2012 Fourth IEEE International Conference On Digital Game And Intelligent Toy Enhanced Learning

# GETOLS: Game Embedded Testing Of Learning Strategies

Gudrun Kellner and Andreas Weißenbacher
Institute of Software Technology and Interactive Systems
Vienna University of Technology
Vienna, Austria
gudrun.kellner@tuwien.ac.at; e0627794@student.tuwien.ac.at

***Abstract*—Learning strategies are typically tested in rather conventional test settings, using questionnaires or observation as a method of information acquisition. However, such conventional test settings are far from the test taker's everyday life situation and might therefore evoke stress. Embedding a learning strategies test into an educational computer game, provides several advantages over the conventional testing procedure: The test takers can stay in their familiar learning environment and even learn something during the process of being tested. We conducted an explorative study by testing 24 pupils, comparing the outcome of a conventional learning strategies test conducted by a psychologist with the results obtained by our GETOLS method. The high similarity of results strongly supports our approach.**

***Keywords—game embedded testing; learning strategies; learning styles; testing***

## I. Introduction

Educational games are being used in different contexts and by different people. Some games aim to teach special learning contents (starting with Chinese history [10] to art education [9]), others intent to train skills [11] or try to trigger the players' reflection on social topics like violence prevention [29] or political efficiency [16]. Teachers are using educational games in schools, children and students play educational games in their spare time, several enterprises as well as the military, use educational games for training purposes. According to Kruse [15], the following specific advantages of e-learning can be identified among others: reduced overall costs, increased retention, on-demand availability, self-pacing, interactivity etc. Furthermore, educational games have been proven to be a source of enjoyment [19] and, if well-made, to evoke a flow feeling within the players [13].

At the same time, taking a psychological test is often perceived as a rather stressful situation [25]. By embedding a psychological test into an educational game, the aim is to combine the advantages of gameplay with the outcome of a learning strategies test. In order to examine this approach, we compared the results achieved via GETOLS method with the ones collected by a psychologist.

This paper answers the following questions: a) Is it possible to embed a psychological test into a game? b) How could a psychological test be adjusted in order to make it feel game-like? c) What are the benefits of game embedded testing?

The paper is organized as follows: Section II searches for existing answers to those questions in related research. Section III explains the methods we used to find our own answers to those questions. Section IV presents the results, which are discussed in section V. Section VI provides a conclusion and an outlook to further research.

## II. Related Research

This section explores related research. First, we give an overview of cognitive styles, especially those that reflect upon sensory preference. Second, we take a look at test theory and psychological testing methods with a focus on personality and/or preferences. Third, we focus on the field of game embedded testing.

### *A. Cognitive style and learning strategies*

Cognitive style "may be defined as an individual's consistent approach to organising and processing information during thinking" [24]. In education, cognitive or learning styles describe different preferences in how learners perceive and retain information. Quite some models have one or more attributes that express perceptual preferences: Several learning style models describe dimensions of information representation that can be classified as perceptual preference for visual input as opposed to verbal input (which might be considered as a person's preference for auditory input if one assumes that content is communicated orally): e.g., the dimension "imager vs. verbalizer" as proposed by *Riding* [6; 24], or the dimension "visual vs. verbal" as proposed by *Felder and Silverman* [8]. *Vester*[28] describes in his model of learning strategies four perceptual preferences, namely visual, auditory, kinesthetic, and text based input. *Dunn and Dunn*'s model includes the "modality preferences" visual, auditory, kinesthetic and tactile preference [6; 7]. An application of this concept to the field of language learning was published by *Reid* [22]. In her model on learning styles she differentiates between four basic perceptual learning channels, visual, auditory, kinesthetic and tactile modalities, and presents a questionnaire focusing on the learner's perceptual preference and his or her preference for individual vs. group learning. According to [7], students achieve significantly better learning results when the students' perceptual preference is matched with educational methods.

978-0-7695-4663-6/12 $26.00 © 2012 IEEE
DOI 10.1109/DIGITEL.2012.10

## B. *How to test learning strategies*

In the field of psychological research, two sorts of variables are tested in order to describe a person: rather permanent and rather transitory attributes. Permanent attributes remain the same over a long period of time and can concern personality, preferences, or fields of interest. Transitory attributes are state-dependent, may change after a short period of time, and represent moods, plans and goals. The choice and number of variables tested in psychological research are strongly dependent on the model the research is built on.

Learning strategies can be categorized as rather permanent attributes. They are developed according to conscious or unconscious interventions by the learner. Furthermore, interventions might be formed by some other external agent [23].Learning strategies might change after a longer period of time or after intense training, but not from one day to the other [22].

Even though learning strategies tests, like most other psychological tests, are traditionally done face to face or via questionnaire, a lot of research has been invested during the last two decades into the topic of shifting the testing procedure to computers. If well implemented, computer based testing can be treated as equivalent to pen-and-paper testing [3].

## C. *Embedded testing*

Testing the learner's learning achievement and learning gains can be done in explicit and implicit manner.

Explicit testing can be evaluated with the help of questionnaires or explicit tasks that need to be fulfilled, and is performed to test the learner's knowledge. Educational games are designed in order to enhance the player's knowledge about a certain topic. Naturally, that knowledge should be tested during or after the game. On the one hand, such testing of knowledge works as repetition and supports learning efficiency; on the other hand it might be used to prove teaching efficiency. [17] [20] provide a skin for explicit knowledge testing within a game.

Implicit testing is a rather new approach in the field of digital learning that analyzes the learner's behavior during the learning process, and rather focuses on the learner's learning approaches and skills. What can be tested and how that can be done is object of investigation in the field of embedded assessment. [27] describe embedded assessment as "the process of measuring knowledge and ability as *part* of a learning activity". Subject of examination are, e.g., problem-solving strategies [12] and higher-order skills [26]. But how is it possible to elicit such preferences? In the field of e-learning, several methods are used to obtain information on the user's learning style: the learner's behavior in web based learning [1; 4] or learning context features [5] can be used as indicators to detect learning styles.

Our approach applies mainly to the first group: Even though the testing phases are well embedded in the storyline and somehow "hidden" in the game, the testing itself is explicit in order to keep the results comparable to the ones collected in the conventional test setting.

# III. METHOD

Before defining the method, we want to sum up the goals of this research:

- to ease the testing process of learning strategies by embedding it into a game
- to choose the learning strategies model most adequate for implementation in a computer game
- to push the testing process into the background of the software in order to make the test takers feel comfortable and not like taking an exam
- to make the game (and hence the testing process) educational and enjoyable

## A. *Testing method*

In order to find the learning strategy test most suitable for application in GETOLS, we analyzed nine learning strategy models. We chose three models for more detailed analysis: *Kolb* [14] presents in his learning style inventory the two axes concrete vs. abstract and reflective vs. active. The four quartiles are classified as four learning styles, namely converging (abstract, active), diverging (concrete, reflective), assimilating (abstract, reflective), and accommodating (concrete, active). *Pask* [18] differentiates in his work two different learning approaches, serialists (partists) and holists (wholists). *Vester* [28] describes in his model of learning strategies four perceptual preferences, namely visual, auditory, kinesthetic, and text based input.

For the classification of the models we address the following five categories of suitability:

- Technical applicability (TA): How complicated is the technical application of the test to a game?
- Game integratability (GI): Is it possible to integrate the theoretical concept into a game?
- Complexity of test evaluation (CTE): How complex is the test evaluation?
- Usefulness of results (UR): Are the results understandable and helpful for the test taker?
- Name recognition (NR): Is the concept well known to potential future users? In order to address a high number of users, the concept should not only be known among researchers, but also among people working in the educational sector.

Every category is assigned one value per model, choosing between the values + (very suitable), ~ (quite suitable), – (not suitable at all).

*Kolb*'s testing method is based on a questionnaire. It is hence very suitable in means of technical application and complexity of test evaluation, but very hard to embed into a game without interrupting the story flow. The results rather inform about abstract strengths than about concrete learning strategies. The model is known to researchers and interested laypersons, but not to the general public.

*Pask*'s testing method is also based on a questionnaire and hence shows the same characteristics as *Kolb* concerning technical applicability, game integratability and complexity of test evaluation. Its results inform about the individual information processing style. The model is known to researchers and interested laypersons, but not to the general public.

*Vester*'s testing method is based on a memory test: for each testing sequence, ten objects are presented in different perceptual modalities (visual, auditory, kinesthetic and text based input) and shall be recalled by the test person after a distraction sequence. The testing is not so easy to realize in means of technical applicability, but it is more suitable concerning the game integratability. The test evaluation is more complex because of the necessity of free word input, which might need special treatment of spelling mistakes and the use of synonyms. The results provide explicit and individual suggestions for improvement on how to learn. The model is very well-known in the didactic sector in German-speaking countries.

An overview of this analysis can be found in Table 1.

TABLE I. SUITABILITY OF DIFFERENT TESTING METHODS FOR GETOLS

| Method | Categories of Suitability | | | | | Sum |
|---|---|---|---|---|---|---|
| | *TA* | *GI* | *CTE* | *UR* | *NR* | *+ = 3*<br>*~ = 1*<br>*– = -3* |
| Kolb | + | – | + | ~ | ~ | 5 |
| Pask | + | – | + | ~ | ~ | 5 |
| Vester | ~ | ~ | ~ | + | + | 9 |

According to this suitability analysis, we opt for the testing method of *Vester*. Even though the technical application is a bit more complicated and needs some further investigation in mapping to digital evaluation, the testing method is easier to integrate into a game. The interpretation of disambiguate input can be solved with more or less work dependent on the degree of automation. Out of the analyzed three models, it is the most popular one and is widely used in the educational sector. The most important point is the usefulness of results, which is assured by the detailed feedback provided to the test taker.

The mapping of object presentation to a digital format is quite easy for visual, auditory and text based modalities, but a challenge regarding the kinesthetic modality. According to [28], not only tactile input, but also the use of an object may be perceived as kinesthetic experience. We therefore produced videos showing the use of objects so as to evoke similar results. For validation of our mapping, we conducted a small study with 20 test persons, performing first the original test with direct kinesthetic input and then our GETOLS video test. For all test persons, the results only differed very little. The results of the study are presented in Table 2.

The scores for the testing method using GETOLS video are slightly higher than the scores for *Vester*'s testing method, which might be explained with a learning effect from the first to the second testing scenario. Still, deviation is of 2 points at most, the average of 0.65 lies beneath one point. We consider the use of videos hence as an adequate mapping of the kinesthetic testing method to a digitally realizable version.

TABLE II. VALIDATION OF GETOLS VIDEO USE FOR KINESTHETIC TESTING

| | Mode of Kinesthetic Testing | | Difference |
|---|---|---|---|
| | *Vester* | *GETOLS Video* | |
| Subject 1 | 6 | 7 | 1 |
| Subject 2 | 8 | 9 | 1 |
| Subject 3 | 5 | 6 | 1 |
| Subject 4 | 6 | 7 | 1 |
| Subject 5 | 4 | 5 | 1 |
| Subject 6 | 5 | 5 | 0 |
| Subject 7 | 7 | 8 | 1 |
| Subject 8 | 3 | 5 | 2 |
| Subject 9 | 3 | 4 | 1 |
| Subject 10 | 5 | 4 | -1 |
| Subject 11 | 4 | 5 | 1 |
| Subject 12 | 7 | 6 | -1 |
| Subject 13 | 8 | 10 | 2 |
| Subject 14 | 6 | 7 | 1 |
| Subject 15 | 5 | 5 | 0 |
| Subject 16 | 3 | 4 | 1 |
| Subject 17 | 4 | 4 | 0 |
| Subject 18 | 6 | 5 | -1 |
| Subject 19 | 7 | 7 | 0 |
| Subject 20 | 6 | 8 | 2 |
| **Average** | **5.40** | **6.05** | **0.65** |

## B. Measures

The outcome of *Vester*'s learning strategies test are four values, pointing out the memory's performance in regard to information presented in four different ways, namely visual (V), auditory (A), kinesthetic (K) and text based (T) input. Each value is in the range of 0 and 10. The test results can be expressed as a four-dimensional vector

$$vresult(t) = [p.V, p.A, p.K, p.T]$$

based on the number of memorized objects per input system.

The vector similarity *vsim(t1,t2)* between two test results was measured as the cosine similarity of the two result vectors, which is defined as

$$\cos(x, y) = \frac{\langle x, y \rangle}{||x|| * ||y||}$$

The value cos(x,y) = 1 indicates complete similarity, i.e. y=ax ε R**.** The value cos(x,y)=0 expresses orthogonality between the vectors x and y. [2]

## C. Hypotheses

We are testing the following three hypotheses:

**Hypothesis A** states that *Vester*'s learning strategies test can also be performed digitally via the GETOLS method. This is the case if hypothesis B can be verified.

**Hypothesis B** investigates the similarity of the two testing methods. According to [3], scores from conventional and computer based testings can be considered as equivalent when the order of the scores is approximately the same and the scores are, or have been made approximately the same by scale adjustment. Hypothesis B.1 states hence that the

score ranking for the four tested perceptual modalities is the same for GETOLS and for conventional testing.
Still, we think that the results achieved via GETOLS might as well be similar concerning scores (Hypothesis B.2a) or be even higher concerning scores (Hypothesis B.2b).

**Hypothesis C** states that testing via GETOLS is more time efficient than the conventional testing.

## IV. The GETOLS Test Setting

We decided to do our testing with 12-13 years old pupils. At that age, adolescents show high interest and engagement when interacting with computers [30], and are old enough to provide useful feedback on the game and the test design. To ensure that all test persons are experienced in the handling of computers, we chose a school type with computer science classes. From a very early stage in game development on, we chose a specific teacher as contact teacher, which allowed us to get feedback from a field expert.

### *A. The Didactic Adventure Game "Save the city"*

Following a suggestion of our contact teacher, we decided to focus on environmental protection as a subtopic of the subject matters "city and ecology", which is taught in biology at 7th grade. The game is about a young boy who is told that he had been chosen to save the city. In order to do so, he needs to complete several missions.

We chose the genre of adventure games as a test setting. Adventure games are intuitive and easy to play. Their focus on the storyline helps to build up suspense and game engagement and to keep up the player's concentration. The contents are prepared according to [28]. The story is made as suspenseful and fascinating as possible so as to enhance learning motivation. We pay attention to address all learning types during the game by not only using visual and kinesthetic stimulation (which are part of adventure games), but also by addressing the auditory system via complete speech recording and providing textual input via subtitles.

The game can be played in about half an hour, which is short enough to be performed during one school lesson. The gameplay evokes a flow feeling that permits to keep up concentration for the duration of the testing. [13]

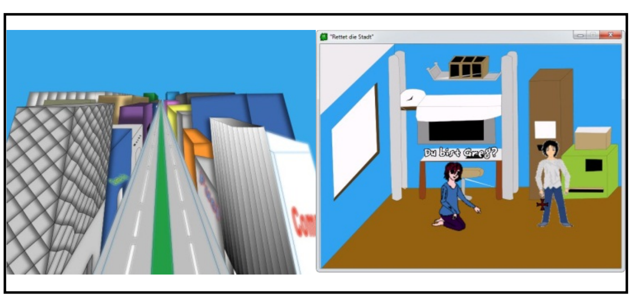

Figure 1. Screenshots of the educational adventure game "Save the city".

### *B. Integrating the Testing Method into the Game*

Therefore it is necessary to not only loosely combine the test and the game, but to really embed the testing sequences into the game. The testing sequences as described in [28] always follow the same test procedures:

- Memorization phase: 10 objects are presented according to the perceptual modality that shall be tested
- Different brain activity: mental arithmetic
- Test phase: test the memorized objects

Due to organizational issues, it is necessary to define a surrounding phase:

- Relaxation: wait until everything is prepared for the next testing phase; this part applies before and after each testing loop

Those patterns are mapped to the game as follows:

The story itself shall enhance engagement and help to focus on the game; furthermore it shall allow relaxation between the testing phases. In order to raise the level of attention, every memorization phase is preceded by an unexpected event in the story.

During the memorization phase, 10 objects are presented according to the perceptual modality that shall be tested: via pictures (visual), dictation (auditory), videos of use (kinesthetic) and as words (textual). Each object is shown for a period of 2 seconds.

The phase of different brain activity uses an adapted concept. According to [28], the goal of this phase is productive distraction. Instead of 30 seconds of mental arithmetic, we implemented four different minigames, each lasting 30 seconds. The tasks of the minigames are to identify differences between two pictures, to reassemble the pieces of a diamond, to find the way through a labyrinth and to click several times at a moving object. While mental arithmetic might resemble an examination situation and hence cause stress, our minigames are designed to be a fun activity of distraction.

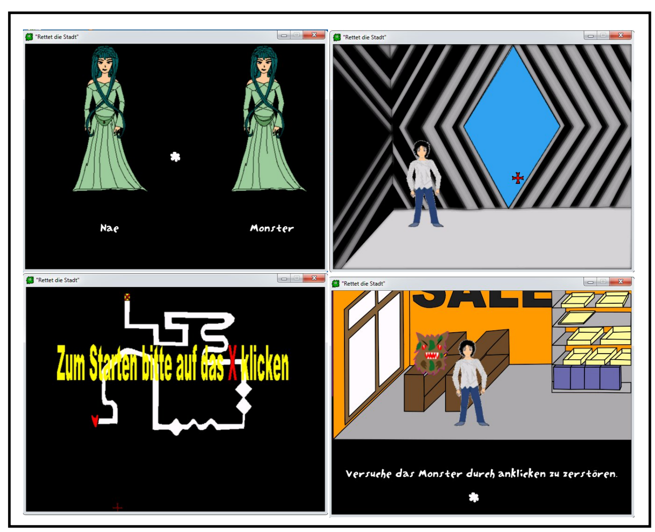


Figure 2. Screenshots of the four GETOLS minigames.

Furthermore, the minigames are part of the gameplay and add another element to the storyline. Figure 2 shows screenshots of the minigames.

In the testing phase, the player is asked to recall as many of the 10 objects presented in the memorization phase as possible and to list by keyboard input.

The testing procedure can hence be depicted as shown in Figure 3.

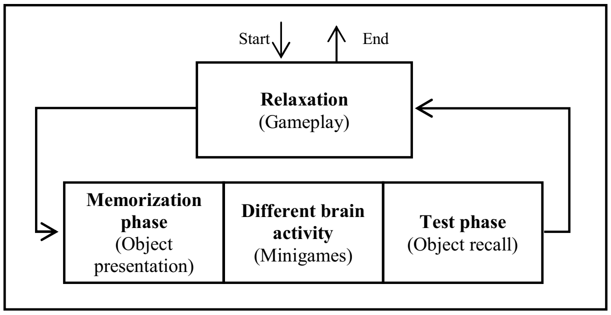


Figure 3. The GETOLS Testing Procedure.

### C. *Implementation and Testing*

"Save the city" was implemented in SLUDGE, an object oriented programming language designed for the programming of adventure games.[1] Animated graphics were produced with the help of Anime Studio Pro.[2]

The test results for each test phase are stored as a list of words and are, at this stage of development, evaluated manually. An automated evaluation can be realized by integration of spell checking and synonym treatment, but might still need manual verification in order to assure correct evaluation of results.

The alpha testing phase was started after implementation of basic functionality. 28 people of different age and with different expertise in the manipulation of computers were asked to play and comment the game. The so obtained feedback led to more explicit information about the time limits during the testing sequence. Furthermore, we reduced the resolution of the videos in order to ensure functionality also on less performant computers.

The first loop of beta testing was already conducted with our final test subjects. Each subject got a very short sequence from the game for testing. In order to avoid measurement bias, the memorization phases were executed with different objects than in the final version. As several test subjects had problems with the use of the inventory during the last minigame, we redesigned the minigame. The second loop of beta testing was conducted under the same conditions as the first one.

In a last beta testing phase, 9 computer scientists tested the game with regard to functionality and usability.

[1]http://opensludge.sourceforge.net/
[2]http://anime.smithmicro.com/

## V. EXPERIMENTAL RESULTS

We opted for repeated measure design. One set of data was recorded via GETOLS method, the second set was collected in a conventional test setting, where a psychologist conducted the learning strategy test. We had 24 test subjects, whereof 18 test subjects were recorded via GETOLS and 16 tested by a psychologist. The conventional test was conducted one week before the GETOLS screening. We chose within-subjects analysis for the comparison of results. As only 10 test persons were recorded in both tests, we base our analysis upon the results of those 10 test subjects.

All test subjects were 12 or 13 years of age.

The following factors were considered when setting up the test:

- Location: The location chosen for the testing was the computer labs of the school. In addition to organizational advantage, the place is known to the subjects, which reduces fear, and can be closed for the public in order to avoid distraction.
- Preparation: Test preparation is finished before the test subject enters the room.
- Test procedure: Each test subject is welcomed individually in order to reduce tension.
- Duration: Even though average testing only takes 45 minutes, testing is scheduled for two hours per person in order to avoid stress.

The results of the testing are shown in Table 3-5.

## VI. DISCUSSION

Looking at the test results, our hypotheses leads to the following conclusions:

The GETOLS test method can be considered equal to the conventional test setting. As can be seen in Table 3, the order of the scores is similar for the two test methods. Similarity was calculated as cosine similarity as described in Section III.B. All results show a cosine similarity greater than 0.8, and 70% have a similarity greater than 0.9. Hypothesis B.1 can hence be verified. Concerning the scores, the similarity of results based on scores is even higher than based on order. The lowest vector similarity of scores is 0.88, and 90% have a similarity greater than 0.9 (see Table 4). Hypothesis B.2a can hence be verified as well. We could yet not prove hypothesis B.2b, which states that test results via GETOLS are higher; this is only the case for 75% of the results (see Table 5), which is not high enough for significant conclusions. A slight scale adjustment might hence be taken into consideration when using the GETOLS method. However, in *Vester*'s learning strategies test, the perceptual modalities' order is more informative than the absolute scores, which makes hypothesis B.1 the most important.

These results also support Hypothesis A: *Vester*'s learning strategies test can thus be performed digitally via the GETOLS method.

TABLE III. COMPARISON OF METHODS: RESULTS BYORDER

<table>
<tr><th></th><th>Method</th><th colspan="4">Perceptual Modality</th><th colspan="3">Similarity: Subjects</th></tr>
<tr><th></th><th>g = GETOLS<br>p = Psychologist</th><th>Text based</th><th>Auditory</th><th>Visual</th><th>Kinesthetic</th><th><g,p></th><th>||g|| * ||p||</th><th>cos(g,p)</th></tr>
<tr><td rowspan="2">Subject 1</td><td>g</td><td>3</td><td>4</td><td>1</td><td>2</td><td rowspan="2">27</td><td rowspan="2">30.000</td><td rowspan="2">0.900</td></tr>
<tr><td>p</td><td>4</td><td>2</td><td>1</td><td>3</td></tr>
<tr><td rowspan="2">Subject 2</td><td>g</td><td>3</td><td>4</td><td>1</td><td>2</td><td rowspan="2">24</td><td rowspan="2">29.749</td><td rowspan="2">0.807</td></tr>
<tr><td>p</td><td>3.5</td><td>2</td><td>3.5</td><td>1</td></tr>
<tr><td rowspan="2">Subject 3</td><td>g</td><td>3</td><td>4</td><td>2</td><td>1</td><td rowspan="2">29.5</td><td rowspan="2">29.749</td><td rowspan="2">0.992</td></tr>
<tr><td>p</td><td>2.5</td><td>4</td><td>2.5</td><td>1</td></tr>
<tr><td rowspan="2">Subject 4</td><td>g</td><td>1</td><td>4</td><td>1</td><td>3</td><td rowspan="2">28</td><td rowspan="2">28.222</td><td rowspan="2">0.992</td></tr>
<tr><td>p</td><td>1.5</td><td>4</td><td>1.5</td><td>3</td></tr>
<tr><td rowspan="2">Subject 5</td><td>g</td><td>1.5</td><td>4</td><td>1.5</td><td>3</td><td rowspan="2">25.5</td><td rowspan="2">29.749</td><td rowspan="2">0.857</td></tr>
<tr><td>p</td><td>4</td><td>3</td><td>1</td><td>2</td></tr>
<tr><td rowspan="2">Subject 6</td><td>g</td><td>2.5</td><td>4</td><td>2.5</td><td>1</td><td rowspan="2">29.5</td><td rowspan="2">29.500</td><td rowspan="2">1.000</td></tr>
<tr><td>p</td><td>2.5</td><td>4</td><td>2.5</td><td>1</td></tr>
<tr><td rowspan="2">Subject 7</td><td>g</td><td>3</td><td>4</td><td>2</td><td>1</td><td rowspan="2">28</td><td rowspan="2">28.983</td><td rowspan="2">0.966</td></tr>
<tr><td>p</td><td>2</td><td>4</td><td>2</td><td>2</td></tr>
<tr><td rowspan="2">Subject 8</td><td>g</td><td>3</td><td>4</td><td>1</td><td>2</td><td rowspan="2">30</td><td rowspan="2">30.000</td><td rowspan="2">1.000</td></tr>
<tr><td>p</td><td>3</td><td>4</td><td>1</td><td>2</td></tr>
<tr><td rowspan="2">Subject 9</td><td>g</td><td>4</td><td>3</td><td>1.5</td><td>1.5</td><td rowspan="2">24</td><td rowspan="2">28.740</td><td rowspan="2">0.835</td></tr>
<tr><td>p</td><td>3</td><td>1</td><td>3</td><td>3</td></tr>
<tr><td rowspan="2">Subject 10</td><td>g</td><td>3</td><td>4</td><td>2</td><td>1</td><td rowspan="2">29.5</td><td rowspan="2">29.749</td><td rowspan="2">0.992</td></tr>
<tr><td>p</td><td>2.5</td><td>4</td><td>2.5</td><td>1</td></tr>
<tr><td rowspan="3">Similarity:<br>Perceptual Modality</td><td><g,p></td><td>76.25</td><td>119</td><td>32.75</td><td>37.5</td></tr>
<tr><td>||g|| * ||p||</td><td>83.285</td><td>132.068</td><td>36.297</td><td>39.481</td></tr>
<tr><td>cos(g,p)</td><td>0.916</td><td>0.901</td><td>0.902</td><td>0.950</td></tr>
</table>

TABLE IV. COMPARISON OF METHODS: RESULTS BYSCORES

<table>
<tr><th></th><th>Method</th><th colspan="4">Perceptual Modality</th><th colspan="3">Similarity: Subjects</th></tr>
<tr><th></th><th>g = GETOLS<br>p = Psychologist</th><th>Text based</th><th>Auditory</th><th>Visual</th><th>Kinesthetic</th><th><g,p></th><th>||g|| * ||p||</th><th>cos(g,p)</th></tr>
<tr><td rowspan="2">Subject 1</td><td>g</td><td>4</td><td>3</td><td>9</td><td>6</td><td rowspan="2">148</td><td rowspan="2">155.82683</td><td rowspan="2">0.94977</td></tr>
<tr><td>p</td><td>4</td><td>7</td><td>9</td><td>5</td></tr>
<tr><td rowspan="2">Subject 2</td><td>g</td><td>5</td><td>3</td><td>7</td><td>6</td><td rowspan="2">99</td><td rowspan="2">105.19981</td><td rowspan="2">0.94107</td></tr>
<tr><td>p</td><td>4</td><td>5</td><td>4</td><td>6</td></tr>
<tr><td rowspan="2">Subject 3</td><td>g</td><td>4</td><td>2</td><td>5</td><td>6</td><td rowspan="2">87</td><td rowspan="2">87.72115</td><td rowspan="2">0.99178</td></tr>
<tr><td>p</td><td>5</td><td>3</td><td>5</td><td>6</td></tr>
<tr><td rowspan="2">Subject 4</td><td>g</td><td>8</td><td>6</td><td>8</td><td>7</td><td rowspan="2">207</td><td rowspan="2">207.42710</td><td rowspan="2">0.99794</td></tr>
<tr><td>p</td><td>8</td><td>5</td><td>8</td><td>7</td></tr>
<tr><td rowspan="2">Subject 5</td><td>g</td><td>7</td><td>2</td><td>7</td><td>4</td><td rowspan="2">104</td><td rowspan="2">118.49895</td><td rowspan="2">0.87764</td></tr>
<tr><td>p</td><td>3</td><td>5</td><td>7</td><td>6</td></tr>
<tr><td rowspan="2">Subject 6</td><td>g</td><td>5</td><td>4</td><td>5</td><td>7</td><td rowspan="2">111</td><td rowspan="2">111.44505</td><td rowspan="2">0.99601</td></tr>
<tr><td>p</td><td>5</td><td>3</td><td>5</td><td>7</td></tr>
<tr><td rowspan="2">Subject 7</td><td>g</td><td>5</td><td>3</td><td>7</td><td>8</td><td rowspan="2">109</td><td rowspan="2">111.12156</td><td rowspan="2">0.98091</td></tr>
<tr><td>p</td><td>5</td><td>3</td><td>5</td><td>5</td></tr>
<tr><td rowspan="2">Subject 8</td><td>g</td><td>5</td><td>4</td><td>8</td><td>6</td><td rowspan="2">110</td><td rowspan="2">110.11812</td><td rowspan="2">0.99893</td></tr>
<tr><td>p</td><td>4</td><td>3</td><td>6</td><td>5</td></tr>
<tr><td rowspan="2">Subject 9</td><td>g</td><td>3</td><td>6</td><td>8</td><td>8</td><td rowspan="2">131</td><td rowspan="2">138.57489</td><td rowspan="2">0.94534</td></tr>
<tr><td>p</td><td>5</td><td>6</td><td>5</td><td>5</td></tr>
<tr><td rowspan="2">Subject 10</td><td>g</td><td>4</td><td>2</td><td>6</td><td>7</td><td rowspan="2">112</td><td rowspan="2">113.64418</td><td rowspan="2">0.98553</td></tr>
<tr><td>p</td><td>5</td><td>3</td><td>5</td><td>8</td></tr>
<tr><td rowspan="3">Similarity:<br>Perceptual Modality</td><td><g,p></td><td>235</td><td>156</td><td>410</td><td>366</td></tr>
<tr><td>||g|| * ||p||</td><td>257.721</td><td>171.216</td><td>433.274</td><td>401.186</td></tr>
<tr><td>cos(g,p)</td><td>0.912</td><td>0.911</td><td>0.946</td><td>0.912</td></tr>
</table>

TABLE V. COMPARISON OF METHODS: SCORE DIFFERENCES

| | Method | Perceptual Modality | | | | | Differences: Subject | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | g = GETOLS<br>p = Psychologist | ***Text based*** | ***Auditory*** | ***Visual*** | ***Kinesthetic*** | ⍰ | ***T*** | ***A*** | ***V*** | ***K*** |
| Subject 1 | *g* | 4 | 3 | 9 | 6 | 5.5 | **0** | -4 | **0** | **1** |
| | *p* | 4 | 7 | 9 | 5 | 6.25 | | | | |
| Subject 2 | *g* | 5 | 3 | 7 | 6 | 5.25 | **1** | -2 | **3** | **0** |
| | *p* | 4 | 5 | 4 | 6 | 4.75 | | | | |
| Subject 3 | *g* | 4 | 2 | 5 | 6 | 4.25 | -1 | -1 | **0** | **0** |
| | *p* | 5 | 3 | 5 | 6 | 4.75 | | | | |
| Subject 4 | *g* | 8 | 6 | 8 | 7 | 7.25 | **0** | **1** | **0** | **0** |
| | *p* | 8 | 5 | 8 | 7 | 7 | | | | |
| Subject 5 | *g* | 7 | 2 | 7 | 4 | 5 | **4** | -3 | **0** | -2 |
| | *p* | 3 | 5 | 7 | 6 | 5.25 | | | | |
| Subject 6 | *g* | 5 | 4 | 5 | 7 | 5.25 | **0** | **1** | **0** | **0** |
| | *p* | 5 | 3 | 5 | 7 | 5 | | | | |
| Subject 7 | *g* | 5 | 3 | 7 | 8 | 5.75 | **0** | **0** | **2** | **3** |
| | *p* | 5 | 3 | 5 | 5 | 4.5 | | | | |
| Subject 8 | *g* | 5 | 4 | 8 | 6 | 5.75 | **1** | **1** | **2** | **1** |
| | *p* | 4 | 3 | 6 | 5 | 4.5 | | | | |
| Subject 9 | *g* | 3 | 6 | 8 | 8 | 6.25 | -2 | **0** | **3** | **3** |
| | *p* | 5 | 6 | 5 | 5 | 5.25 | | | | |
| Subject 10 | *g* | 4 | 2 | 6 | 7 | 4.75 | -1 | -1 | **1** | -1 |
| | *p* | 5 | 3 | 5 | 8 | 5.25 | | | | |
| Average | ***g*** | 5 | 3.5 | 7 | 6.5 | 5.5 | 0.2 | -0.8 | 1.1 | 0.5 |
| | ***p*** | 4.8 | 4.3 | 5.9 | 6 | 5.25 | | | | |

Furthermore, the GETOLS test method is more time efficient than the conventional testing approach: While the conventional testing needs a human test supervisor and can only be executed in a serial manner, the GETOLS method allows parallel testing, which shows time efficiency effects starting from a test group of more than two test takers. In our experiment, testing 16 persons with the conventional testing method took 130 minutes, whereas testing via the game "Save the city" took 45 minutes for 18 test persons and would not take longer even for a bigger group (of course, the group size is limited according to the number of computers available).

As suggested by [28], the phase of different brain activity was accomplished with mental arithmetic exercises in the conventional test setting. According to the teacher, that phase was experienced by most of the test subjects as an examination situation and hence as quite stressful. This might be a reason for the lower test scores in the conventional test setting. Oral feedback by the test subjects confirms that assumption, which corresponds to the findings of [25]. Furthermore, several test subjects pointed out that playing the minigames was much more fun. We plan to test those aspects in a second experimental run.

*Limitations:* So far, all our studies were conducted with a low number of test persons. Therefore, our results cannot yet provide a significant proof of our method, but rather offer a first insight. At the moment, our GETOLS method has only been realized with one testing method, namely *Vester*'s learning strategies test. Nevertheless, other test methods could be implemented by the same principle, most likely accompanied by the need of adjusting the storyline of the game to the testing method.

On the infrastructural side, a computer is needed for conducting the test. However, computers are easily available and are thus a good substitute for the conventional testing material that needs to be assembled and prepared first in order to conduct a test in the conventional way.

## VII. CONCLUSIONS AND OUTLOOK

In this paper, we proposed to embed a well-known learning style test into a rather unconventional test setting by integrating it into an educational computer game. We opted for *Vester*'s learning strategies test as the most suitable for such an embedded testing approach.

Our empirical validation shows a high similarity of results obtained with the two different test methods.

According to the findings of this study, the questions we raised in the introduction are answered as follows: (a) As the testing of hypotheses A and B shows, it is possible to embed a psychological test in a game. Results are similar in terms of scores and score order for both testing methods. (b) We chose the testing method of *Vester* after analyzing several learning styles testing methods by the criteria of technical applicability, game integratability, complexity of test evaluation, usefulness of results, and name recognition. (c) The benefits of game embedded testing are certainly a more time efficient testing. Furthermore, the feedback of our test subjects indicates a higher fun factor and less stress. We are planning to test those effects in a second test run.

The obtained results are quite encouraging: Our next steps are to repeat the study with a bigger number of test samples, to support automated test evaluation, to monitor reactions of stress and engagement for both testing methods, and to investigate the relation of test results and in-game behavior.

## ACKNOWLEDGEMENTS

Thanks a lot to Thomas Pölzl for conducting the psychological tests and to Wolfgang Lechner for his support as our contact teacher. Funding for this research was partly provided by the fFORTE WIT – Women in Technology Program of the Vienna University of Technology.

## REFERENCES

[1] N. Atman et al: Learning Styles Diagnosis Based on Learner Behaviors in Web Based Learning. In: Gervasi, O. et al. (Eds.): ICCSA 2009, Part II, pp. 900–909, 2009.

[2] R.Y. Bayardo, Y. Ma and R. Srikant, "Scaling Up All Pairs Similarity Search", Proceedings of WWW 2007, ACM, New York, 2007.

[3] A.C. Bugbee, "The Equivalence of Paper-and-Pencil and Computer-Based Testing," In: Journal of Research on Computing in Education 28(3): 282–299, 1996.

[4] Y. Chang, W. Kao, C. Chu and C. Chiu,"A learning style classification mechanism for e-learning," In: Computers & Education 53, pp. 273–285, 2009.

[5] M. Chi, K. VanLehn, D. Litman and P. Jordan: Inducing Effective Pedagogical Strategies Using Learning Context Features. In: De Bra, P., Kobsa, A., Chin, D. (Eds.): UMAP 2010. Springer, Berlin; pp. 147–158, 2010.

[6] F. Coffield, D. Moseley, E. Hall and K. Ecclestone: Learning styles and pedagogy in post-16 learning. A systematic and critical review. LSRC, London , 2004.

[7] R. Dunn: Commentary: Teaching Students Through Their Perceptual Strengths or Preferences. In: Journal of Reading 31/4, pp. 304–309, 1988.

[8] R. Felder and L. Silverman: Learning and Teaching Styles In Engineering Education. In: Engr. Education, 78/7, pp. 674–68, 1988.

[9] J. Froschauer, M. Arends, D. Goldfarb and D. Merkl, "Towards an Online Multiplayer Serious Game Providing a Joyful Experience in Learning Art History", In Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES 2011), 2011, pp. 160–163.

[10] C. Hui Chung Yan and F. Tam, "Learning Chinese History through Digital Game," DIGITEL, 2010 Third IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, 2010, pp.156–160.

[11] E. Ju and C. Wagner, "Personal computer adventure games: their structure, principles, and applicability for training," SIGMIS Database, vol. 28, 1997, pp. 78–92.

[12] D. J. Ketelhut, B. Nelson and C. Schifter, „Situated assessment using virtual environments of science content and inquiry", In Proceedings of AERA, 2009.

[13] K. Kiili, "On educational game design. Building blocks of flow experience," PhD Thesis, Tampere Universityo f Technology, Finland, 2005.

[14] D.A. Kolb, "Experiential learning: Experience as the source of learning and development", New Jersey, Prentice-Hall, 1984.

[15] K. Kruse. (2002, 10.09.2011). The Benefits and Drawbacks of e-Learning. Available: http://labsel.pesarosviluppo.it/docindexer/Uploads%5C213-Beginner%20Basics.doc

[16] H. Ling Yi, "Designing Serious Games to Enhance Political Efficacy and Critical Thinking Disposition for College Students: The Case of Taiwan," VS-GAMES, 2011 Third International Conference on Games and Virtual Worlds for Serious Applications, 2011, pp.148–151.

[17] P.E. Morris and C.O. Fritz, "The Name Game: Using Retrieval Practice to Improve the Learning of Names," Journal of Experimental Pychology 6(2):124–129, 2000.

[18] G. Pask, "Styles and strategies of learning," In The British Journal of Educational Psychology 46: 128 –148, 1976.

[19] K. Poels, Y. de Kort and W. Ijsselsteijn, ""It is always a lot of fun!": exploring dimensions of digital game experience using focus group methodology," Proceedings of the 2007 conference on Future Play, ACM, New York, 2007.

[20] S. Ramani, V. Sirigiri, N. L. Panigrahi and S. Sabharwal, "Games as Skins for Online Tests," DIGITEL, 2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, 2008, pp.90–92.

[21] K. Rapeepisarn, K. Wong, C. Fung, and M. Khine, "The Relationship between Game Genres, Learning Techniques and Learning Styles in Educational Computer Games," in Technologies for E-Learning and Digital Entertainment. vol. 5093, Z. Pan, et al., Eds., ed: Springer, Berlin, 2008, pp. 497–508.

[22] J.M. Reid,"The Learning Style Preferences of ESL Students". In: TESOL Quarterly, 21/1, pp. 87–111, 1987.

[23] Robotham, (1999, 10.09.2011) "The application of learning style theory in higher education teaching,", available: http://serprofessoruniversitario.pro.br/m%C3%B3dulos/c%C3%A9rebro-e-aprendizagem/application-learning-style-theory-higher-education-teaching-0

[24] E. Sadler-Smith and R. Riding: Cognitive style and instructional preferences. In: Instructional Science 27, pp. 355–371, 1999

[25] I.G. Sarason: "Stress, anxiety, and cognitive interference: reactions to tests", Journal of Personality and Social Psychology 46(4):929–938, 1984.

[26] V. J.Shute, M.Ventura, M. I. Bauer and D. Zapata-Rivera, "Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow," in U. Ritterfeld, M.J. Cody and P. Vorderer, (Eds.), The Social Science of Serious Games: Theories and Applications. Philadelphia, PA: Routledge/LEA, 2008.

[27] J. S. Underwood, S. Kruse, and P. Jakl, "Moving to the Next Level: Designing Embedded Assessments into Educational Games," in Design and Implementation of Educational Games: Theoretical and Practical Perspectives, P. Zemliansky and D. Wilcox, Eds., ed Hershey, New York: IGI Global, 2010, pp. 126–140.

[28] F. Vester, "Denken Lernen Vergessen. Was geht in unserem Kopf vor, wie lernt das Gehirn, und wann lässt es uns im Stich?," dtv, München, 2011.

[29] O. Wolters, "Elektronische Spiele: Wachstumsmarkt mit großer Wertschöpfung," in Spielend die Zukunft gewinnen. A. Picot, S. Zahedani and A. Ziemer, Eds., ed: Springer, Berlin, 2008,pp.25–36.

[30] „KIM-Studie 2010: Basisstudie zum Medienumgang 6- bis 13-Jähriger in Deutschland," Medienpädagogischer Forschungsverband Südwest (ed.), 2011.