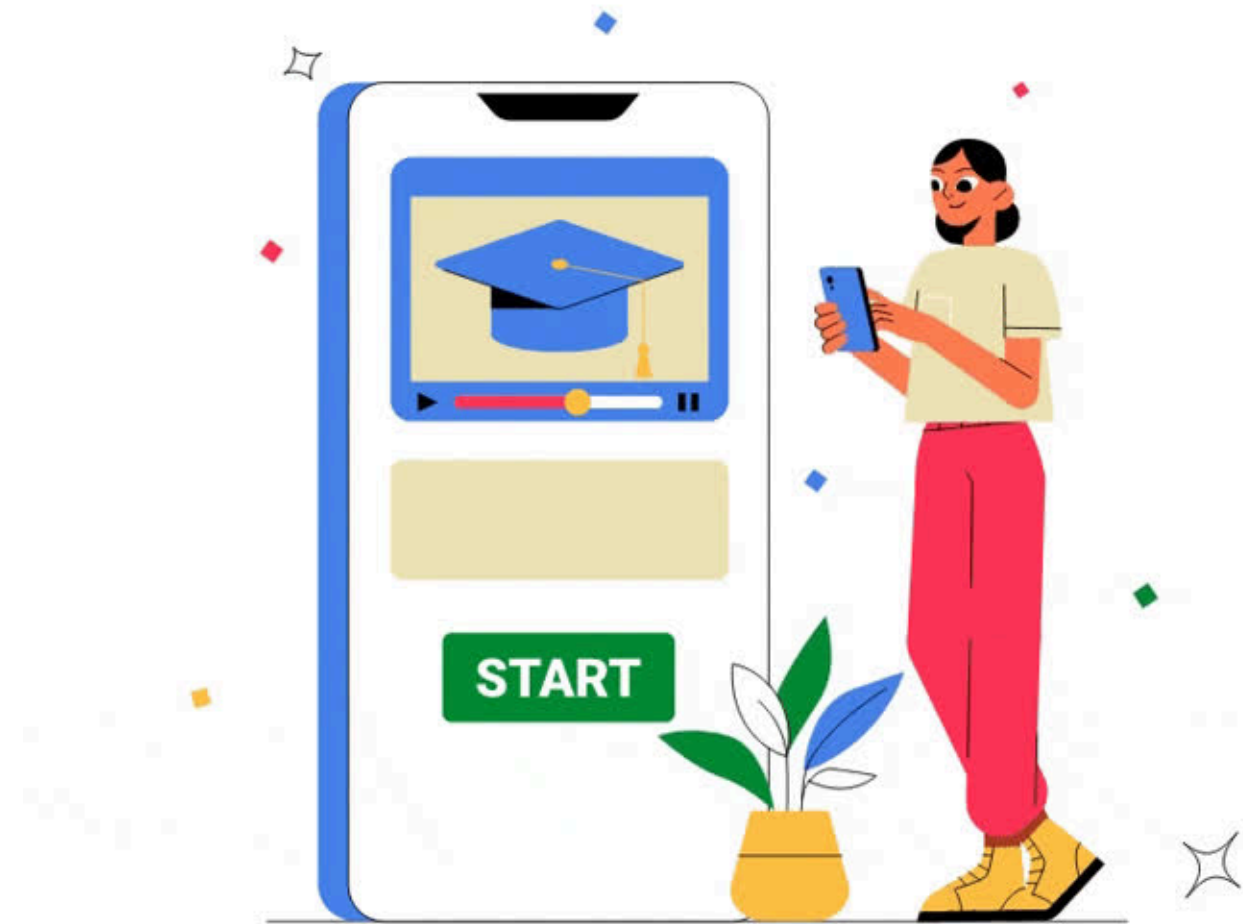




Co-funded by
the European Union



VYTAUTAS
MAGNUS
UNIVERSITY
M C M X X I I



STEALTH ASSESSMENT in SERIOUS GAMES

Ameny Rjiba

ameny.rjiba@gmail.com

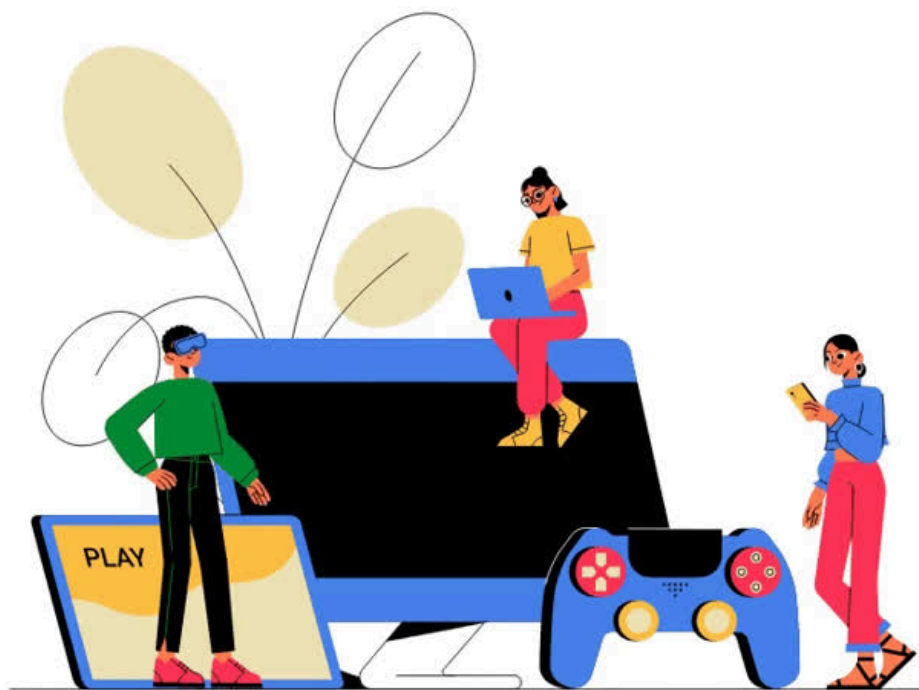
GBL

Serious Games

Known to : Engage learners
Promote knowledge acquisition
Foster 21st Century skills

They go beyond entertainment

Integrate educational objectives with interactive
gameplay



BUT ONE QUESTION REMAINS:

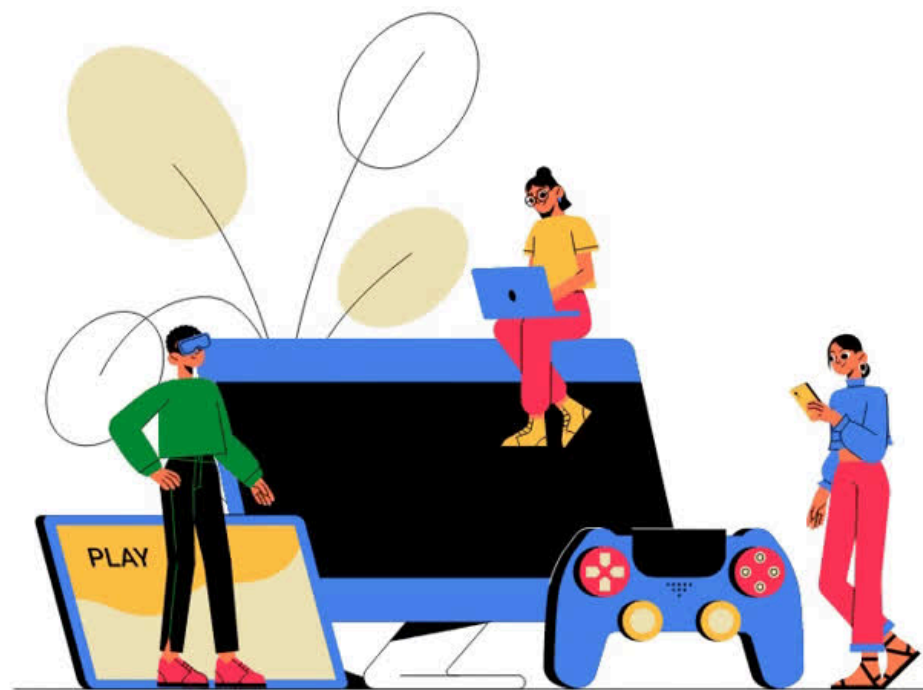


We talk a lot about
using serious games
for learning

How often do we
use them for
assessment?

WORDCLOUD

Stealth Assessment



What words come to mind when
you hear Stealth Assessment ?



WHAT IS STEALTH ASSESSMENT?

Embedding formative assessment seamlessly into gameplay so that players are assessed continuously.

Without interrupting the learning experience.



ANALOGY : FITNESS SMART WATCH

Your watch quietly tracks every heartbeat, step, and movement all day.

It doesn't just log data, it gives you real-time prompts ("Stand up now," "Slow your pace," "Time for a breath session")

You never stop what you're doing to fill out a survey or take a test

Assessment and feedback happen seamlessly



ANALOGY

CONTINUOUS DATA CAPTURE:

Just as your watch records every step and your heart rate throughout the day, a serious game logs every click, retry, and time-on-task.

INSTANT FEEDBACK

When your heart rate spikes, your watch vibrates and says 'Take a moment to breathe.'

In stealth assessment, the game might drop a hint or adjust the next challenge the moment a learner struggles.

NO INTERRUPTION

You don't exit your workout to fill out a questionnaire similarly, learners stay in the flow of the game rather than pausing for a separate test

WHY STEALTH ASSESSMENT?

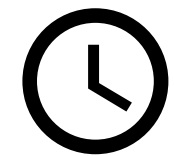
Traditional assessments (quizzes, exams, and surveys) don't translate well into game environments.

They pull learners out of the immersive experience, disrupt flow, and miss the rich behavioral data that games generate.

Stealth assessment was developed to address these gaps.



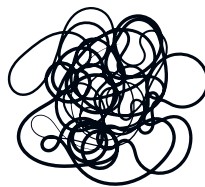
TRADiTiONAL



After learning (summative)



Obvious “test”



Can trigger test stress



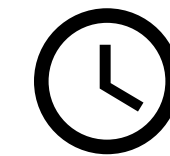
Delayed, often just a score



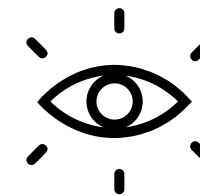
Surface knowledge



STEALTH



During learning (formative)



Invisible, part of the game



low anxiety



Real-time hints, adaptivity



Also complex skills (problem-solving, creativity)

WHY

Why It Matters - Key Benefits

Engagement & Flow

Keeps learners “in the zone” by merging play and assessment.

Richer Data on Skills

Captures process data (retries, time on task) for 21st-century skills like persistence and systems thinking

Adaptive Support

Automatically adjusts difficulty, gives hints exactly when needed.

Low Stakes

Because it “feels like play,” students take risks and learn from failure without anxiety

WHERE

Where Does Stealth Assessment Work Best?

Single-player serious games :

Ideal for assessing individual cognitive skills like problem-solving, persistence, and decision-making.

Example: A logic puzzle where retries and strategy changes are tracked.

Immersive environments (AR/VR)

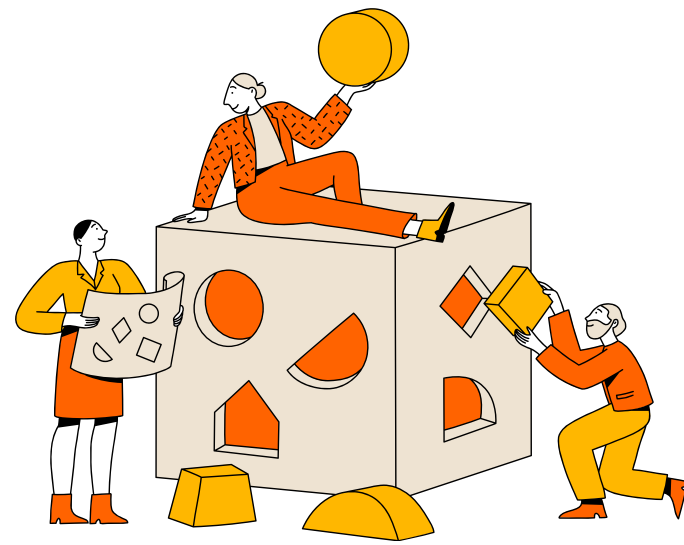
Allow for assessing spatial reasoning, adaptability, and real-time decision-making in simulated real-world contexts.

Example: Virtual labs or emergency simulations.

Collaborative multiplayer games

Excellent for evaluating teamwork, communication, shared planning, and coordination.

Example: Co-op missions where success depends on role distribution and information sharing.



What Can Stealth Assessment Measure?

THINKING SKILLS

PROBLEM-SOLVING

LOGICAL REASONING

**MEMORY AND
RECALL**

**SELF-MANAGEMENT
SKILLS**

PLANNING AHEAD

**REFLECTING ON
CHOICES**

**STAYING ORGANIZED
AND FOCUSED**

SOCIAL SKILLS

COLLABORATION

EMPATHY

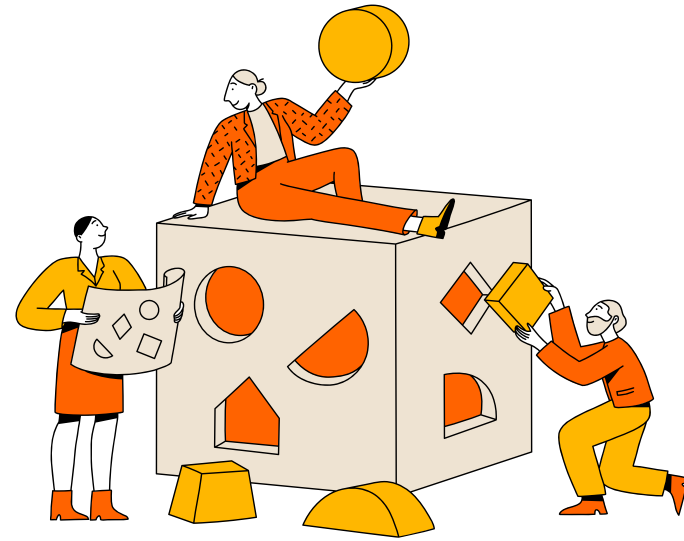
**NEGOTIATION AND
COMPROMISE**

BEHAVIORAL TRAITS

PERSISTENCE

**ATTENTION TO
DETAIL**

**WILLINGNESS TO
TAKE RISKS**



Sample Data You Can Log

TIME ON TASK

**SEQUENCE OF
ACTIONS**

USE OF HELP/HINTS

**FREQUENCY OF
TRIAL/ERROR**

**COMMUNICATION IN
MULTIPLAYER**

**TOOL/RESOURCE
USAGE**

HOW

How to Build a Stealth Assessment

The most adopted framework: **Evidence-Centered Design (ECD)**

Evidence-Centered Design (ECD) is a framework developed to guide the creation of valid, reliable assessments in games.

Introduced by Mislevy et al. (2003)

Adopted for stealth assessment by Valerie Shute (2005, 2011) to assess learning invisibly during gameplay.



EVIDENCE-CENTERED DESIGN HAS THREE CORE COMPONENTS:

Competency Model

Evidence Model

Task Model

1ST STEP

Competency Model

What do you want to measure?

This is about defining the target skill or knowledge you care about. Think of it as your learning goal.

Collaboration (Can players work effectively as a team?)

Critical Thinking (Can they analyze and solve novel problems?)

Persistence (Do they keep trying after failure?)

2ND STEP

Evidence Model

What behaviors in the game reflect that skill?

You identify observable actions that give clues about the hidden skill.

Collaboration: frequency of helping teammates, chat messages, shared decision-making.

Persistence: number of retries, time spent before quitting, willingness to try alternate solutions.

Critical thinking: how many paths explored, use of in-game tools, reflection before choice.

3RD STEP

Task Model

What kind of game task will make those behaviors show up?

Now you design or select game challenges that naturally push players to demonstrate the skill.

Persistence, give a puzzle that requires multiple failed attempts before success.

Critical thinking, present a problem with multiple solution paths and limited guidance.

Collaboration, create a mission where one player holds information the other needs.

CONCRETE WALK-THROUGH :

(PERSiSTENCE EXAMPLE)

Competency: Persistence, sticking with a challenge despite failure.

Evidence: Count **how many times** the player retries a level before quitting or requesting help.

Task: A “locked door” puzzle that can be solved by experimenting with different key combinations.

Implementation:

Log each retry, time between retries, and whether a hint was requested.

If retries exceed threshold, dynamically offer a scaffold (e.g. a clue pop-up).

Inference:

High retry counts + low hint use → strong persistence; low retries + early hint use → lower persistence.

CONCRETE WALK-THROUGH :

(COLLABORATION)

Competency: Productive collaboration

Evidence: Number and quality of chat messages, Frequency of shared resource use, Response time to teammate requests

Task: Two-player mission where one holds a key and the other must solve a puzzle, requires communication.

Implementation:

Instrument chat logs for help-offers versus off-topic chat.
If collaboration stalls, game prompts (“Ask your partner for the key”).

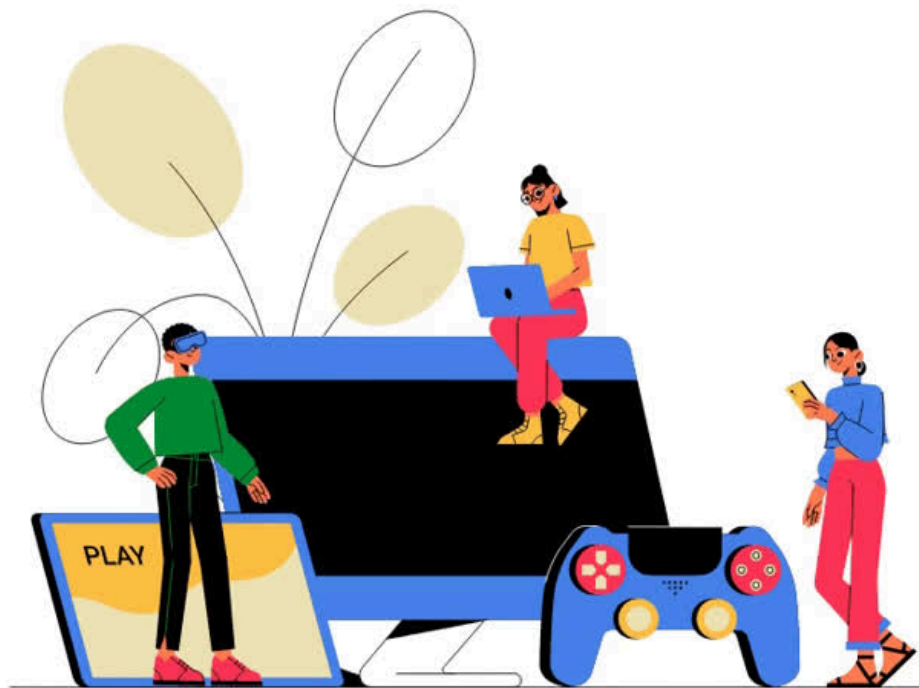
Inference:

Balanced, task-focused dialogue + mutual help behaviors → high collaboration competence.

QUIZ

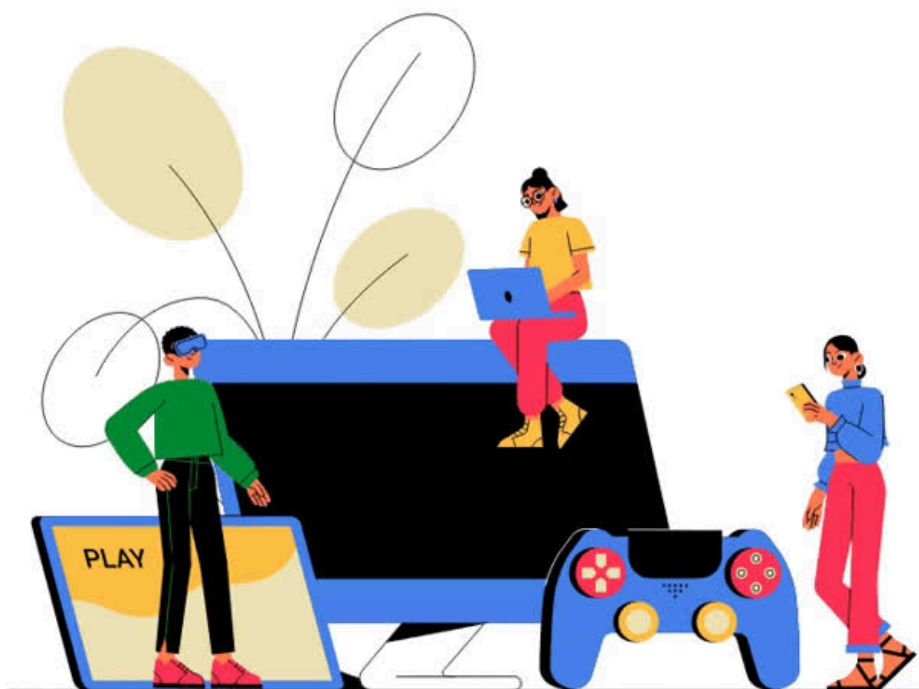
Is It Really Stealth Assessment?

Which of the following scenarios
reflect true stealth assessment
principles?



PLAY

Play a game : escape room



POLL

FEW QUESTIONS TO ANSWER



You thought you were just escaping a lab.
But all along...
You were being **assessed**.

Each decision you made gave clues, not about
your knowledge, but about your skills:
How you **think**,
How you **adapt**,
How you **collaborate**,
And how you **lead under pressure**.

Options:

- A) Start backing up digital data immediately.
- B) Alert the team and coordinate physical backup efforts.
- C) Investigate the source of the outage.

Skill Mapping:

- A → Time management & prioritization
- B → Collaboration & leadership
- C → Problem-solving under uncertainty

Options:

- A) Mediate and find middle ground.
- B) Support the cautious team member.
- C) Side with the risk-taker and proceed.

Skill Mapping:

- A → Conflict resolution & diplomacy
- B → Ethical decision-making
- C → Risk-taking behavior & assertiveness

Options:

A) Report the breach immediately.

B) Quietly close it and move on.

C) Skim it to understand its importance, then close it.

Skill Mapping:

A → Integrity & responsibility

B → Passive rule compliance

C → Curiosity vs ethical restraint

Options:

A) Try to fix it yourself using the manual.

B) Improvise an alternative solution.

C) Halt progress and wait for help.

Skill Mapping:

A → Technical resourcefulness

B → Creativity & adaptability

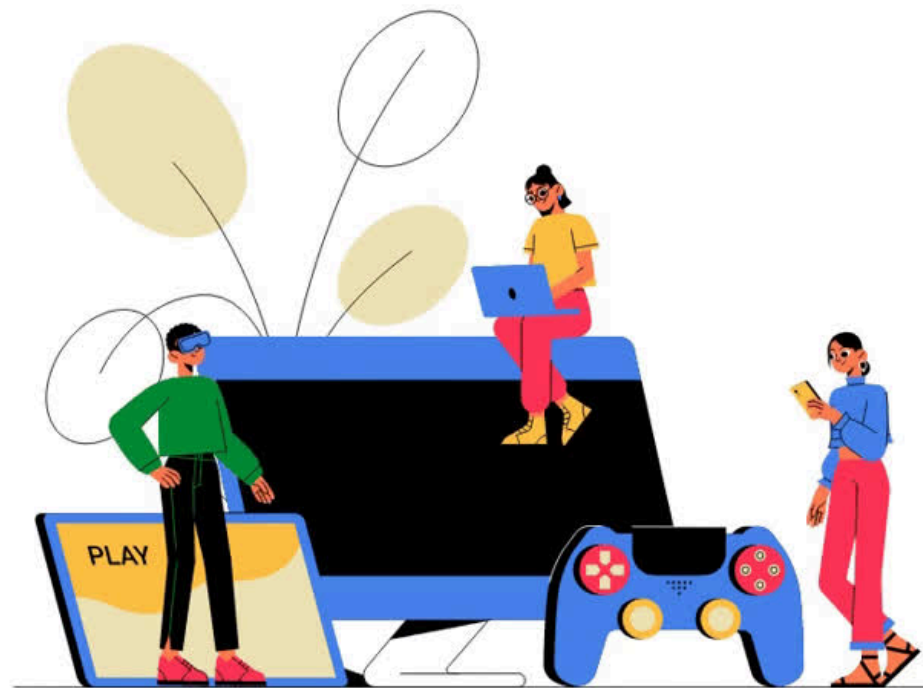
C → Rule-following under pressure

TRY

Build a Mini Stealth Assessment

Pick one course you teach.
Imagine a challenge scenario related to that topic (AI can help)

Give your students 2–3 options
For each option, decide what skill or competency it stealthily reveals



Genially



Thank you for your
attention

Ameny rjiba
ameny.rjiba@gmail.com