Grant Agreement No. 687676

Innovation Action

ICT-20-2015

D6.1 Evaluation Guidelines

Due date	Month 14
Actual date	Month 18
Deliverable author(s)	Marius Preda, Mustafa Ali Türker, François Mohier
Contributors	Sylvester Arnab, Jannicke Baalsrud Hauge, Theo Lim, Tomasz Skupiński, Sadi Türeli, İlker Helvaci, Sertan Okan, Mihaela Garabet, Radu Jugureanu, Mihaela Coman, Olivier Heidmann, Claire Dutertre, Raphael Attias, Aparajithan Sivanathan, Scott McGibbon, Leonel Morgado, Baltasar Fernandez Manjon, Neil Judd
Partner(s)	SIVECO
Version	V1.2
Status	Final
Dissemination level	PU

Project Coordinator

Coventry University

Sylvester Arnab

Priory Street, Coventry CV1 5FB, UK

E-mail: s.arnab@coventry.ac.uk

Project website: http://www.beaconing.eu

This project has received funding under the Horizon 2020 Framework Programme of European Union – Grant Agreement No. 687676

Version control				
Version	**Date**	**Author**	**Institution**	**Change and where applicable reason for change**
V1	19.03.2017	Mustafa Ali Türker Sadi Türeli İlker Helvaci Sertan Okan	SEBIT	Outline version
V2	21.04.2017	Mustafa Ali Türker Sadi Türeli İlker Helvaci Sertan Okan	SEBIT	Rough Draft version
V3	28.04.2017	Mihaela Garabet Radu Jugureanu Mihaela Coman Marius Preda	SIVECO	Draft version
V4	10.05.2017	Sylvester Arnab	COVUNI	Edit/review
V5	11.05.2017	Mihaela Garabet Mihaela Coman Marius Preda	SIVECO	Update
V6	11.05.2017	François MOHIER Olivier Heidmann Claire Dutertre Raphael Attias	ORT	Update/Review/Comments
	12.05.17	Theo Lim Aparajithan Sivanathan Scott McGibbon	HWU	Edit/Review/Comments Update 3.1.3
	30.05.17	Leonel Morgado	INESC TEC	Edit/Review/Comments
	30.05.2017	Jannicke Baalsrud Hauge	BIBA	Update chapter 3, 4, 6 comments
V7	2.06.2017	Mustafa Ali Türker	SEBIT	Complete chapter 3, revise other chapters
V8	07.06.2017	Mihaela Garabet Marius Preda	SIVECO	Review comments/ completing sections
V9	21.06.2017	Ali Turker	SEBIT	Inputs for 3.3 Pedagogical evaluation plan during spring 2018 pilots
V9	21.06.2017	Marius Preda	SIVECO	Update section 3.3
V9.03	28.06.2017	Lawrence Howard	HFC	Edit / Review / Comments

Quality control				
QA Version	**Date**	**QA Responsible**	**Institution**	**Change and where applicable reason for change**
V9.2	26/06/2017	Jayne Beaufoy	COVUNI	Language check and formatting
V9.3	26/06/2017	Olivier Heidmann	ORT	First reviewer
V9.3	28/06/2017	Lawrence Howard	HFC	Second reviewer
V1.0	28/06/2017	Jannicke Baalsrud Hauge	BIBA	QM- sent back for updates
V1.1	29/06/2017	Marius Preda	SIVECO	Update

V1.2	30/06/2017	Jannicke Baalsrud Hauge	BIBA	Reopened, removed section 2.1 since too similar to D6.2 chapter 2

Release approval				
Version	**Date**	**Name**	**Institution**	**Role**
1.1	29/06/2017	Jannicke Baalsrud Hauge	BIBA	QM
1.2	30/06/2017	Jannicke Baalsrud Hauge	BIBA	QM

Statement of originality
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

Table of Figures

Table of Tables

LIST OF ABBREVIATIONS

STEM - Science, Technology, Engineering and Math

ICT - Information and Communications Technology

FP6, FP7 - Framework Programme

VET – Vocational Education and Training

EUN - European Schoolnet

SIPTEC - Systemic, Institutional, Pedagogical, Technological, Economic, and Cultural

P2P - Peer-to-peer

MoE – Ministry of Education

TPACK - Technological Pedagogical Content Knowledge

RAT - Replacement, Amplification and Transformation

ODS - Open Discovery Space

ISE - Inspiring Science Education

LM-GM - Learning Mechanics–Game Mechanics

ECD - Evidence-Centred Designs

IRT - Item Response Theory

INESC - Instituto de Engenharia de Sistemas e Computadores

IMS – Instructional Management Systems

DSM - Delivery System Model

ISTE - International Society for Technology in Education

COSVR195 - Produce standard architectural stone enrichments

LRS - Learning Records Store

xAPI – Experience Application Programming Interface

TAM3 - Technology Acceptance Model 3

DoA – Description of Actions

WP - Work Package

TRL – Technology Readiness Level

BM - Block matching

ESSO – Environment-Strategy-Structure-Operations

PESTEL – Political, Economic, Social, Technological, Environmental, Legal

EBIT - Earnings before interests and taxes

ROI - Return On Investment

B2B - Business-to-business

B2C - Business To Consumer

B2G - Business to Government

C2C – Customer-to-Customer

ISO - International Organization for Standardization

CPT - Conditional Probability Tables

EXECUTIVE SUMMARY

This deliverable is one of the two deliverables of Task 6.1.

BEACONING (Breaking Educational Barriers with Contextualised, Pervasive and Gameful Learning) project will provide different learning scenarios supported by different technologies for teaching and learning in an inclusive society focusing on 21st century skills, competencies, strategies, learning outcomes and learning disabilities.

The large-scale pilots are aimed at validating the scalability of integrated solutions in different real-life educational contexts by addressing large user groups. In contrast to WP5 small-scale pilots, the large-scale pilots are designed top-down starting with the engagement of networks of schools through to educational NGOs and educational experts. The target audience will be of significant size (5000), involving large and heterogeneous groups of learners with their teachers. Specific studies, including leveraging online surveys, enables a wider network analysis on the objectives of establishing the large-scale pilot target participant groups, addressing crosscutting school topics as well as intra- / inter school relations. The pilots will primarily focus on engineering, entrepreneurship and on developing STEM and digital skills.

The aim of this deliverable is to provide a plan for the evaluation of the whole project: timelines, approaches and responsibilities are described in order to provide a sustainable backbone during the whole project lifetime.

On the other hand this deliverable intends to give guidelines to all involved stakeholders to have a common approach to tackle the task of evaluating a large-scale learning project.

The different learning paths identified as samples under WP3 have been the basis used to define an architecture developed in WP4. This top-down approach will ensure knowledge transfer from one community to another as well as additional actors in the value chain. Large scale pilots will be mainly deployed in 5 different countries: France, Israel, Greece (ORT) and in other ORT worldwide countries such as South-Africa, Italy, Bulgaria (ORT Network), Turkey (SEBIT) and Romania (SIVECO) involving a total of 5000 users. T6.1 will be implemented using outcomes from WPs 3, 4 (needs, user models, and specs) and WP5 (small scale pilots) in which the BEACONING platform will have already been technically and practically tested by small amount of users. The partners will use the gathered information and feedback in order to fine-tune the technical specifications of the educational content/tool.

The large pilots will inform upon how innovation can be made upon ICT and education to create fit-for-purpose digital technologies for learning. It should give us a view on how to remove obstacles for ubiquitous learning. It should also provide insights to the likelihood of uptake as a business and its adoption in WP7[1].

[1] Grant Agreement-687676-BEACONING

1 INTRODUCTION

1.1 BACKGROUND

This document details the guidelines designed to evaluate large-scale pilots of the BEACONING game-based learning platform.

The main aim of this report is the formulation of evaluation and usage guidelines for the training framework, both for teachers and learners. Methods and techniques developed to conduct verification and validation of the components, services, and tools provided by the BEACONING platform are included. The training framework, complete with exemplified solutions and guidelines, is to provide teachers and learners with an effective use of the BEACONING platform and its educational content.

1.2 ROLE OF THIS DELIVERABLE IN THE PROJECT

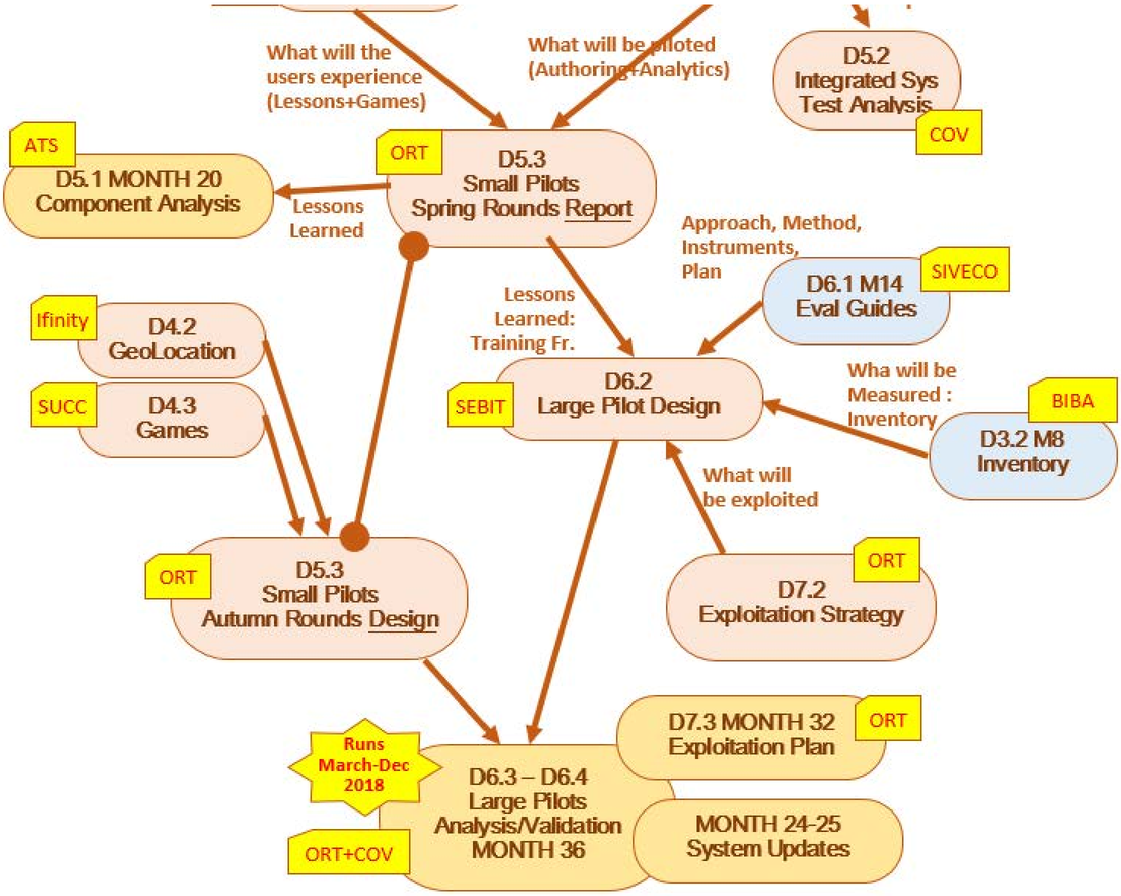


Figure 1 Qualitative and Quantitative Evaluation Methods from Action Research[2]

[2] EUN Validation Framework: fcl.eun.org/validation-service

1.3 APPROACH

Conducting an evaluation through an external user group ensures an international footprint and ecological validity of the BEACONING concept. Objective feedback from diverse cultural, learning, and professional environments, across diverse learning contexts will provide insights to the learning needs of individuals aged 15-24 in general or vocational education, including individuals with special needs. A combination of quantitative and qualitative methods will be used to evaluate the knowledge developed as a result of engaging with BEACONING activities, also informed by the small scale pilots and guidelines in WP5. Tests and other quantified methods will be used to assess the level of knowledge and quality of the experience before and after the learner, accompanied by the teacher, engagement with the gamified approach. Qualitative methods in the forms of interviews with selected users (learners and teachers) will be used to evaluate the degree to which critical thinking or abstract thinking capacity has evolved as a result of engaging with learning games. Finally, information on learner performance will be gathered through the learning analytics background services.

Measures will include:

- indicators on engagement: number of learners using the platform; number of times accessed; total time spent on play-learn resources; completion rates; sense of achievement;
- indicators on effectiveness of the tools towards skill building: correlation between completion rates and professional achievement / satisfaction / progress (such as promotions) /aspiration; confidence level of learners on mastering specific tasks; evaluation of the development of soft skills such as executive functions, critical and analytical thinking, problem solving, and entrepreneurial thinking; correlation between skills developed and set learning objectives for specific cases;
- indicators on usability: ease of use of services; ease of integration into learning; attractiveness of learning activities; interface usability; adaptiveness to learner patterns collected through learning analytics;
- indicators on effectiveness of knowledge transfer: ability to explain newly developed knowledge to peers; capacity to apply knowledge to professional scenarios; sense of achievement;
- indicators on the impact of the project activities towards skill building among targeted stakeholders: number of learners, teachers, educational institutions, and other stakeholders reached through evaluation and dissemination3.

1.4 STRUCTURE OF THE DOCUMENT

The structure of this deliverable consists of:

Section 1 – an introduction for describing the background, the approach and the structure of this deliverable

Section 2 – describing the European benchmark for evaluation of large scale school pilots as part of the research and technology development Framework programme (iClass Project in FP6, iTec Project in FP7, Next-Tell Project in FP7) and, also, as part of the Competitiveness and Innovation framework Programme (Open Discovery Space and Inspiring Science Education).

[3]

Section 3 – is dedicated to the pedagogical value evaluation guidelines. Pedagogical evaluation approach consisting in: assessment in games, Evidence-Centred Design in BEACONING K12 Pilots, Evidence-Centred Design in BEACONING VET Pilots and examples from Gamified Lesson Paths, the method and the Pedagogical Evaluation Plan durIng Spring 2018 pilots.

Section 4 – is dedicated to acceptance and impact evaluation approach with method and the Plan during Spring 2018 pilots

Section 5 – describing the evaluation guidelines for location-based activities

Section 6 – describing the Business model evaluation guidelines

Section 7 – conclusions of this analysis

2 EUROPEAN BENCHMARK FOR EVALUATION OF LARGE SCALE SCHOOL PILOTS

Task 6.1 has identified three Framework Programme projects that have significant relevance. The relevant property is that all these projects implemented a novel technology enhanced learning solution in large scale school pilots and evaluated the outcomes. In D6.2 we review the setup of their pilots. Here in D6.1, we review the evaluation approaches and methods.

2.1 AS PART OF THE RESEARCH AND TECHNOLOGY DEVELOPMENT FRAMEWORK PROGRAMME

Three projects, one from each of the last 3 framework programmes are reviewed. More details can be found in D6.2.BEACONING evaluation approaches builds upon the different partners experiences from these three projects.

2.2 AS PART OF THE COMPETITIVENESS AND INNOVATION FRAMEWORK PROGRAMME

Two projects are identified in this programme which implemented a novel technology enhanced learning solution in large scale school pilots and evaluated the outcomes.

2.2.1 Open Discovery Space

CIP-ICT-PSP-2011-5, Coord: Intrasoft, 2000 schools

Overall, the aim of the **Open Discovery Space** consortium is to mobilise **10,000 teachers** and **40,000 students** in the framework of the proposed activities (requirements elicitation, implementation and assessment, validation), during the life cycle of the project.

These teachers act as innovation leaders in their communities spreading the ideas and the vision of the Open Discovery Space project.

The evaluation is based on quantitative and qualitative tools. A quantitative tool portal analytics was used while case studies produced qualitative data. Together with the questionnaires that produced both types of data a comprehensive picture of the ODS project was produced[4].

[4] http://opendiscoveryspace.eu/sites/ods/files/d5_2_evaluation_plan_updated.compressed.pdf

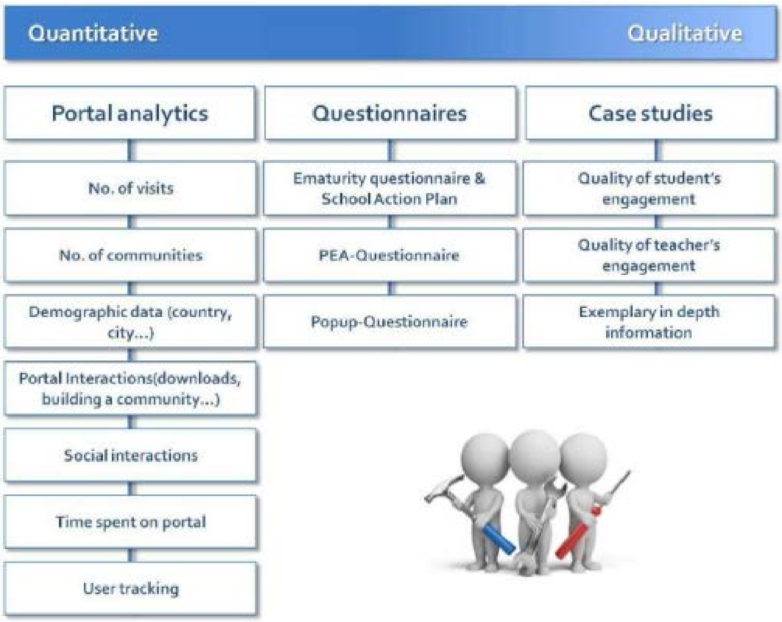


Figure 2 Quantitative and qualitative tools in the ODS project

2.2.2 Inspiring Science Education (ISE)

CIP-ICT-PSP-2012-6 , Coord: Intrasoft, 5000 schools

---SIVECO, Romania

Inspiring Science -Large Scale Experimentation Scenarios to Mainstream eLearning in Science, Mathematics and Technology in Primary and Secondary Schools is a project aimed at large-scale take-up of educational opportunities amongst European science teachers. Pilot activities took place in 5000 primary and secondary schools in 15 European countries. During these pilots, teachers accessed interactive simulations, educational games and eScience applications and integrated them with extra-curricular activities such as field trips to science centres and discovery parks, and virtual visits to research centres. Teachers also had the possibility to access remote and online labs and relevant scenarios for their use in the school classroom. Students are inspired to use eTools and digital resources to learn Science, Technology, Engineering and Maths (STEM related subjects) in a practical, competitive and exciting way.

The evaluation of the activities was based on the questionnaires to assess constructs in a pre/post design linked to background data. The teachers completed the questionnaires before carrying out a learning scenario and after running a learning scenario when they are asked to estimate how effective this scenario was for their students to learn about the inquiry circle, to be motivated for using eLearning tools, their ability to use eLearning tools, and its effect on their teaching practice.

The aim is to assess the influence of the ISE approach on affective constructs like interest and motivation as well as on their knowledge about the nature of science[5].

The validation of the ISE approach is mainly based on cases studies, teacher interviews and video documentations.

[5] http://inspiringscience.eu/sites/default/files/outcome/ise_d8_2_-_inspiring_science_education_evaluation_report.pdf

Web metric of the ISE portal indicate the acceptance of the proposed pedagogical framework of ISE and illustrate changes in the users´ behaviour. The benchmark data of the portal recorded: the number of sessions, the number of page views, the mean duration of a visit, the exit rate, the number of new sessions.

3 PEDAGOGICAL VALUE EVALUATION GUIDELINES

The pedagogical evaluation will be committed at the first round of large-scale pilots. The project aims to impact STEM competencies. The "learning requirements" and examples of such competencies can be found in D3.3 – Learning Environment System Specification. D3.3 lists

- Critical Thinking
- Complex Problem-Solving
- Computational Thinking
- Active Learning
- Adaptive Thinking

among others and provisions flexibility for instructional designers to define competencies as well as the corresponding evidence in the authoring templates. The authoring tool enables using the "LM-GM model to translate and implement the high-level pedagogical requirements into low-level game mechanics." On such mechanics the assessment framework will link the expected evidences to the competencies. Deliverable 3.3 states that "It is expected that the proposed BEACONING Taxonomy will enable users of the BEACONING Authoring Tool to create scenarios with associated missions and quests, and link these scenarios to specific competencies and skills to show learners' progression at a specific level. The evidence that learners should provide for their assessment will be linked with specific weights/measures that will be different for each gamified activity as the level of difficulty for each activity will be different."

D3.3 delivered in M8 was a preliminary version and the iterations/updates will be documented in D3.7 in M25. By that time the gamified lesson plans for piloting will be available and common competencies among them can be identified so that each pilot can ensure an evaluation is made about their progress.

Note that, such generic competencies are not possible to measure via standard tests and performance based assessment techniques are very hard to design and implement. Game-based learning is an opportunity not only to improve these competencies but also to measure them directly, during game play. The following subsection describes the adopted evaluation approach which exploits the interactive nature of gaming to harvest evidences of competence. Once the evidence is collected, they can be linked to the competencies using a rubrics based approach, or by developing software components that keep track of likelihood of competence.

3.1 PEDAGOGICAL EVALUATION APPROACH

Competencies can be assessed during performance. One approach is called "complex assessment" where authentic activities are designed which demand certain competencies so the students explicitly and consciously demonstrate competency during the activity. However, complex assessment suffers from inter-rater reliability problems when scored by humans and validity claims when scored automatically[6]. Game-based learning activities that would demand similar competencies provide an opportunity for assessment. Since the student's focus is on game play and learning needs to succeed in the game, such evaluation would be implicit, by means of the evidence (traces) that are left by the players. This approach is candidly called

[6] D. Williamson, I. Mejar & R. Mislevy (2006), Automated Scoring of Complex Tasks in Computer-Based Testing, pg.51 accessed on 22nd May 2017 at https://www.researchgate.net/profile/Isaac_Bejar/publication/236167620_Automated_scoring_of_complex_tasks_in_computer-based_testing/links/0deec53b4446cd0d8e000000.pdf

"stealth assessment" and it refers to game challenges that are subject to Evidence-Centred Designs (ECD). Therefore, assessment is woven directly and invisibly into the fabric of the learning or gaming environment[7].

3.1.1 Assessment in Games

Implicit Assessment during performance upon learning has been studied under many names:

- Mastery Assessment
- Evidence-Centred Assessment
- Stealth Assessment
- Embedded Assessment
- Rubric-Based Assessment
- Outcome-Based Assessment

The common property of all these techniques is that they rely on an evaluation expert to make diagnostic reasoning in advance about a certain progress of a certain competency as being the cause of a certain behaviour or outcome that may be observed. ECD takes this paradigm one step further and provisions the design of educational activities so that good (valid) evidences are designated to occur.

As the gamers interact with a game, the values of game-specific variables change, some of which may as well be educationally-relevant (e.g. number of attempts, problem solving and comments during gaming*)*. If that is the case, in addition to checking a game-specific variable (such as health status), players could also check their current levels of critical, computational or adaptive thinking. Each of these competencies can be further broken down into constituent knowledge and skill elements (e.g., teamwork may be broken down into cooperating, negotiating, and influencing/leadership skills), for which evidences from play can be linked. If the levels of those competencies got too low, the player would likely feel compelled to take action to boost them." [8]

Nevertheless, it is still a matter of chance that the progress in a competency is the one and only reason of an observation. Therefore, ECD employs probabilistic measures to infer with a certain degree of likelihood that an observation is the effect that is caused by progress. This is achieved by assigning prior probabilities to that cause. Prior probabilities can be calculated by counting past data or even by asking the expert who surmises the cause-effect relationship in the first place. They are updated as observations start to flow in.

The beliefs of experts about unobservable variables that characterise the knowledge, skills, and/or abilities of students are the basis of assessment. However, these beliefs are at best probabilistic and accumulating evidence across task performances is better be used to update them. This is the role traditionally associated with psychometric models, such as those of item response theory (IRT) and latent class models. The "items," however, in this case are "evidences."

[7] V. Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M-W. (2016). Advances in the science of assessment. *Educational Assessment*, *21*(1), 34-59. accessed on 22nd May 2017 at http://myweb.fsu.edu/vshute/pdf/asstPPF.pdf

[8] V. Shute & M. Ventura (2013), Stealth Assessment in Digital Games, MIT Press

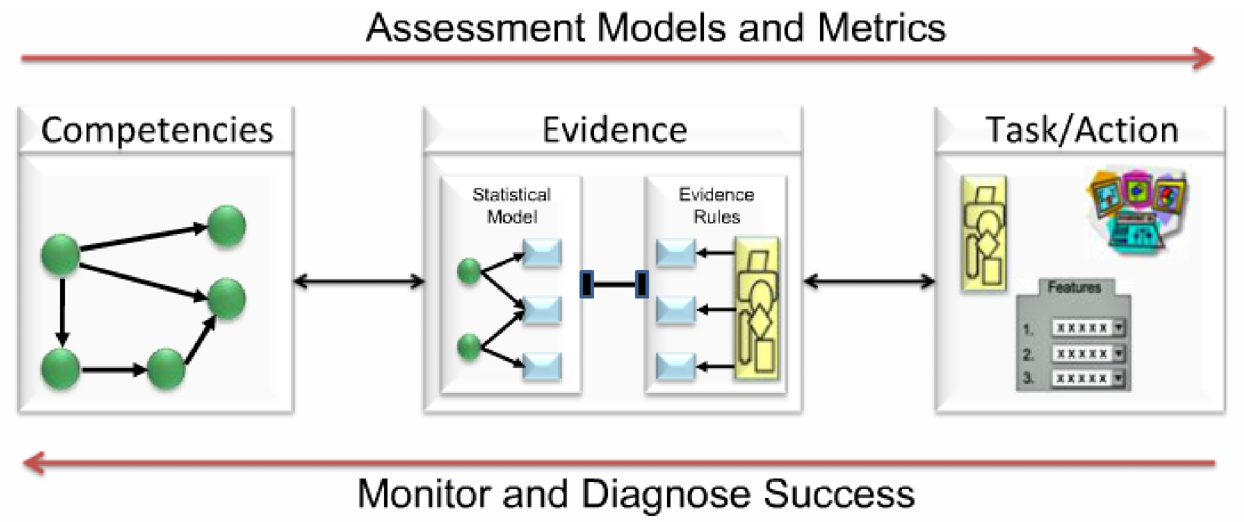


Figure 3 Evidence Centred Design

The above figure shows the two-step approach of ECD. "The red arrow heading left-to-right shows reasoning about assessment design (competency to evidence to task model) and the arrow going from right-to-left demonstrates reasoning about a person's performance."[9] The competencies in the constructed model are *random variables* that correspond to learner's attributes such as skills, knowledge, and abilities, so it is also called Student Model.

After a competency structure is constructed, an evidence model which links to it needs to be created. This model has some rules to extract, collect and synthesise evidences, plus a statistical model (a set of conditional probabilities). Then, the task model is created which harbours those evidences. At the run time of the tasks, evidences show up and their statistical model is used to update the competency model for individual students. As observations are made, the statistical model is updated to be used when the next set of evidence comes in.

3.1.2 Evidence-Centred Design in Beaconing K12 Pilots

One particular approach to evidence-centred design (ECD) is called the Triadic Certification model, which combines the competences with the challenges designed for the serious games on a matrix that matches the needs and levels.

[9] V. Shute, Y. Kim, & R. Razzouk (2016), Evidence Centred Design for Dummies, accessed on 22nd May, 2017 at http://myweb.fsu.edu/vshute/ECD%20for%20Dummies/ECD%20for%20Dummies.swf

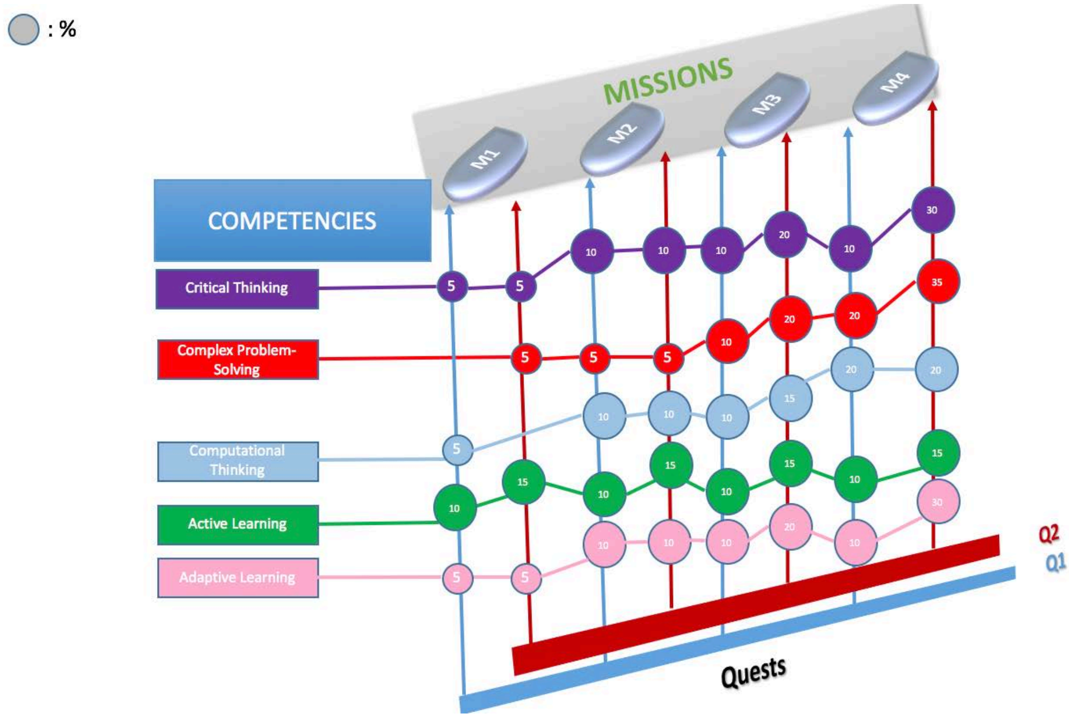


Figure 4 Triadic Certification approach[10]

Proposed by INESC, triadic certification accepts quest completion as partial evidence to acquiring a certain set of competencies. This approach is also being built into the Authoring Tool of the BEACONING platform.

[10] R. Babtista, A. Coelho & C. Carvalho (2015), Relation Between Game Genres and Competences for In-Game Certification, In Proc. 5th Intl. Conf. on Serious Games, pp. 28-35.

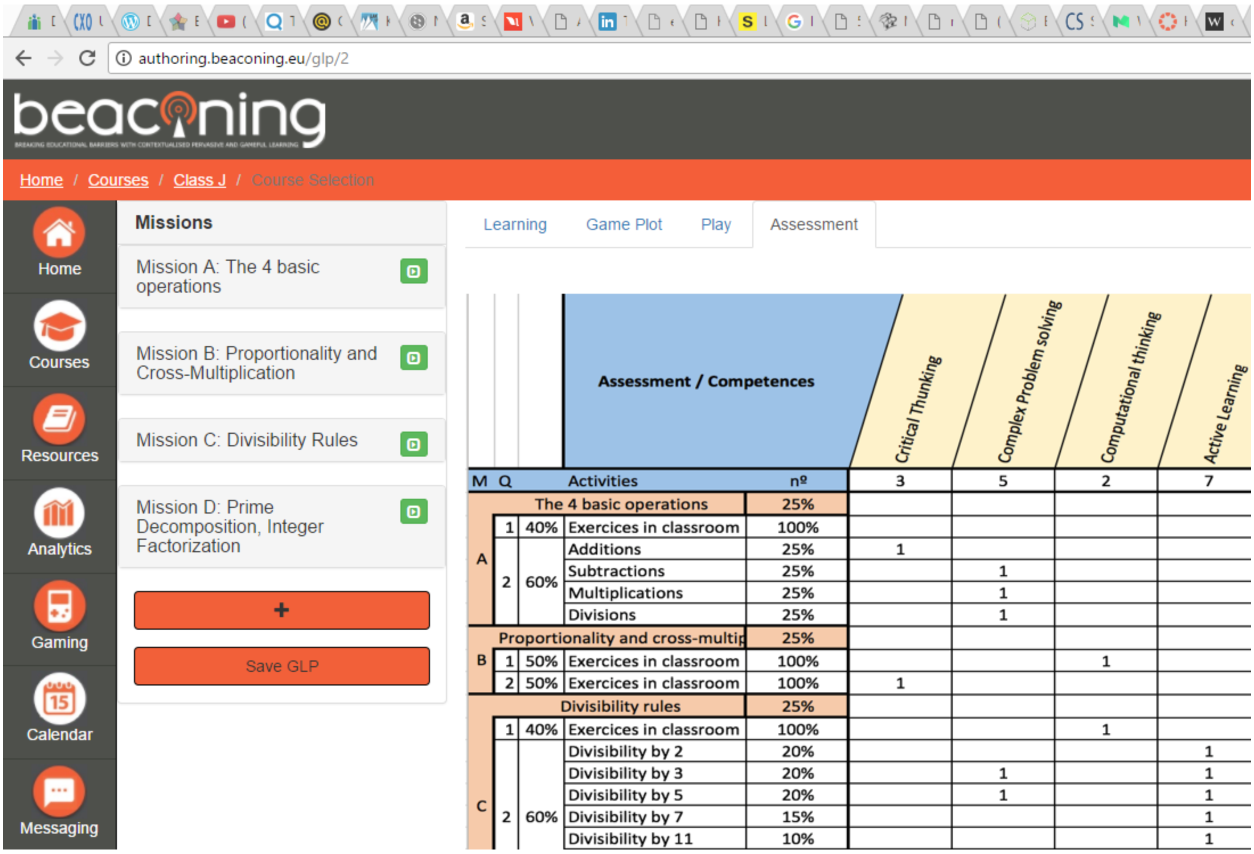


Figure 5 Screenshot from BEACONING Authoring Tool (mock-up screen at the time of taking)

The percent link between quest completion and the teacher who guides their class during the play-lesson activity finalizes competency acquisition, at least at the expected level of achievement. Beyond quest completion, the evidences that are demanded from the participants can also be utilised as part of a rubric for competency assessment. Although it would be the teacher's responsibility to employ such rubrics, if common evidences can be identified as part of a BEACONING taxonomy or if the piloting partner has designed the pilot accordingly, the platform can have additional functionalities to track the competencies.

According to V. Shute & M. Ventura, educational measurement refers to the application of a measuring tool (or standard scale) to determine the degree to which educationally-valuable knowledge, skills, and other attributes have been, or are being acquired. It involves the collection and analysis of learner data. On the other hand, assessment involves more than just measurement, it also involves interpreting and acting on information about learners' understanding and/or performance relative to educational goals[11].

Quest completion is an evidence of progress, but it is a challenge for educators who want to employ or design games to support learning to make valid inferences. In ECD, every evidence can be treated as an item of measurement, which informs the design of valid assessments and yields real-time estimates of students' competency levels[12].

[11] V. Shute & M. Ventura (2013), *Stealth Assessment in Digital Games,* MIT Press

[12] R. Mislevy, R. Almond, & J. Lukas (2003) A Brief Introduction to Evidence-centered Design, Educational Testing Service, Research Report. Accessed on 22nd May, 2017 at https://www.ets.org/Media/Research/pdf/RR-03-16.pdf

Evidences are behaviours or performances that should reveal the tracked competencies. The next section provides sample gamified lesson plans where the evidences are used as part of a rubric for evaluation, as well as a sample which aims to use additional analytics functionality for automated evaluation. For such automated evaluation, evidences must be derived from data in the following typical steps

1. Data retrieval
2. Feature extraction
3. Evidence synthesis
4. Evidence accumulation

These steps represent a funnelling graph, which receives many data points and processes them to yield the evidences. Naturally, if the task models in Evidence-Centred Design are crafted to *elicit the expected behaviours that comprise the evidence,* then this graph can get shorter.

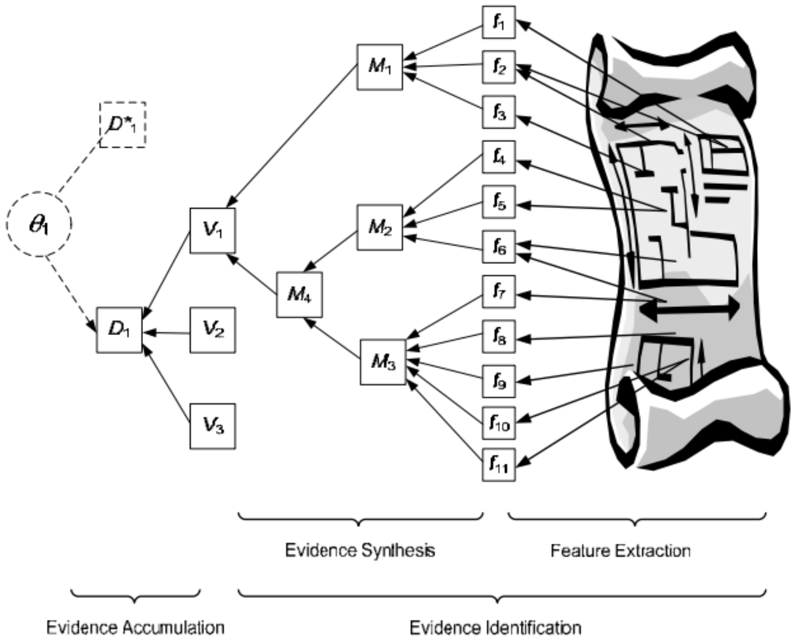


Figure 6 Evidence funnelling graph as represented in Braun (2006) [18]

The evidence funnelling process (called response processing in the IMS specification) is responsible for identifying the key features of the observable task outcomes. These outcomes can be work products such as a sequence of actions, time-based challenges or a short answer. Note that, such outcomes can also be used to provide the student with task-level feedback, which further strengthens ECD.[13]

In case of rubric based evaluation or simple check on quest completion, the Statistical Model that links the accumulated evidence to the Competency Model is just a linear sum of scores or weights marked by the teacher. However, in specific pilot implementations more complicated but more rigorous statistical models such as a Bayesian Belief Networks can be employed.

[13] R. Almond, R. Mislevy, L. Steinberg, D. Yan & D. Williamson (2015), Bayesian Networks in Educational Assessment, Springer, pg. 36

In case of using the probabilistic approach, a Bayesian network of conditional probabilities has to be updated as the evidence accumulates. A conditional probability is the likelihood that a student (can be an average student, student type or a particular student per se) is at one of the designated levels of a competency or a subcompetency.

The subcompetencies add further levels of depth to the student competency model. Generic competencies such as critical thinking or problem solving have been studied in various large-scale applications. Therefore it is fairly easy to construct a Belief Network subcompetencies and evidences about them.

	(1) Establishing and maintaining shared understanding	**(2) Taking appropriate action to solve the problem**	**(3) Establishing and maintaining team organisation**
(A) Exploring and Understanding	(A1) Discovering perspectives and abilities of team members	(A2) Discovering the type of collaborative interaction to solve the problem, along with goals	(A3) Understanding roles to solve problem
(B) Representing and Formulating	(B1) Building a shared representation and negotiating the meaning of the problem (common ground)	(B2) Identifying and describing tasks to be completed	(B3) Describe roles and team organisation (communication protocol/rules of engagement)
(C) Planning and Executing	(C1) Communicating with team members about the actions to be/ being performed	(C2) Enacting plans	(C3) Following rules of engagement, (*e.g.*, prompting other team members to perform their tasks.)
(D) Monitoring and Reflecting	(D1) Monitoring and repairing the shared understanding	(D2) Monitoring results of actions and evaluating success in solving the problem	(D3) Monitoring, providing feedback and adapting the team organisation and roles

Figure 9: Matrix of Collaborative Problem Solving. PISA 2015 definition of this competency is "the capacity of a person to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution"[14].

Once the Belief Network of competencies to be tracked is complete, it contains the entire set of competency variables, all of the evidences sought from every task and conditional probabilities that link them. While experts could provide their beliefs as priors for these probability tables, at each game level or student role, it may require considering a large number of configurations. However, design patterns or heuristics can be used under certain assumptions to select parameterisations. For example, well-defined evidences can be exclusive enough to consider conjunctive, disjunctive, or compensatory design patterns[15].

[14] OECD PISA Group (2017), PISA 2015 Collaborative Problem Solving Framework, accessed on 22nd May, 2017 at https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

[15] R. Almond (2015), An IRT-based parameterization for conditional probability tables. In J. M. Augusta and R. N. Carvalho, (Eds.), (Proceedings of the 2015 Bayesian Modeling Application Workshop at the 2015 Uncertainty in Artificial Intelligence conference. Accessed on 22nd May, 2017 at http://pluto.coe.fsu.edu/RNetica/

For updating the belief network based on evidence accumulation process, probabilities of the proficiency variables will be adjusted using the Bayes theorem. The values before the evidence arrives are called **priors** while the updated values are called **posterior**. Note that the posterior distribution after processing a batch of evidence becomes the prior distribution for the subsequent batch. At any point in time, there is a **current** proficiency distribution, which can be used to produce scores or to make inference claims about the student. Therefore, this approach is also called Ongoing Assessment or **Continual Assessment**[16].

The BEACONING platform provisions a technology-rich environment where high quality, ongoing, unobtrusive (implicit) assessments are feasible and can be aggregated to set expectations on the evolving competency levels of a student. Moreover, these assessments can be aggregated across many students to inform higher-level decisions about the on-going educational practice (e.g. from student to class to school to region to country).

3.1.3 Evidence-Centred Design in Beaconing VET Pilots

A proposed evaluation for BEACONING VET play-learn solution adapted is the delivery system model (DSM)[17] given the gaming basis is a platform of practice and experience. Here, the DSM comprises three base models: Student, Task and Presentation. Evidential models capture and assess transactions between and across DSM base models. They are serving as indicators to provide validation and verification of systems-level (**Figure 9 DSM and evidence units at systems level.**) and operational-level (**Figure 10**) performance. Systems-level evaluation focuses on functionality or usability, i.e. the technical elements of the system. **Figure 9 DSM and evidence units at systems level.** groups the DSM models as Learner, Lesson Designer, and Game Assembly, each with a corresponding evidence model. Operational-level evaluation concerns digital pedagogy, i.e. the effectiveness of BEACONINGs game-pedagogy to deliver STEM education and develop competencies. **Figure 9 DSM and evidence units at systems level.** presents the operational groups as Learner, Lesson Path, and Game, with corresponding evidence models.

[16] K. DiCerbo, V. Shute & Y. Kim (2016), The Future of Assessment in Technology-Rich Environments: Psychometric
Considerations, J.M. Spector et al. (eds.), Learning, Design, and Technology, pp 1–21, Accessed on 22nd May, 2017 at http://myweb.fsu.edu/vshute/pdf/DSK.pdf

[17] Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. ETS Research Report Series, 2003(1).

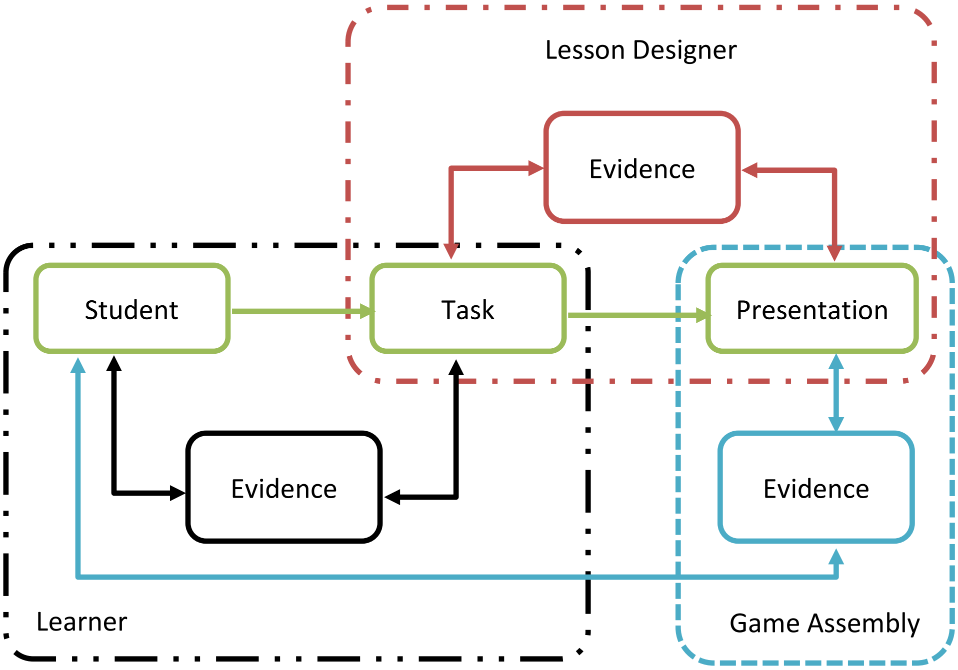


Figure 7 DSM and evidence units at systems level.

The Student model generates information pertaining to proficiency of applied knowledge, skills, and abilities. Proficiency variables are generally the performance criteria associated with the learning outcomes for a Task; for the VET example, these correspond to Table 1**: Sample systems, operational level variables, and indicative evidence**.

At systems level, Task model variables describe key pedagogical features implemented by the lesson designer to meet the learning outcomes. Pedagogical requites dictate that in-game interactions pertaining to learning has to be driven by extracting the relationship between games and learning mechanics (LM-GM)[18] at the abstract and concrete level; the fundamentals behind what links a game design pattern to a pedagogical pattern. The Presentation model variables at systems-level describe how tasks appear in the assembled game, i.e. the LM-GM mappings that organise the material to be presented and captured.

---

[18] Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., Freitas, S., Louchart, S., ... & De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, *46*(2), 391-411.

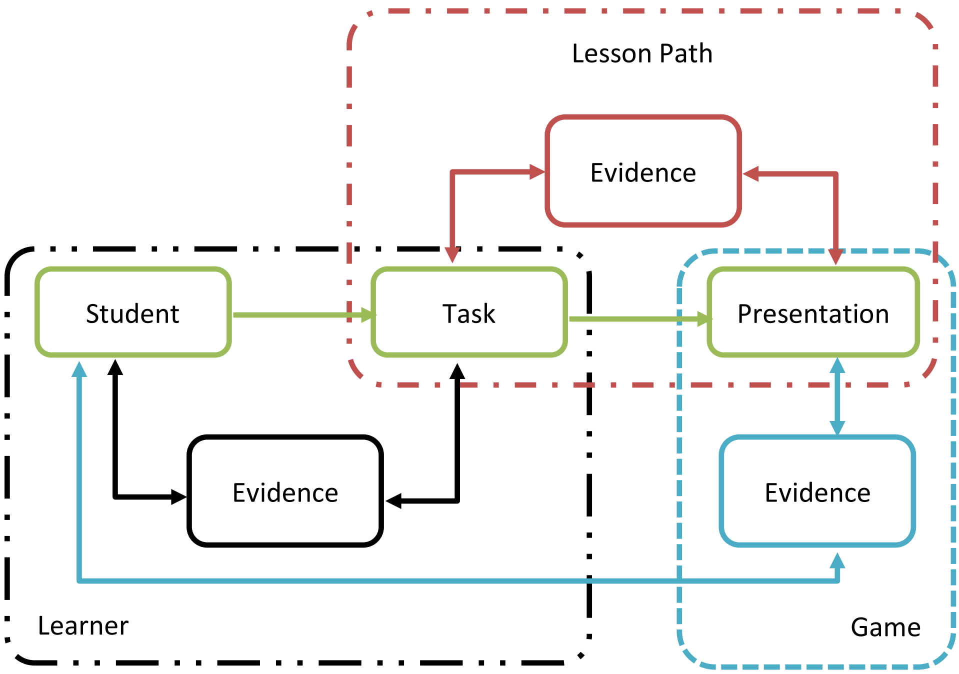


Figure 8 DSM boundaries demarking generic operational-level assessment

At operations level, Task model variables describe key learning features that align with concrete LM-GM elements to execute scaffold learning, tracking response time, and to score performance during simulation-based assessment of problem solving. Presentation model variables represent materials, stimulus, and prompts, including instructions for a task. They assess the presentation process ability to display the task to the student and capture the results when the student performs the task.

All evidence models contain instructions or rules on how to update information based on variables that define the performance of a transaction. Evidence models comprise of composite variables to be observed programmatically (e.g. a game trace) and/or by users and administrators (human-in-the-loop). The variables summarise the primary outcomes of a DSM model, providing information to update an assessment, progression or feedback at that level. Evidence models are data-driven and comprise some mathematical form such as Bayesian IRT or parametric function that accumulates evidence across tasks. For example, the Learner evidence model synthesizes Student variables with LM-GM Task variables to provide evidence associated with the pedagogical pattern. At the operational level (Figure 9 **DSM and evidence units at systems level.**), the Learner evidence model synthesises Student variables with the concrete LM-GM Task variables. Here, the evidence model assesses the Lesson Path features important to the task, such as interactivity, scaffolding and scoring mechanisms. Likewise, a Lesson Path is assessed for its presentation and execution to meet the learning outcomes.

Globally, the International Society for Technology in Education (ISTE)[19] frames its benchmarks for digital literacy around six standards: creativity and innovation; communication and collaboration; research and information fluency; critical thinking, problem solving, decision-making; digital citizenship; and technology operations and concepts. The goal of technology to access opportunities to participate and develop competencies shifts from traditional literacy

[19] International Society for Technology in Education (2007) ISTE NETS: Advancing Digital Age Learning. Iste.org/nets.

taught in classrooms to one that is participatory through blended digital spaces[20]. A BEACONING VE-STEM game should assess knowledge and understanding of artisan and STEM skills. With a simulation-based problem-solving game, the evaluation emphasises the lesson design, environment, technology and learning experience across multiple platforms and formats through digital technologies. Table 1 lists sample variables taken in light of ISTE and education through digital media. These are suitable for the lesson path on stone masonry and generic to be applicable to a wide range of VET courses.

Table 1: Sample systems, operational level variables, and indicative evidence

Level	Variable	Evidence
Systems	Networking | Collaboration	Ability to search for, synthesise, and share information
	Negotiation | Community invariance | Multi-objective	Ability to span across diverse communities, discerning, composite multiple perspectives, provide alternative to norms, leading to goal objectives
	Transmedia	Ability to support, integrate, use multiple media platforms to create game-based learning to tell a narrative across time Ability to use different tech media platforms or break down the barriers between the lessons, story and reality by bringing the narrative out into the real world Ability to create or support alternative reality games (ARGs), where learners engage with narrative/game elements and characters using real world locations as part of the lesson path
	Appropriation	Ability to sample and remix media, content, digital pedagogical, interactive technology meaningfully Ability to adapt to learner preferences and environment for the purpose of multimodal learning formats
	Simulation | Presentation	Ability to construct dynamic lesson and learning models of real-world processes
Operational	Play	Capacity of lesson paths to enable learners to experiment, explore, interact with and/or between physical-real surroundings as a form of problem-solving
	Appropriation |	Ability for learners to sample, evaluate the reliability

[20] Jenkins, H., Purushotma, R., Weigel, M., Clinton, K., & Robison, A. J. (2009). Confronting the challenges of participatory culture: Media education for the 21st century. MIT Press.

	Intelligence	and credibility of different information sources Capability for learners to analyse and translate presented in-game data into information to application
	Cognition | Judgment	Ability to interact meaningfully with tools that expand mental capacities Ability to pool, discern and compare knowledge towards building subject mastery/STEM competency Ability to interpret and construct links with real-world processes
	Multitasking	Ability to scan one's environment and shift focus as needed to salient details Ability to search for, synthesise, and disseminate information
	Negotiation	Ability to communicate and interact across diverse communities, multiple perspectives, grasping and following alternative norms

3.1.4 Examples from Gamified Lesson Paths

Some of the BEACONING gamified lesson plans are given in D4.8 Gamified lesson plans (M18). These plans are designed to include gaming activities that also yield evidences of competency achievement. The mapping of this evidence to a set of target competencies and is a part of the authoring process supported by the BEACONING authoring tool. However, the process by which this mapping to levels of achievement is evaluated can be achieved by quest completion rates (triadic approach), by rubrics (that teachers can use) or by probabilistic measures (such as Belief Networks), depending on the pilot study.

Example Gamified Lesson Path for Triadic Approach: Studying the power plants – Physics-High school

In this plan, players are split into small groups in order to participate in a contest having to do some missions against time which are mini-games/challenges about the impact of energy production over the environment (See Appendix A). These challenges involve quizzes, comparisons and matching that would require critical thinking and collaborative problem solving competencies. As the plan proceeds, quest completion rates are tracked and displayed on the BEACONING dashboard.

Example Gamified Lesson Path for Rubrics-based Evaluation: Health and Safety controls for 9" Angle Grinder in stone cutting use

Players learn as individuals, with game-mechanics to promote and develop errorless learning. All BEACONING VET lesson paths embed errorless learning, taking into account STEM competency development across a wide spectrum of trade skills. The general form within each lesson path is to make aware and highlight the importance of Health and Safety when working in and around a construction site, built environment or restorative site. All play learning objectives and outcomes must conform to, or meet, accreditation standards, industry standards and occupational standards. In stone masonry lesson paths, Mission and Quest designs must demonstrate performance and knowledge criteria in all cases. VET learning designers should consider lesson paths that can transit across formal, informal and workplace learning. Mini games (quests) support each mission's Learning Aim and Objectives through specific defined learning outcomes. Mission A: Establishing pre-requisites; Mission B: Stone masonry standards and interpreting information; Mission C: Adopting industry relevant, safe, and healthy working practices; Mission D: Quantification of resources, trade skills, VE-STEM; Mission E: Applying tools (moving, handling, using, storing) and knowledge on operational and occupational safety.

There are various criteria for stone masonry skills and competencies assessment; learning designers should work closely with teachers and industry to implement the correct level of study. One particular set of performance criteria is given by COSVR195 Produce standard architectural stone enrichments (Source: National Occupational Standards). These criteria serve as rubrics for evaluating mastery level (See Appendix A)

Example Gamified Lesson Path for Probabilistic Evaluation: Middle School Maths for celebrity house design

This gamified lesson plan follows the "learning as a game" concept and engages students with a hands-on design task where they would need to use Math topics such as fractions, measures, ratios, metric system. They will design the construction plan of a house and a garden by using paper and pencil before implementing their design in a PC constructions game (Minecraft Education Edition). They will then need to update their plan upon a customer change request. Finally, they will present their designs. Meanwhile, the teacher tracks identified evidences of competency throughout the lesson about which prior probabilities are obtained from expert judgement. At the end of the plan, these probabilities are used to calculate likelihood of achievement. The tracked competencies are not limited to Math skills. A Belief Network about Critical Thinking is also tracked via evidences (See Appendix A).

Example Gamified Lesson Path: Algebra

This specific Lesson Path represents the most basic use of the Beaconing Platform, following a linear path of quiz-like activities delivered through Mini Games. The potentiality of the Beaconing Platform to be context sensitive here leans on its on-demand nature and on the networking of data across its whole community, more than on the activities themselves.

3.2 METHOD

A four step approach is recommended to construct the competency, evidence and task models:

> **Step 1** – Build the Student Competency Model for the generic competencies given earlier in this section. If a probabilistic approach will be used, then while building the graph structure, capture the conditional probability prior values as beliefs of the experts who prepare the belief structure.
>
> **Step 2** – Select a Gamified Lesson Path, examine the games, activities, mini games in the plan, as well as the indicated assessment events in order to build an Evidence Model where each evidence links to a node or nodes in the Competency Model. The links can be quest completion rates or some rubrics or probabilistic.
>
> **Step 3** – Analyse the mapping of the Evidence Model to the Competency Model to discover if there are any major gaps in substantiating some of the (sub)competencies.

Step 4 – Revise the Gamified Lesson Path by adding/modifying activities/quests that fill in the gaps and form the Task Model.

Upon preparing the three models and building the links in-between this Evidence-Centered Design has to be implemented. The implementation can be with paper and pen tools or simple spreadsheet software or it can be done with available software packages.

The implementation steps are recommended by Almond, Steinberg, and Mislevy as a Four-Process Architecture, namely Activity Selection, Presentation, Response Processing and Summary Scoring[21]. Although this architecture is informative, our main purpose is not to build an assessment system but rather evaluate pedagogical value of the BEACONING platform, so our implementation method steps are proposed as

Step 1: Feature extraction, Evidence Synthesis and Evidence Accumulation. BEACONING data is kept in a Learning Records Store (LRS) as xAPI statements[22]. The LRS covers the feature extraction step. Depending on the design of the pilot study, evidence synthesis and accumulation steps has to be implemented, either by updating the software or by other means available to the piloting partner.

Step 2: Calculate observables from the work product, and fill in the new information in the Evidence Model. Measurements need to be able to handle any number of attempts and revisions of solutions during game play. This can be achieved by comparing the outcomes of consecutive actions/events. Operational constraints (e.g. time) may also impose a limit on the number of attempts allowed.

Step 3: Use the evidence to update the values for tracked competency variables. The variables can be monitored via the BEACONING platform or externally depending on the pilot design.

3.3 PEDAGOGICAL EVALUATION PLAN DURING SPRING 2018 PILOTS

Thesteps of how we intend to apply the proposed methods is presented below:

1. September 2017: Develop simple competency maps for STEM skills (eg. Using ISTE standard) and enter evidences for them in Gamified Learning Activities
2. October-November 2017: Test the evaluation approach within small pilots and their outcome. Use the feedback for possible refinements
3. January-February 2018: Finalise authoring tool features if the ISTE standard would be built in there. Otherwise, update Teacher Training material.
4. March 2018: Finalize play-lesson plans with learning designers (or pilot teachers) to ensure that the plans that will be exposed at the large pilot are all evidence-centered.
5. April 2018: Launch first set of controlled experiments.
6. May 2018: Launch second set of controlled experiments with refinement at step 4

[21] R. Almond, L. Steinberg, and R. Mislevy (2002) , Four-Process Architecture for Assessment Delivery, with Connections to Assessment Design, Educational Testing Service, Technical Report, Accessed on 22nd May, 2017 at https://www.education.umd.edu/EDMS/mislevy/papers/ProcessDesign.pdf

[22] A. Serrano-Laguna, I. Martines-Ortiz, J. Haag, D. Regan, A. Johnson, B. Fernandez-Manjon (2016), Applying standards to systematize learning analytics in serious games, Computer Standards & Interfaces. Available at http://www.e-ucm.es/drafts/e-UCM_draft_297.pdf

Detailed information's regarding the location, the testing population characteristics, the content and the duration, are described in D6.2.

4 ACCEPTANCE AND IMPACT EVALUATION GUIDELINES

The acceptance and impact evaluation will be committed at the second round of large-scale pilots. Acceptance evaluation will be done using Technology Acceptance Model 3 structure. TAM3 construct is shown to be valid with high reliability. To employ the TAM3 construct, a survey of 12 questions is administered to users who had enough time with a certain technology. Each question corresponds to a factor that is shown to relate and effect acceptance through a number of latent variables.

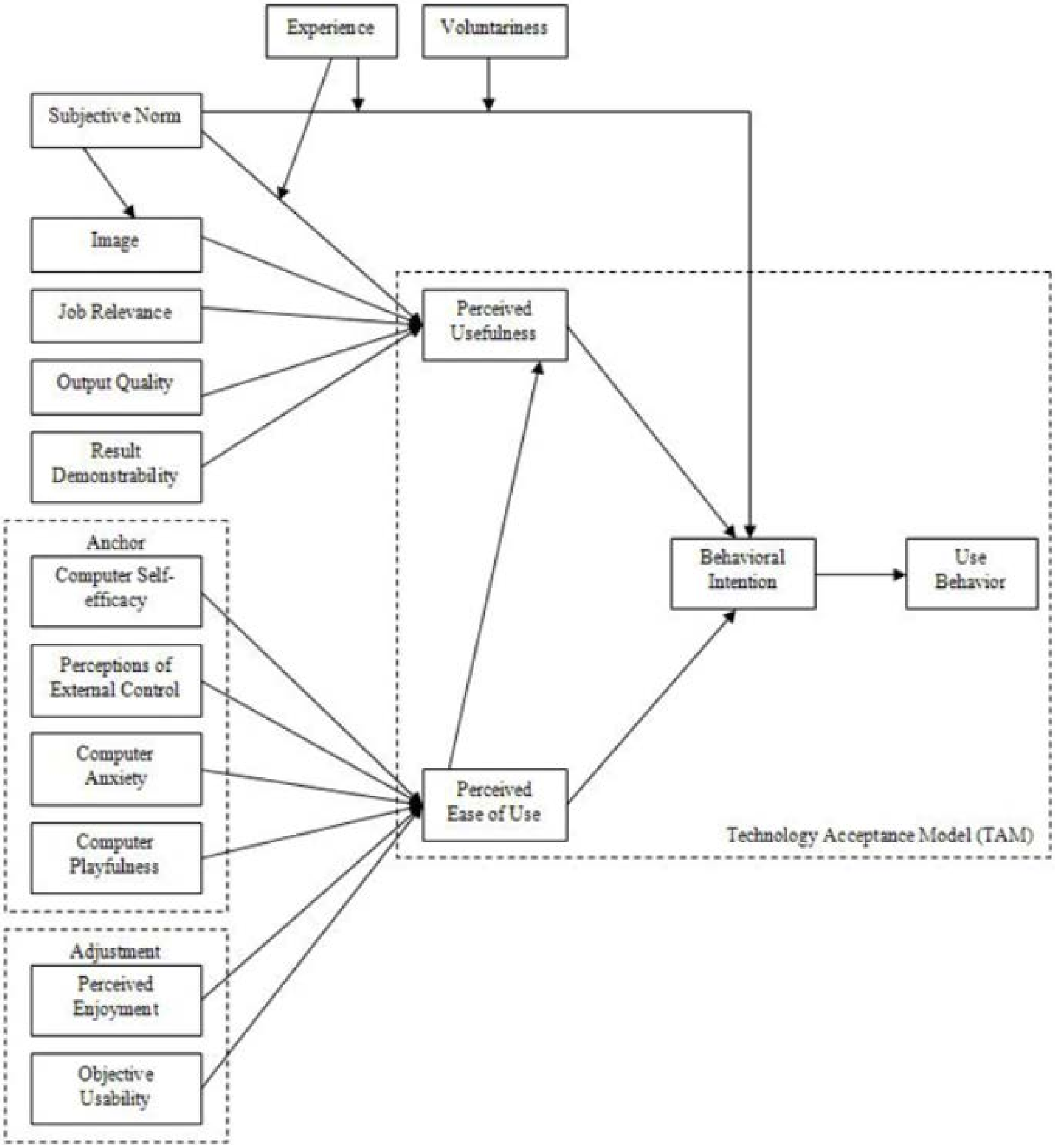


Figure 9 Technology Acceptance Model 3

As such the TAM3 construct is a structural equation[23] which models the acceptance behaviour of users. If the "assumption" that the users will accept, i.e. intend to use BEACONING platform,

[23] R. B. Kline (2016) Principles and Practice of Structural Equation Modeling, Fourth Edition.

then the factors that are surveyed vary in a similar way (high covariance). **Structural Equation Modelling** is to test a "theory" by specifying a model that represents predictions of that theory among plausible constructs measured with appropriate observed variables[24] (Hayduk et al. 2007). Note that the word "theory" is used in a broad sense including any kind of (probabilistic) causal relationship.

We will then expand the TAM3 construct adding new variables that reflect the impact assumptions about BEACONING. "HOW" the impact will happen is listed in DoA Part B, Section 2. These are

1. Contribute to the objectives of the "Opening up Education" initiative:
    a. Facilitate open, connected and contextual learning;
    b. Transfer holistic play-learn approach across learning domains;
2. Enhance the development of digital learning and teaching resources, including for children and adults with mental or physical disabilities:
    a. Modular process for scalability and sustainability;
    b. Contextualised learning and assessment;
    c. User experience design;
    d. Blending spaces and contexts;
    e. Empowerment of key actors;
3. Speed up the rate of adoption on technologies for the modernisation of education and Training.
4. Enable faster ways of testing fundamental business hypothesis (including continuous development and testing with users) and increased skills capacity:
5. Facilitate the emergence of innovative businesses and create a digital learning ecosystem in Europe:
6. Increase the number of public-private partnerships addressing technological challenges for modernizing and improving education and training:
7. Reinforce European leadership in adaptive learning technologies for the personalisation of learning experiences:

For such activities to cause an impact, firstly some conceptual assumptions on the "value" of BEACONING should hold. These are listed as follows:

1. Using BEACONING Ecosystem and App, both students and teachers experience blissful productivity with their mobile devices, regardless of their gender, abilities or even certain disabilities.
2. The Authoring Tool makes it easier for teachers to teach and evaluate STEM competencies.
3. The Authoring Tool conducts the teachers preferences for adaptivity effectively for i) struggling learners ii) context iii) on STEM areas and in particular with lesson paths that will cover inter-disciplinarity STEM features (such as coding and robotics or coding and Mathematics).
4. Analytics Dashboard provides formative feedback on not only curricular competencies but also gaming, learning skills and cognitive faculties.
5. Analytics functionality certifies expertise correctly by direct measurement and continual assessment.

[24] Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three—Testing the theory in structural equation models! Personality and Individual Differences, 42, 841–850.

4.1 IMPACT EVALUATION APPROACH

We will use explicit evaluation using Structural Equation Modelling [25] for testing if value theories on BEACONING platform and components hold. Since all value theories relate to perceived ease of use and perceived usefulness we can update the structural equation in TAM3 and design a new model, rather than developing a new structural model about the value of BEACONING. The evaluation will take place in four stages:

Stage 1 – Obtain the demographics of the participants: Different populations among the participating students, teachers, and schools will be identified and tagged on the anonymous data (attending elite vs. less selective schools, attending students with STEM vs. humanities and social science majors, attending students who struggles vs those who are on track). The segmentation that profiles different populations at different kinds of institutions will provide useful information, and insight into the aggregate experience of participants as well as the variations in the experiences of different profiles.

Stage 2 – Allow enough time with the system: The information system whose acceptance is to be modelled has to be used long enough by the users to have the users shape their ratings to statements such as "Using the system would improve my job performance." Although the impact of eLearning systems on performance takes many years, basic indicators for positive affect surface in typically 4 weeks. This is because involving a technology component in a social process would naturally transform that process in time. TAM3 reveals which aspects of the system are critical in shaping users intentions to sustain the actual use of the system for so long.

Stage 3 – TAM3 Survey: A standard survey for capturing the TAM3 input factors is below:

PERCEIVED USEFULNESS		**1**	**2**	**3**	**4**	**5**	**6**	**7**		**NA**
1. Using the system in my job would enable me to accomplish tasks more quickly	unlikely								likely	
2. Using the system would improve my job performance	unlikely								likely	
3. Using the system in my job would increase my productivity	unlikely								likely	
4. Using the system would enhance my effectiveness on the job	unlikely								likely	
5. Using the system would make it easier to do my job	1unlikely								likely	
6. I would find the system useful in my job	unlikely								likely	
PERCEIVED EASE OF USE		**1**	**2**	**3**	**4**	**5**	**6**	**7**		**NA**
7. Learning to operate the system would be easy for me	unlikely								likely	
8. I would find it easy to get the system to do what I want it to do	unlikely								likely	
9. My interaction with the system would be clear and understandable	unlikely								likely	
10. I would find the system to be flexible to interact with	unlikely								likely	
11. It would be easy for me to become skillful at using the system	unlikely								likely	
12. I would find the system easy to use	unlikely								likely	

[25] V. Shute & M. Ventura (2013), *Stealth Assessment in Digital Games,* MIT Press

Figure 10 A Standard TAM3 Survey with Likert-Type Scale

This standard survey will be expanded by items that capture data about additional factors that we will derive from the value theories about BEACONING.

Stage 4 – Model Formation: The model formation will be made in two steps as proposed by Anderson (1988)[26]. The first step is to check the internal consistency of the extended TAM3 model in application to BEACONING. This step will test the strength of the model. To do so, sampling adequacy will be checked using Kaiser-MeyerOlkin measure and statistical significance will be checked using Bartlett's test of sphericity. If these metrics prove that the measurements are adequate and significant, Cronbach's alpha and % variance explained will be calculated using factor analysis in order to find which factors in the extended TAM3 model has the greatest correlation and explanatory power to measure the participants behavioral intentions towards using BEACONING. In the second step, correlation analysis will be carried out to discover which anchors have the greatest effect on the behavioral intentions to use. Pearson correlation coefficients will be analyzed for this purpose. At this step, further inquiries can be made to discover if the results change depending on different user populations.

4.2 METHOD

1. Start with the TAM3 construct and the survey of 12 questions on subjective norms.
2. Make additional hypothesis such as "playfulness increases perceived ease-of-use"[27].
3. Add new "latent variables" to TAM3 (such as playfulness) and add new statements to the survey.
4. Run the survey to collect data and check your hypothesis by calculating factorial loadings.
5. If hypothesis holds (if the loading of playfulness on the perceived ease-of-use was high enough) use the new construct thereon, else update and run new survey.

4.3 IMPACT EVALUATION PLAN DURING SPRING 2018 PILOTS

1. January-March 2018: Develop additional hypothesis.
2. April 2018: Develop new surveys, translate and setup at Survey.
3. May-June 2018: Use the new construct at the spring large-pilots to validate the new structure.
4. October-November 2018: Use Survey in autumn large-pilots.
5. December 2018: Use the data and the new construct to validate impact.

[26] J.C. Anderson & D. W. Gerbing (1988), "Structural equation modelling in practice: a review and recommended two-step approach," Psychological Bulleting, Vol 103, pp. 423-441

[27] A. Padilla-Melendez et al, (2013) "Perceived playfulness, gender differences and technology acceptance model in a blended learning scenario" Computers & Education, Volume 63, Pages 306–317

5 EVALUATION GUIDELINES FOR LOCATION-BASED ACTIVITIES

This part of document aims to describe evaluation processes which might be introduced when taking in consideration location-based activities.

Firstly, it is important to underline, that most of the time location based activities are supplementary to other learning paths and rarely a learning path will be designed with those activities at its core. What’s also important from the evaluation point of view is the fact that location-based services are integrated with other minigames and challenges. Thanks to Context Aware Games other minigames can be introduced to students in the more engaging context-based way.

Inside BEACONING platform there are 7 Context Aware Games which will be used to setup location-based activities. Depending on the scenario of the game, activity will have a different goal and will bring different value to students and teachers. List of games available inside BEACONING platform:

- Follow the path
- Treasure hunt
- Capture a flag
- Rat race
- Conquest
- Jigsaw
- Stratego

The detailed description of Context Aware Games and more information about Location Based Component can be found in D4.2- Location Based Component file.

Context Aware Games will most likely highly benefit from teachers and students feedback as it is quite a new approach to teaching and learning. That makes this feedback really important and potentially beneficial for future development to location based approach in these areas. Important matters that should be checked during the evaluation process of location-based activities:

- What are learning effects from paths, which include location-based activities compared to those which do not have such content?
- How do teachers rate experience of using Location Based Component to create location-based activities?
- What is the opinion of students about context-based way of learning and how do they compare it to previous learning experiences?
- Which position tracking technologies are teachers most likely to use?
- Is there any other added value provided by this way of learning and teaching which teachers and students can point out?

6 BUSINESS MODEL EVALUATION GUIDELINES

This chapter describes the evaluation approach the BEACONING consortium will apply in order to assess the business models developed in WP7. As described in D7.2, there will be different business models for different partners, depending on which product or service they intend to offer, and in some cases, a partner will operate with different business models. Whereas WP7 deals with the planning and development of new business models both for the different components as well as for the BEACONING platform, this task will provide guidelines on how to evaluate these models. Business models serve as a contextual property to various piloting activities so that, by the end of the project, each BEACONING component can be shown to reach a TRL mostly between 6 and 9. This means that quite a few components have a pre-competitive maturity level and to some extent a prototype, whereas other components will be ready for market introduction.

In addition, it has to be mentioned that according to the input from the consortium members regarding exploitation plans (D7.2) not only the business models for the technical components and the BEACONING ecosystem (i.e. products) needs to be evaluated, but also that of different learning services, since some of the partners intend to develop services. A key advantage of the BEACONING ecosystem is the interoperability aspects which allow easy integration of components/ products/ services (also third parties) leading to a large variation in possible business models. This, as well as the fact that part of our solutions will be prototypes, limits our possibilities to use common evaluation approaches for business models. We have therefore considered using a mixed approach, which will better cover all the uncertainties and differences. The evaluation approach should be applicable for all variations we can have within the consortium.

This variation is typically for research projects and a previous project (GALA NoE) has presented both success and failure cases for transferring the outcome of such projects into successful examples of innovation uptake[28]. Therefore, the starting point for our evaluation approach is based on the suggestions for key criteria for success in combination with the usage of a Business Canvas. Business Canvas is actually an approach for planning and developing future business models, consisting of 9 building blocks shown in **Figure 13**. These building blocks will deliver a complete view on the impacts of business strategies on a business model. Every business can be reflected with this model, but depending on strategy and the relation between the components, the weight of the building blocks will be different. This is also an advantage for the evaluation ensuring that no important issue for the business model is left out. Instead of the Business Canvas, it is also possible to use other BM approaches like ESSO, however the canvas is quite straight forward.

[28] Garcia Sanchez, R; Baalsrud Hauge, J; Kraume, H. (2014) D4.11 business modelling and implementation report 3, GALA NoE; https://studylib.net/doc/10221730/d4-11_for-internal-r

<table>
<tr><td rowspan="2">Key Partners</td><td>Key Activities</td><td rowspan="2">Value Proposition</td><td>Customer Relationships</td><td rowspan="2">Customer Segments</td></tr>
<tr><td>Key Resources</td><td>Channels</td></tr>
<tr><td colspan="3">Cost Structure</td><td colspan="2">Revenue Streams</td></tr>
</table>

Figure 11 Business Model Canvas (after Osterwalder & Pigneur, 2010)[29]

In order to use the business canvas as an evaluation tool and not only for planning, it is necessary to evaluate the nine blocks according to different dimensions, in principle this approach is similar to a PESTEL analysis, but with different dimensions. In our case, we would look into the economic (all factors for calculating EBIT, ROI, ROCIW etc.), technology, stakeholders and Learning (results will be provided from the evaluation described in chapter 3. In the evaluation of the business model, we would only look if the evidence of learning for a specific group is given, not the learning itself). This results in a matrix which will give an overview of critical factors. In addition, it is necessary to take into account if we are talking about B2B, B2C or B2G (business to government). It is unlikely that there will be any C2C solutions within BEACONING, but B2C is not unlikely for those institutions (like some of the research institutions and universities that have suggested to sell BEACONING as a learning Service)

In the next step each critical factor will be further assessed as a part of a risk analysis (see chapter 6 at Osterwalder and Pigneur (2010). Important in this part, is that we use a symmetric risk definition according to the ISO standard, so that the opportunities are also sufficiently considered. The outcome will be a set of scenarios in which the variables can be weighted differently (i.e. like using scenario analysis as a forecasting tool) and therefore support the evaluation of different options and trends. The drawback of this approach, like all multi-variate approaches, is that the process has to be applied to each specific component as well as for all dimensions. Since this is a manual process, it will therefore only be carried out for the most likely scenarios.

[29] A. Osterwalder & Y. Pigneur (2010) *Business Model Generation,* Wiley

7 CONCLUSIONS

This deliverable gives guidelines to all involved stakeholders to have a common approach to tackle the task of evaluating a large-scale learning project. We described the European benchmark for evaluation of large scale school pilots as part of the research and technology development Framework programme (iClass Project in FP6, iTec Project in FP7, Next-Tell Project in FP7) and, also, as part of the Competitiveness and Innovation framework Programme (Open Discovery Space and Inspiring Science Education).

We also dedicate attention to the pedagogical value evaluation guidelines, to the acceptance and impact evaluation approach with method and to the Plan for the Spring 2018 pilots. Pedagogical evaluation approach consists in assessment in games, Evidence-Centred Design in Beaconing K12 Pilots, Evidence-Centred Design in Beaconing VET Pilots and examples from Gamified Lesson Paths, the method and the Pedagogical Evaluation Plan during Spring 2018 pilots.

Regarding the evaluation, we add the correct processes which might be introduced when taking in consideration location-based activities..

The deliverable closes with the Business model evaluation guidelines, more exactly with the evaluation approach which the BEACONING consortium will apply in order to assess the business models developed in WP7.

8 APPENDIX A – DETAILED GAMIFIED LESSON PATHS

Coding and robotics

Description: In today's world robots and automated systems occupy an ever growing space, either as every-day life facilitators or replacements for human workers. Basic knowledge base about robotics and coding is slowly becoming a basic skill. This activity will allow the users to discover and apply entry level notions of coding by using the LEGO Mindstorm solution. The lesson path is divided in four missions presenting basic robotic knowledge. The lesson starts with a simple robot able to move all around under the direct orders of the user and continues with a more evolved version of the robot able to move by itself according to the programming it received. Then the robot becomes more sophisticated by being equipped with a whole range of different sensors (light/touch/colours/etc.) improving its capabilities. Finally, all robots
world, everyone leaves a certain amount of information about themselves on the Internet, with the data either produced by them or entered by others. All those fragments of information put together constitute a digital identity which, especially where that information is publicly available, can be used by others to discover that person's civil identity. In this sense, a created by the students will be pitted against each other in a competition.

https://onedrive.live.com/view.aspx?cid=e7ce1adadc77d64b&page=view&resid=E7CE1ADADC77D64B!1882&parId=E7CE1ADADC77D64B!668&authkey=!AJrRO_oM70cBcJI&app=PowerPoint

Basic algebra

Description: Today's world requires knowledge based on solid basic skills such as mathematic literacy. An alarming picture of the current situation in math literacy at high school level is provided by the PISA and TIMSS international surveys. Some basic mathematical notions learned in primary and secondary schools are still not fully mastered by some students by the time they arrive in high school. This situation will lead those students being unable to correctly follow the math curricula, especially in algebra, should they not able to build on this pre-existing knowledge. This activity will address the problem by providing activities designed to work on essential algebraic notions needed throughout the high school curricula.

https://onedrive.live.com/view.aspx?cid=e7ce1adadc77d64b&page=view&resid=E7CE1ADADC77D64B!1881&parId=E7CE1ADADC77D64B!668&authkey=!AJrRO_oM70cBcJI&app=PowerPoint

Digital Identity

Description: In today's digital digital identity is a version, or facet, of a person's social identity. The ramifications of the concept of digital identities, both legal and social, are a complex and challenging topic. Furthermore, somebody owns digital identity is persistent. It is important to know how to manage it at any point of someone's social life: at school, at work, looking for work, being retired, etc. This concept is especially important for people being vulnerable and at risk, like NEETS. This lesson plan is about learning the concept of digital identity, how to protect it but also how to improve it. This way we're covered, you have a small recap about what they are and where to find them.

https://onedrive.live.com/view.aspx?cid=e7ce1adadc77d64b&page=view&resid=E7CE1ADADC77D64B!671&parId=E7CE1ADADC77D64B!668&authkey=!AJrRO_oM70cBcJI&app=PowerPoint

Middle School Maths for celebrity house design

Taylor Swift

Taylor Swift

@TaylorSwift

Ana Sayfa

Hakkında

Fotoğraflar

Store

Download on iTunes

Videolar

Etkinlikler

Beğenenler

Gönderiler

Notlar

Sayfa Oluştur

Gönderiler

Taylor Swift

2 June

Hey you, are you a good designer? I want an impressive pool at my charming house with trees and lighting and a shadowy chilling spot. What you need to know is,

- My land is 4 dunams and it is surrounded by trees that are spaced at 10m inbetween.
- My house is a single flat square building whose area is $400m^2$.
- To the right of my house I have a $30m^2$ garage for my cars.
- My pets are like my kids. Meredith and Olivia Benson are my cats and they live at home with me, but my Dobermanns Bug ve Baby have their own $2m^2$ cabins at opposites corners of the garden (those two always fight with each other).

I am looking forward to your designs.

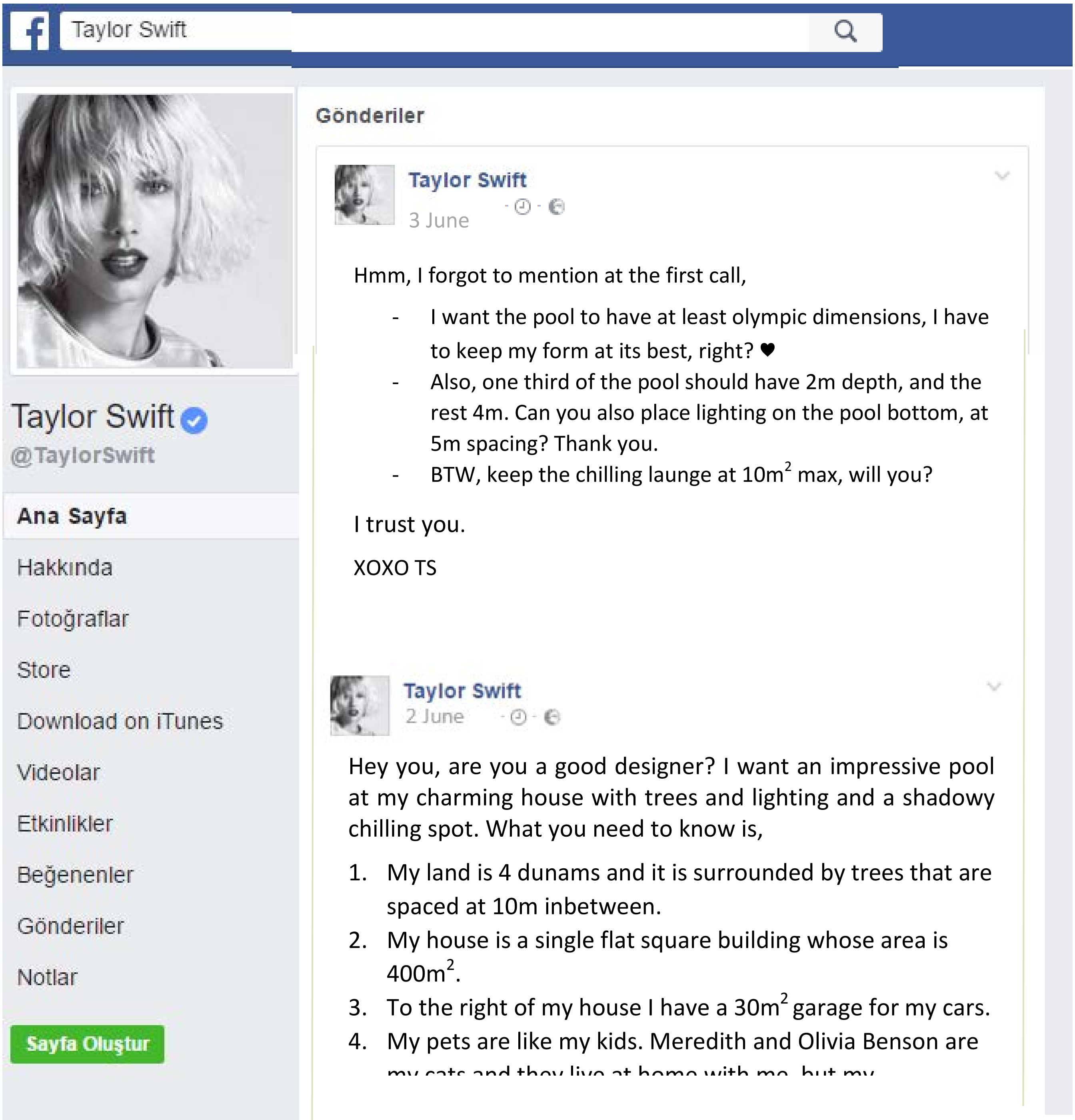
I wish to have dinner with the designer I like, at my home, at the pool side.

XOXO

TS

Taylor Swift

3 Mayıs



A list of evidences are identified and an Evidence Model is constructed that links to a simple Belief Network about Critical Thinking, which is a poly-tree of out-degree 3.

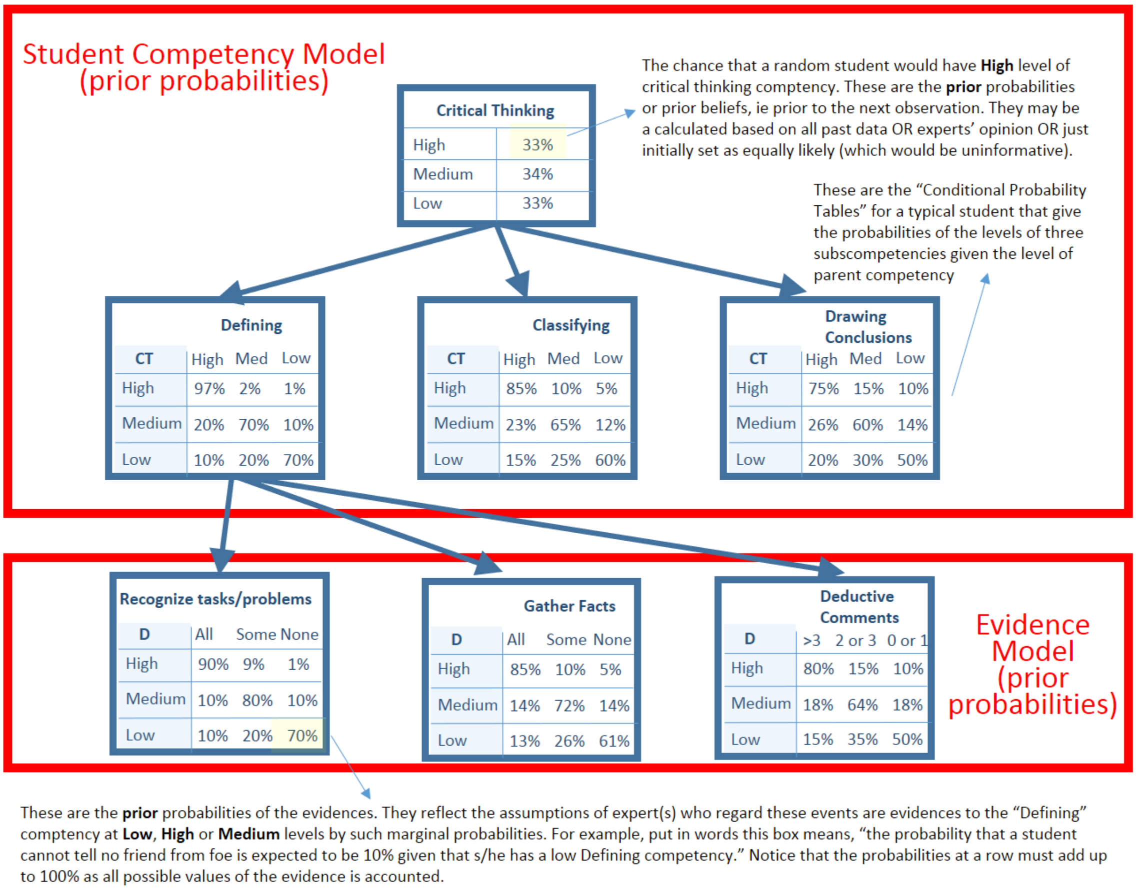


Figure 12 Example Student Competency Model and the Evidence Model of one of the subcompetencies in the model

Notice in **Figure 14** that all variables are discrete and having just three levels. All the Conditional Probability Tables (CPTs) are expert estimates considering a random or typical middle-school student. As new evidence arrives, the Evidence Model CPTs will be used to update the Defining subcompetency CPT, which in turn can be used by the Pearl's algorithm to update the Critical Thinking table. As a result the beliefs about such evidences pointing to a progress in "defining" propagates above to the main competency that is tracked and perhaps displayed at the analytics dashboard.

Gamified Lesson Path: Studying the power plants – Physics- High school

The players are split into small groups in order to participate in a contest where they have to do some missions against time.

Environment: inside or outside the school, at home, in the power plants

Tasks: given across missions

Interaction: walking around, doing tasks through mini-games/challenges.

The lesson starts by studying the power plants for assessing the impact of energy production over the environment, followed by requiring students to participate in a competition where they will have some missions against time.

Quest 1. After studying the types of power plants, students will receive 3 challenges:

- *Drag It! - with the cursor different types of power plants (with specific icons) over the map of Romania* - the game where students are provided with a map with missing elements. When the game starts, these elements start to drop from the activity header. Students will have to drag each element and drop it in the corresponding place, against time.

The game mechanics

- Learning designers can set a background picture and then position a set of items on that picture
- Students will see these items dropping and will have to drag each item and place it in the right place
- A timer will be presented; students will have to provide correct answers to win time and be able to place all items on the map
- The number of correct answers will be displayed
- A solution will be available for students when they run out of time; even if they run out of time, they will still be able to finish the game and place all elements in the right place, but the correct answers will not be counted after the time expires
- Students will have to select a level when they start the game: easy, medium and difficult; each superior level means less time to provide the correct answer
- A leaderboard will be displayed to show how students are ranking for that activity
- At the end, with the leaderboard, students will get instructional feedback with what they did wrong;

Player interactions

- Students will take the dropping item and place it over the blue dot; When an item is over a red dot, a sound is played and the blue dot turns yellow;
- If the answer is right, the item is positioned there;
- If the answer is wrong, the item drops rapidly and cannot be dragged again; the blue dot will turn red;
- An item that was not answered correctly, will be dropped again in order to be placed in the right place;

- *Match it! - A game where students have to match text labels (components of a power plant) with images to win time and manage to finish the activity.*

The game mechanics

- Learning designers can add a number of images to the scene and attach to each image a label
- Students will have to match the label with the image
- Doing so will win time for learners, helping them to finish the activity before the time expires
- If the time expires, correct answers won't count any further
- The number of correct answers will be displayed
- A solution will be available for students when they run out of time; even if they run out of time, they will still be able to finish the game and place all elements in the right place, but the correct answers will not be counted after the time expires
- Students will have to select a level when they start the game: easy, medium and difficult; each superior level means less time to provide the correct answer
- A leaderboard displays how students rank for that activity.

Player interactions

- Students have to drag one label over an image or vice-versa
- If the match is right, the items will flash and will disappear
- If the match is wrong, the border will go red, the user will lose a number of seconds and the items will bounce like they are rejecting each other

- *Millionaire Quiz - students will be provided with 10 -15 questions, with each category of questions being harder than the previous one, regarding the power plants.*

The game mechanics

- Learning Designers can add 15 question of their own choice
- Students will have to provide a correct answer for each question;
- If they get one wrong answer, they lose, but will still be able to complete the activity and see whether they would have won;
- A timer will be presented;
- The progress within the game will be displayed;

Player Interactions

- Students have to click the right answer;
- If the answer is right, it will flash in green for two seconds and switch to the next activity;
- If the answer is wrong, it will flash red for two seconds and a sound will be played; students will be able to continue, but without winning points;

In this case the game is pushing students' limits to improve persistence, being accompanied by the great sense of satisfaction they will get upon successful completion of the quizzes.

Gamified Lesson Path: Health and Safety controls for 9" Angle Grinder in stone cutting use – VET

Tasks: Health and safety topics are across missions, in correspondence to the work tool, processes and procedures for stone masonry. The exemplification here is Mission D: Quantification of resources, trade skills, VE-STEM (Table 2).

Environment: inside or outside the VET college, at home, at workplace.

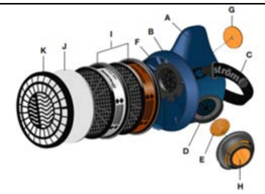
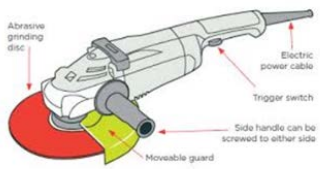
Interaction: The planned three modalities are mobile interactive apps, pc-based and cyber-physical system.

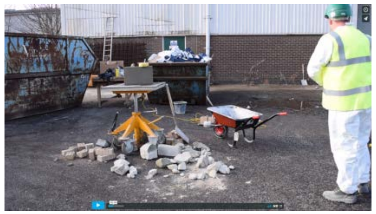
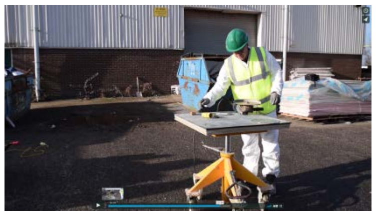
Number of Missions: 5

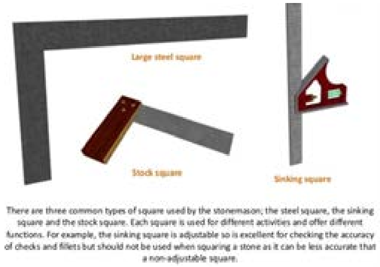
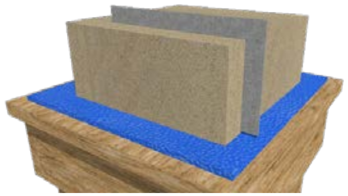
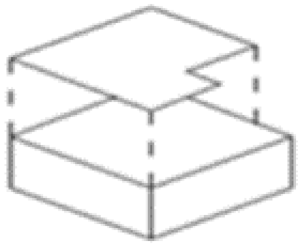
Table 2: Mission D with COSVR195 and SCQF Level 6 mapping.

LEARNING Aim: Mission D Quantification of resources, trade skills, VE-STEM	**Performance Criteria (PC) and Applied Knowledge and Understanding (K)**	**Lesson Plan Activity** Excess stone removal from a Quoin	**Game content**	**Essential learning Objectives (Errorless learning)** **Guided learning hours; 9**
Select the required quantity and quality of resources for the methods of work	PC 3: Selection of resources K9 the characteristics, quality, uses, sustainability, limitations and defects associated with the resources and how defects should be rectified K10 how the resources should be used and how any problems associated with the resources are reported	Q&A session covering the previous lecture to assess the learner's appreciation of the subject Ensure students understand the need to recognise and identify various;	*Which PPE/RPE tools are required for this mission?* 	**Health and Safety (RPE/PPE, Work Environment, etc..);** Top 9 things to be aware of: 1. Personal protection equipment 2. Dust suppression (Local exhaust ventilation (LEV; water suppression) 3. Carry out task specific risk assessment 4. Compile task specific method statement 5. Ensure the work area is regularly inspected and dust and cuttings are cleared away. 6. Engage in appropriate operator training by attending suitable off/on-the-job training 7. Respiratory protective equipment (RPE) with an APF

This project has received funding under the Horizon 2020 Framework Progra the European Union – Grant Agreement No. 687676

	K11 the organisational procedures to select resources, why they have been developed and how they are used K12 the hazards associated with the resources and methods of work and how they are overcome Select the most appropriate templates to enable production of replacement stone.	3.1 materials 3.2 tools and equipment Students are given a video demonstration of these areas and are briefed on the different types and their uses. Video + Q&A session covering the correct selection of resources associated with own work	  *Which hand and/or power tools are required for this mission?* - marker/ scriber, steel square sinking square and a steel ruler - Cutting tools (Hammer, pincher, cluerer, teethtool, grinder, blade) 	protection rating of 20-40 8. Operatives to be clean shaven to maximise protection of RPE 9. Face fit tested to ensure the RPE affords each individual the anticipated level of protection. **Safe working methods** (warm up/down, frequent breaks,etc.) Top 5 things to be aware of: 1. Pre-work protocol; warm up of body to alleviate possibilities of muscle injury/strain; Physical work-outs required to warm-up the body/muscles before starting + warm down after task completion 2. Provide scheduled breaks dependant on length of time operative use of grinder 3. Check for others in the vicinity and ensure safe distances are maintained. 4. Good posture to avoid fatigue 5. Healthy nutrition to provide energy **Machine controls** (Management Controls + Operator Controls); Top 10 things to be aware of: 1. Carry out a pre-use check – mains lead, switch functionality, no missing parts/modifications, blade and guard are secure, 2. Guards, fittings and blades not to be modified by operators at will. 3. Circular Diamond Blades designed only to cut not to be used for buffing/smoothing + for correct type of stone 4. Keep the mains supply lead under control (110v + residual current device (RCD)

				5. Ensure that the power flex and any extension cords are behind you and can't fall into the path of the cutting disk. 6. Portable appliance testing (PAT) recording 7. Carry out an after-use check – mains lead, switch functionality, no missing parts/modifications, blade and guard are secure. 8. If the blade jams, release the trigger immediately to allow it to come to a stop before removing it from the cut. 9. Make sure the grinder has stopped fully before placing down
Comply with organisational procedures to minimise the risk of damage to the work and surrounding area	PC 4: Minimise the risk of damage K13 how to protect work from damage and the purpose of protection K14 why disposal of waste should be carried out safely and how it is achieved prior to commencing	Group discussion on the student's knowledge of the H&S material. Lecturer provides class with videos/possible VR site visit and asks students to identify various H&S/OHS/ Sustainability methods/practices 4.1 protection of the work and its surrounding area from damage 4.2 minimise damage and maintain a clean work space 4.3 disposal of waste in accordance with legislation	  	**Health and Safety (**RPE/PPE, Work Environment, etc..);(as above) **Machine controls** (Management Controls + Operator Controls);

Comply with the given contract information to carry out the work efficiently to the required specification	PC 5: Meet the contract specification K15 how methods of work, to meet the specification, are carried out and problems reported K16 how maintenance of tools and equipment is carried out	Interactive quiz on work skills to measure, mark out, cut and finish masonry component and relevant terminology. Students are given a video demonstration of; 5.1 demonstration of work skills to measure, mark out, cut and finish 5.2 use and maintain hand tools and/or ancillary equipment 5.3 produce basic chamfers/checks/jointing to given working instructions for natural stone components, shaped true and square The students are then to complete the task by following a set of sequential instructions laid out by the lecturer.	*Is there a specific size or just a square? Why?*   *Which templates are required for this mission? How many zinc sheets should be placed? In which position?*  *What marking out procedure is*	**Machine handling** (operator's best working zone, body position, holding of machine, etc.); Top 10 things to be aware of: 1. Cut stone either at low level and mid-level 2. Hold the tool with a firm grip using both hands, one on the body of the grinder and trigger, the other hand on the side handle. 3. Keep two hands on machine at all times, one on the trigger switch and the other on the front handle + behind cutting edge 4. Position body to the closed side of the guard which provides more protection 5. Keep your feet spread apart. This braces your body better and makes it less likely for you to be thrown off balance. 6. When cutting on the ground, keep your legs out of the way in case the grinder slips. Do not hold stone with feet; Make sure you elevate the piece so that it is off the ground. 7. Ensure the workpiece is appropriately supported, considering its size, shape and the location of the angle grinder. 8. Avoid Kickback or jamming (the angle grinder suddenly thrusts back towards the operator) by not forcing the machine against the rotation of the blade 9. When cutting a hard stone allow eight of machine + your body to help 10. When cutting a soft sandstone take the bearing of the machine weight **Know Your Technique** (direction of cut, etc.) Top 6 things to be aware of: 1. Ensure dust is directed away from the body

			needed? From which datum surface prior to commencing? *Is there a specific number of chamfers? Where should they be placed?* Scribe on working lines of waste removal. Follow pre-scribed lines with angle grinder to full depth of blade. *Which angle and depth pf cut is the best/recommended if it varies from one type of cut to another?*  *Which tools are required? Why care must be taken not to pinch below scribed guidelines?* *Which angle grinding process needs repeated transversely or longitudinally? What is the basis of this specification?*	2. Never force the blade when using an angle grinder to cut stone. 3. Keep the grinder + blade at a right angle (90 °) to the work surface to produce a straight line 4. When cutting an angled surface do not exceed 20 ° 5. Minimise the series of the drafts/ruts required for stone if removing waste, however consider stone height and length. 6. Follow pre-scribed lines with angle grinder to required depth **Machine controls** (Management Controls + Operator Controls);(as above) **Health and Safety (**RPE/PPE, Work Environment, etc..);(as above) **Safe working methods** (warm up/down, frequent breaks,etc.)