huynh-Kim-Bang, Wisdom, & Labat (2010).

Design patterns can help to follow a sensible and succeeding design process. Such recommendations come from Plass, Horner, and Kinzer (2015), Marne et al. (2012), Mildner and Müller (2016), Roungas (2016), and many more. Siriaraya, Visch, Vermeeren, and Bas (2018) highlight the special role of interactive loops between game designers and domain experts. In their colorful paper, these authors describe the serious game design process in form of a cookbook as (i) the specification of desired transfer effects, (ii) the analyses of the target user's world, (iii) the actual game design, and (iv) the evaluation of effects. Lameras et al. (2017) published a comprehensive set of rubrics to link game mechanics to educational purposes. Ravys et al. (2016) published a meta-review on the success factors of serious game elements, uncovering five key elements: (i) backstory and production, (ii) realism (iii) artificial intelligence and adaptivity (iv) interaction, and (v) feedback and debriefing. These elements elaborate on practical guidelines to design successful educational games.

A special role in the context of the project takes Evidence Centered (Assessment) Design (ECD). The approach provides language, concepts, and knowledge representations for designing and delivering educational assessments, organized around the evidentiary argument - the inference of what learners know and can do from the performance they exhibit in assessments (Mislevy, Steinberg, & Almond, 2003). In essence, ECD decomposes the assessment design in domain model, task model, evidence model, linked by an assembly model that specifies how the models work together to generate sufficient evidence to form a valid assessment. These models then guide the development of the operational machinery of assessment, such as tasks, rubrics, and statistical models. Mislevy and colleagues (2014) proposed an extension called Evidence- Centered Game Design (ECgD) to align the game and the evidentiary reasoning mechanism. Specially for the use in serious games featuring Bayesian assessment approaches (such as *Physics Playground*), a 10-step design procedure has be described. These steps are organized into four distinct iterative phases: (i) designing levels (i.e., task authoring), (ii) creating assessment models, (iii) putting all assessment models together using Bayes nets as assessment machinery, and (iv) evaluating and refining the assessment models (a detailed description of the ECD procedure is provided by Kim, Almond, & Shute, 2016)[6].

Exemplarily, the design of a single[7] new game level would follow these procedures:

> Step 1: Identification of learning goal - we take the hammer and feather drop experiment on the moon (cf. Box 1).
>
> Step 2: Identification of competencies involved in the learning scene (cf. Table 1).
>
> Step 3: Development of a domain model (cf. Figure 2).
>
> Step 4: Defining an evidence model (i.e., identifying game mechanics that teach and assess the competencies specified in step 2). This occurs by selecting and combining game mechanics from Table 2 appropriate for the competencies.
>
> Step 5: Composing the new game level and graphical design.
>
> Step 6: Iteratively revising and optimizing the design based on design patterns and in exchange with domain experts, psychometricians, technicians, and users.

A new game level might look like this (Figure 7). The level is on moon, the task is to trigger a button at the ground as quickly as possible, to hit the red balloon with the cannon before the balloon flies away. The player has to decide whether to direct the green ball to the feather or to go the longer way to the heavy iron ball. The same level could take place on Earth. In a manipulation level, the same situation could be used, allowing the player to adjust and experiment with air resistance. Once a prototypical level is established, in a further step, learning supports must be designed (see next section).

Figure 7. A prototypical new game level.

---

[6] https://www.researchgate.net/publication/287799011_Applying_Evidence-Centered_Design_for_the_Development_of_Game-Based_Assessments_in_Physics_Playground

[7] Please note that certain steps described in this document (for example, task, assembly, learner models) operate on the level of a set of takes or the entire game.

7. Learning Support / Feedback

Alongside the design of game levels, a repository of learning supports are designed. The domain, task, and evidence models provide insights into admissible competence and performance states, which allows developing a set of appropriate learning supports (cf. Table 4). The existing learning support for Physics Playground are listed in Table 3).

Table 3. Learning supports in Physics Playground

Type	Description	Example
Worked examples	Show students how an expert would solve the level; this refers to the didactic principles of modelling.	
Physics facts	The game features a set of terms based on the physics experts' recommendations of the physics covered through gameplay.	
Animations	Animations explaining the physics phenomena.	For example, a ball rolling off a a slope and hitting an object.
Hewitt videos	Hewitt videos are cartoon animations explaining general physics competencies such as Newton's first, second, and third laws of force and motion, created by Paul Hewitt.	
Advice	The game features 2-3 advices (i.e., short sentences explaining the physics competency underlying a specific game level) created by the domain experts.	"Remember, an object will keep doing what it's already doing until an unbalanced force acts on it".
Interactive definitions	Based on the set of physics facts a set of cloze tasks are available to practice physics skills.	
Formulas	Where applicable, symbolic representations or formulas are available for learning supports.	
Glossary	For the physical terms, a glossary is available.	
Game tutorials	For the main game mechanics (e.g., ramps, levers, springboards) tutorial levels are available.	
Game tips	To directly access information, game tips are available covering controls, quick reminders of the tutorial levels, and access to the student's backpack (i.e., progress tracking and incentive system).	

Table 4. Additional learning supports.

Type	Description	Example
Adaptive level sequencing	Depending on the learner model, the sequence of presented levels is tailored to the competencies and the performance of a particular student (cf., Wiemeyer, Kickmeier-Rust, & Steiner, 2016).	Individual sequence of presented game levels.
Dynamic difficulty adjustments	Depending on the learner model, the difficulty of a presented levels is tailored to the competencies and the performance of a particular student (Ang & Mitchell, 2017; Wiemeyer, Kickmeier-Rust, & Steiner, 2016; Zohaib, 2018).	Adding or removing sub-tasks in game levels; altering the difficulty of parameters in physics formulas.
Problem solving support	When the adaptive game system detects a student being stuck in a task or sub-task (e.g., based on the progress through the problem space or a time lag), the game may trigger tailored hints, help analyzing the problem, or implementing a plan (Kickmeier-Rust & Albert, 2010; Reddy & Panacharoensawad, 2017)	“Why don’t you try adjusting parameter x”, “Think about which forces exert the ball in the situation”, “Do you think you have considered all forces exerting the ball?”
Praise	Although some authors (incl. Hattie & Timperly, 2007) argue that praise offers little instructional value, specifically in the context of serious games, where motivation is key and that is a setting without the guidance of a teacher, praise may provide feedback of how well a student is doing (perhaps in comparison to others) and increases motivation.	“You did very well!”, “Cool, you did better than most other players”.
Motivation support	Achievement motivation and the intrinsic motivation to play a game are considered distinct advantages of serious games. Prior research has shown positive effects of positive feedback on motivation and performance (e.g., Burgers et al, 2015; Kickmeier-Rust, Hillemann, & Albert, 2014).	Competition, score boards, rewards, unlockables, cheer: “Great, do you think you can solve a level faster next time?”
Time pressure	Time pressure can be a motivating factor for certain (!) students. Essentially, this factor is related to flow theory (Csíkszentmihály, 1996) and the theory of optimal arousal (Yerkes-Dostson law; Yerkes & Dotson, 1908).	Time limit for a level; stopwatch timer

Feedback is likely the most complex method, covering various goals and operating on various levels. Hattie and Timperley (2007) argue that the “main purpose of feedback is to reduce the gap between current understandings and performance and a goal” (p.86). These authors propose a model that includes *feed up* (where am I going), *feed back* (how am I going), and *feed forward* (where to next), each of them operating on the levels of task, process, individual, and self-regulation. A multitude of research specifies the relevant approaches to feedback in detail, for example Jones (2005),v an der Kleij, Vermeulen, Schildkamp, & Eggen (2015), or Yi (2017). Narciss (2006, 2013) proposed the *Interactive Tutoring Feedback Model* that includes context conditions of feedback (learner, feedback source, domain), characteristics of feedback (function, contents, presentation), and effects of feedback (Figure 8). The model operates on three levels, that is, cognitive, meta-cognitive, and motivational levels. Keuning, Jeuring, and Heeren (2016) conducted a meta-review on feedback development and highlights key requirements for the realization of feedback in educational tools. Serious games is a

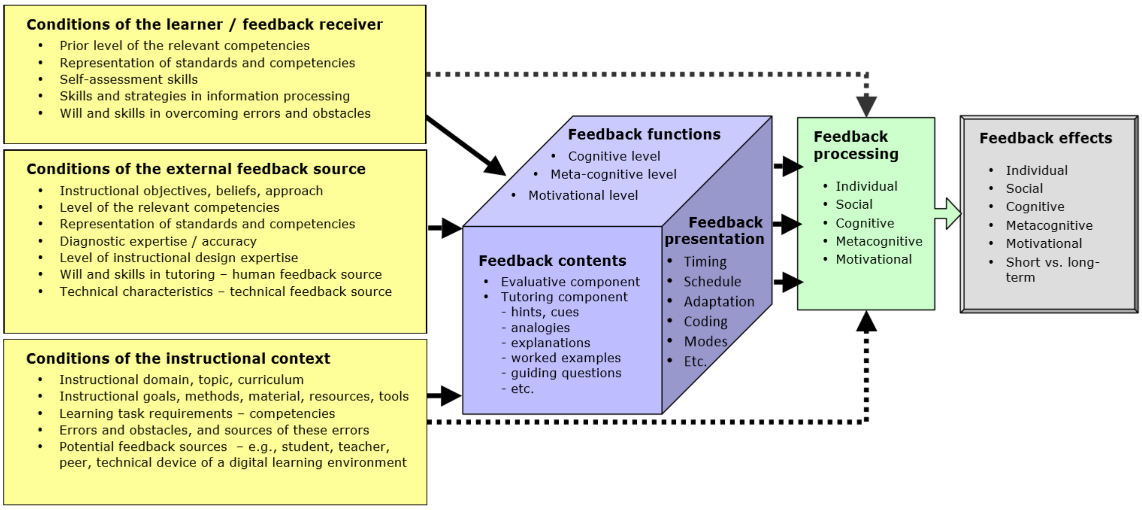


Figure 8. Susanne Narciss' feedback model (Narciss, 2006).

medium that requires special considerations for feedback due to its fragile immersive and motivational nature and aspects of competition (cf. Graesser, 2017). Nadolski and Hummel (2016) highlight the challenges in the game context, particularly the provision of unobtrusively, seamlessly integrated feedbacks in order not to destroy the "delicate balance between learning and playing" but also the cost factor of technical implementations. The authors provide clear guidelines on the implementation of successful feedback strategies in serious games.

Accounting for the existing set of learning supports and frameworks for learning support and feedback, we can establish a repository of learning supports. Based on the student's performance, the learner model is updated. The ideal case is that a student masters a level instantaneous. This is interpreted as an indicator (evidence) that this student holds all the relevant competencies. In case of failure, the exhibited actions determine the most likely competence state of the student (cf. Figure 2). This means that with an increasing number of played levels the believe model (probability distribution; cf. Figure 4) is updated, allowing the game to determine the competencies a player holds and which she doesn't hold. The assembly model holds certain threshold values for competence probabilities. If the probabilities fall below these thresholds, tailored learning supports and feedbacks are triggered. On this basis, the most appropriate learning support, specifically feedback and hints, are triggered. As we have demonstrated in prior research (e.g., Kickmeier-Rust, Hillemann, & Albert, 2014), triggering learning supports must not break the flow of the game. Thus, the assembly model holds rules about when to trigger a support function (including the number of unsuccessful actions, the time of being idle, probability thresholds, etc.).

In addition to automatically triggered support and feedback, Physics Playground features the help functions illustrated in Figure 9.

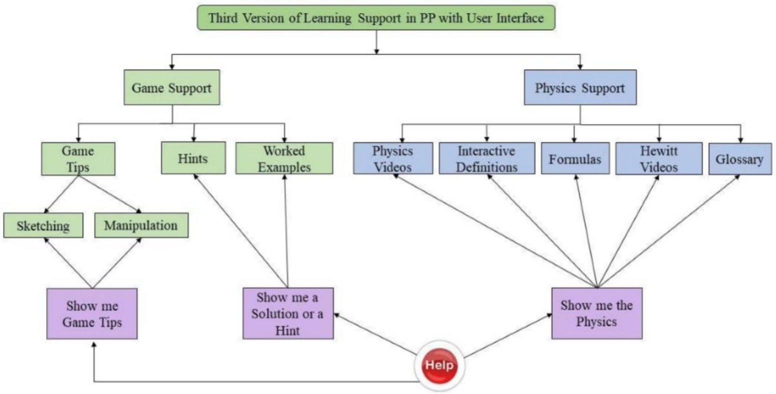


Figure 9. Existing learning supports in Physics Playground (Shute, 2020).

8. Iterative Evaluation and Revisions

The entire process occurs in a mutual exchange of different disciplines and in an iterative evaluation and revision strategy. Once a new game level is designed and implemented, it is subject to user evaluations. In form of pilot studies or in form of focus groups. The insights of these evaluations, in turn, inform the revision and optimization of game levels and the related learning supports.

9. References

Almond, R. G., Mislevy, R. J, Steinberg, L.S., Yan, D. & Williamson, D.M. (2015). *Bayesian Networks in Educational Assessment*. Brelin: Springer.

Ang, D., & Mitchell, A. (2017). Comparing Effects of Dynamic Difficulty Adjustment Systems on Video Game Experience. In *CHI PLAY '17: Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, October 2017, pp. 317–327.

Augustin, T., Hockemeyer, C., Kickmeier-Rust, M. D., Podbregar, P., Suck, R., & Albert, D. (2013). The simplified updating rule in the formalization of digital educational games. *Journal of Computational Science*, 4(4), 293-303.

Augustin, T., Hockemeyer, C., Suck, R., Podbregar, P., Kickmeier-Rust, M. D., & Albert, D. (2015). Individualized skill assessment in educational games: The mathematical foundations of partitioning. *Journal of Mathematical Psychology*, 67, 1-7.

Burgers, C., Edena, A., van Engelenburg, M.D., & Buninghb, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior*, 48, 94-103.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical Analysis of Q-matrix Based Diagnostic Classification Models. *Journal of the American Statistical Association*, 110(510), 850–866.

Csíkszentmihályi, Mihály (1996), *Creativity: Flow and the Psychology of Discovery and Invention*, New York: Harper Perennial.

Graesser, A.C. (2017). Reflections on serious games. In H. van Oostendorp and P. Wouters (Eds.) *Instructional techniques to facilitate learning and motivation of serious games* (pp. 199-212). Cham: Springer.

Hattie, J., & Timperley, H. (2007). The Power of feedback. *Review of Educational Research*, 77, 81-112.

Huynh-Kim-Bang, B., Wisdom, J.O., & Labat, J.-M. (2010). Design Patterns in Serious Games: A Blue Print for Combining Fun and Learning. Retrieved March 5, 2020 from https://www.semanticscholar.org/paper/Design-Patterns-in-Serious-Games-%3A-A-Blue-Print-for-Huynh-Kim-Bang-Wisdom/

Jones, C.A. (2005). *Assessment for learning*. London, UK: Learning and Skills Development Agency.

Keuning, H., Jeuring, J.,& Heeren, B. (2016). *Towards a Systematic Review of Automated Feedback Generation for Programming Exercises – Extended Version.* Technical Report UU-CS-2016-001, Utrecht University, The Netherlands. Retrieved March 10, 2020 from http://www.cs.uu.nl/research/techreps/repo/CS-2016/2016-001.pdf

Kickmeier-Rust, M.D. (in press). *Micro Learning Spaces: Using Competence Models as Anchors for Multi-Modal, Multi-Source Learning Analytics.* To appear in Assessment in Higher Education.

Kickmeier-Rust, M.D., & Albert, D. (2010). Micro adaptivity: Protecting immersion in didactically adaptive digital educational games. *Journal of Computer Assisted Learning*, 26 (2), 95-105.

Kickmeier-Rust, M. D., Hillemann, E.-C., & Albert, D. (2014). Gamification and Smart Feedback: Experiences with a Primary School Level Math App. *International Journal of Game-Based Learning*, 4(3), 35-46.

Kim, Y.J., Almond, R.J., & Shute, V. (2016). Applying Evidence-Centered Design for the development of game-based assessments in Physics Playground. *International Journal of Testing*, 0, 1-22.

Kuczmann, I. (2017). The structure of knowledge and students' misconceptions in physics. *AIP Conference Proceedings*, vol. 1916, 050001-1- 050001-6.

Lameras, P., Arnab, S., Dunwell, I., Stewart, C., Clarke, S., & Petridis, P. (2017). Essential features of serious games design in higher education: Linking learning attributes to game mechanics. *British Journal of Educational Technology*, 48(4), 972-994.

Marne, B., Wisdom, J., Huynh-Kim-Bang, B., & Labat, M.-J. (2012). The six facets of serious game design: A methodology enhanced by our design pattern library. In Proceedings of the 7th European Conference of Technology Enhanced Learning, EC-TEL 2012 (pp. 208-221), September 18-21, 2012, Saarbrücken, Germany. *Lecture Notes in computer Science*, 7563. Cham: Springer.

Mildner, P., & Müller, F. (2016). Design of Serious Games. In Dörner R., Göbel S., Effelsberg W., Wiemeyer J. (Eds.), *Serious Games*. Cham: Springer.

Mislevy, R.J., Oranje, A., Bauer, M., von Davier, A.A., Hao, J., et al. (2014). *Psychometric considerations in game-based assessment.* Redwood City, CA: GlassLab.

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-67.

Nadolski, R.J., & Hummel, H.G. (2016). Retrospective cognitive feedback for progress monitoring in serious games. *British Journal of Educational Technology*, 48(6), 1368-1379.

Narciss, S. (2006). Informatives tutorielles Feedback. *Pädagogische Psychologie und Entwicklungspsychologie*, Band 56. Münster: Waxmann.

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *Digital Education Review*, 23(1), 7-26.

Plass, J.L., Horner, B.D., & Kinzer, C.K. (2015). Foundations of Game-Based Learning. *Educational Psychologist*, 50(4), 258-283.

Ravyse, W.S., Seugnet Blignaut, A., Leendertz, V., & Woolner, A. (2017). Success factors for serious games to enhance learning: a systematic review. *Virtual Reality*, 21, 31-58.

Reddy, M.V, & Panacharoensawad, B. (2017). Problem-Solving Difficulties and Implications in Physics: An Empirical Study on Influencing Factors. *Journal of Education and Practic*e, 8(14), 59-62.

Roungas, B. (2016). A model-driven framework for educational game design. *International Journal of Serious Games*, 3(3), 19-37

Shute, V.J., Smith, G., Kuba, R., Dai, C.-P., Rahimi, S., Liu, Z., & Alom, R.G. (2020). The design, development, and testing of learning supports for the Physics Playground game. *International Journal of Artificial Intelligence in Education*, 37 pages. Preprint available online at http://myweb.fsu.edu/vshute/pdf/IJAIED2020.pdf

Shute, V.J., & Ventura M. (2013). *Stealth Assessment. Measuring and Supporting Learning in Video Games.* Cambridge, MA: The MIT Press.

Siriaraya, P., Visch, V., Vermeeren, A., & Bas, M. (2018). A cookbook method for persuasive game design. *International Journal of Serious Games*, 5(1), 37-71.

van der Kleij, F., Vermeulen, J., Schildkamp, K., & Eggen, T. J. (2015). Integrating data-based decision making, Assessment for Learning and diagnostic testing in formative assessment. *Assessment in Education*, 22(3), 324-343.

Wiemeyer, J., Kickmeier-Rust, M. D., & Steiner, C. M. (2016). Performance assessment in serious games. In R. Dörner, S. Göbel, W. Effelsberg, and J. Wiemeyer (Eds.), *Serious Games: Foundations, Concepts and Practice* (pp. 273-302). Berlin: Springer.

Yerkes RM, Dodson JD (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18 (5): 459–482$

Yi, S. (2017). *The role of feedback in game-based learning: A review of the literature.* Thesis presented at the University of Texas. Retrieved March 10, 2020 from https://repositories.lib.utexas.edu/handle/2152/60446

Žavcer, G., Mayr, S., & Petta, P. (2014). Design patterns canvas: An introduction o unified serious game design patterns. *Interdisciplinary Description of Complex Systems*, 12(4), 280-292.

Zohaib, M. (2018). Dynamic Difficulty Adjustment (DDA) in Computer Games: A Review. *Advances in Human-Computer Interactio*n, v2018, 12 pages.