Discovery simulations and the assessment of intuitive knowledge

J. Swaak & T. de Jong
Telematica Instituut and University of Twente, Enschede

Abstract The objective of the present work is to have a closer look at the relations between the features of discovery simulations, the learning processes elicited, the knowledge that results, and the methods used to measure this acquired knowledge. It is argued that discovery simulations are ‘rich’, have a relatively low transparency, and require active involvement of learners. Discovery simulations are suited to support data-driven, partly implicit learning. Discovery learning leads to intuitive knowledge. To complement this conceptual investigation, a series of five experimental studies is described. In all five studies, learners were pre-tested and post-tested with several knowledge measures. Central to the set of tests was one with the objective of measuring intuitive knowledge. One conclusion of these experimental studies is that assignments contribute most clearly to the instructional effectiveness of simulations. Another conclusion is that the intuitive knowledge tests seem able to measure the results of learning with discovery simulations.

Keywords: Assessment; Discovery learning; Explicit knowledge; Intuitive knowledge; Physics; Post-secondary; Pre-/post-test; Simulation

Introduction

This paper is about learning with discovery simulations. More precisely, the current work investigates the instructional effectiveness of discovery simulations and examines what learners gain and do not gain from interacting with discovery environments.

Learning in discovery environments, such as simulation environments, is supposed to lead to the acquisition of knowledge that is qualitatively different from knowledge that is acquired in more traditional instructional situations. In traditional instruction, learning usually involves explicit transfer of a body of knowledge, with an emphasis on the analytical aspects of the domain and on the reproduction of the knowledge taught. The learning process invoked by discovery environments may be less explicit, and the resulting knowledge may have a less explicit character. Thomas & Hooper (1991), for example, classified and analysed simulation studies. The conclusion most relevant for the present work is that: ‘the effects of simulations are not revealed by tests of knowledge’ (p. 479). They further conclude that simulations can best serve as ‘experiencing programs’ which give students the opportunity ‘to
Thomas and Hooper do not, however, indicate what they mean by ‘tests of knowledge’ and ‘experiencing programs’, why and how they think an intuitive understanding is gained, and how the intuitive understanding can be assessed.

The objective of the present work is to have a closer look at the relations between the features of discovery simulations, the learning processes elicited, the knowledge that results, and the methods used to measure this acquired knowledge. A series of five experimental studies was carried out. In all five studies, learners were pretrained and post-tested with several knowledge measures. Central to the set of tests was one with the objective of measuring intuitive knowledge.

Features of discovery simulations

In the current studies, discovery simulations were developed with the SMISLE and SimQuest authoring systems (see de Jong et al., 1998). The core of the instructional environments in the current work always entails a simulation of a physics topic. The goal for the learners is to discover relationships between variables of the simulated domain. The physics topics are: collisions between two particles, harmonic oscillations (mass-spring systems), and electrical circuits. In the simulations, learners can manipulate variables and perceive the consequences of their manipulations in dynamic outputs. Furthermore, the simulation environment of the current work is enriched with instructional measures including explanations, assignments, and model progression. ‘Explanations’ (in the studies) give a description of a variable in the domain or give an equation used in the domain. ‘Assignments’ are small exercises that aim to help students to structure their learning, and point them to important phenomena in the domain. For example, some assignments ask the student to investigate the relationship between two variables of the domain by manipulating them in the simulation. Other assignments require students to predict a situation based on the relation between two variables in the domain. ‘Model progression’ means that the domain is offered in small sequential steps, where, in each step, new variables are added to the model. This instructional measure also offers structure for the learning process (see White & Frederiksen, 1990).

Apart from (physics) topics and instructional support that may be specific for the simulations of the current work, generic features of simulations can be discerned. This work maintains that discovery simulations have three general characteristics.

Richness: First, it is postulated that these types of learning environments can be described as ‘rich’ environments. In a rich environment:

- a large amount of information can be extracted by the learner;
- this information can be obtained in several ways: the information is usually displayed in more than one representation; a dynamic, graphical representation of the output is generally present next to animations and numerical outputs. More specifically, this latter component can be described as ‘perceptual richness’.

Low transparency: A second characteristic of simulation-based environments

---

1The SMISLE System was an authoring environment for developing Multimedia Integrated Simulation Learning Environments. The follow-up of the SMISLE environment was the SimQuest authoring environment, which was developed in the SERVIVE project. The SMISLE and SERVIVE projects were partly sponsored by the EC (DG XIII) in its Telematics programme. In 2000 SimQuest received the ‘European Academic Software Award’.
concerns the relatively low transparency of the learning environment (as compared to text books or hypertexts, etc.). The less transparent the discovery environment, the less the learner has a ‘direct view’ of the variables and relations, and consequently the more information is to be inferred or extrapolated.

**Active interaction:** The third characteristic of the learning context of this work involves the interactive aspect of the discovery simulations. The learning session entails an interaction with a simulation-based environment. Learners are not supposed to passively absorb information on the domain from the computer screen, but rather they are expected to perform several different actions (i.e. do experiments) to make up their own ‘meaningful’ learning session.

**Processes of discovery learning and intuitive knowledge**

Research into discovery learning is not new. A number of researchers have studied discovery learning processes in the context of computer simulations (e.g. Friedler *et al.* 1990; Schauble *et al.*, 1991; Glaser *et al.*, 1992; see de Jong & van Joolingen, 1998; for an overview), and substantial research is available on scientific reasoning and discovery skills in general (e.g. de Groot, 1961; Lawson, 1985). One conclusion from these studies was that two approaches to discovery learning could be distinguished: a top-down method, or concept-driven way, and a bottom-up approach or data-driven way of discovery. In the concept-driven approach prior knowledge plays a central role. In the data-driven approach features of the environment (e.g. a simulation interface) are of central importance.

In this context, Norman (1993) reasons that some environments lead one toward an ‘experiential’ mode of learning while in other contexts learning can be characterised as ‘reflective.’ The experiential mode of learning is ‘one of perceptual processing ... pattern-driven or event-driven’ (p. 26). The reflective mode of learning is ‘that of concepts, of planning and reconsideration’ (p. 25). Norman argues that ‘rich, dynamic ... environments ... lead one toward the experiential learning mode’ (p. 25). More specifically, Rieber (1996) explains that external representations like animations, may trigger implicit learning.

The learning modes as described above also seem to relate to the transparency of the environment with which learners perform a task. Reber *et al.* (1980) argue that for low salient (i.e. aspects of the problem are hidden) complex tasks, an implicit learning mode is best, while high salient, relatively simple tasks are best performed in an explicit way.

In this work it is argued that the rich, low transparent, interactive simulation environment is suited to support the data-driven, partly implicit processes of discovery. In this context, the data-driven processes can more specifically be described as action-driven and perception-driven. Discovery is always a combination of concept-driven with action-driven and perception-driven processes. It is argued that especially the action-and perception-driven elements in discovery, which are partly implicit, lead to intuitive knowledge.

Despite the under-representation of serious efforts to assess intuitive knowledge, research on interacting with complex simulation systems (e.g. Broadbent *et al.*, 1986; Berry & Broadbent, 1988; Hayes & Broadbent, 1988; Leutner, 1993), complemented by literature on intuitive knowledge (e.g. Polanyi, 1966; Westcott, 1968; de Groot, 1986; Fischbein, 1987), sketches at least five basically stable notions on the intuitive quality of knowledge.

The first is that the intuitive quality of knowledge is only acquired after using knowledge in perceptually rich, dynamic situations. It is postulated that if knowledge is used in rich contexts, implicit learning processes are elicited which lead to intuitive knowledge. This idea is in agreement with Fischbein’s perspective on the acquisition of intuitions. He states that they ‘can never be produced by mere verbal learning . . . but that they only can be attained as an effect of direct, experiential involvement of the subject in a practical or mental activity’ (Fischbein, 1987; p. 95).

A second finding is that intuitive knowledge is difficult to verbalise. A rather important hypothesis is that in the interaction with a simulation environment learners are invited to follow an implicit learning mode which leads to knowledge that is hard to verbalise (for related opinions on implicit learning and resulting knowledge see Berry & Broadbent, 1988; Reber, 1993).

A third feature of intuitive knowledge entails the importance of perception. Though visualisation is critical, Fischbein states that an intuition is not just a perception, but more like a theory. He calls an intuition ‘the analog of perception at the symbolic level.’ According to Fischbein the visualisation may or may not be mediated by an external representation. Fischbein, moreover, recalls that ‘what one cannot imagine visually is difficult to realise mentally’ (p. 103).

The role of anticipations make up the fourth notion related to intuitive knowledge (e.g. Fischbein, 1987; p. 61). With regard to anticipation, de Groot (1986; p. 71) refers to Jung’s description of intuition as ‘was etwas werden kann’ (i.e. ‘what something could become’) and states that this anticipatory element is always present. He explains that intuitive expectations anticipate what will or may happen and that intuitive judgements or evaluations anticipate the outcome of a more complete argument.

The fifth concept is that the access in memory of knowledge with an intuitive quality is different from the access in memory of knowledge without this quality. It is speculated that this differential access exists next to differences in verbalisation. It is hypothesised that the action-driven and perception-driven elements in learning ‘tune’ the knowledge and give it an intuitive quality. Though intuitive knowledge is hard to verbalise (and as a consequence in explicit tasks sometimes labelled ‘inert’) the access to knowledge with an intuitive quality is assumed to be ‘smoother’ than the access to knowledge with a more declarative quality. In other words, the intuitive quality causes the access to the knowledge in memory to be more efficient.

To recapitulate, active experience is indispensable for the acquisition of intuitive knowledge, and low verbalisability, perception, quickness, and anticipation are the most frequently cited observations in relation to intuitive quality of knowledge. A certain coherence can be indicated between the characteristics of discovery simulations, implicit learning, and the acquisition and features of intuitive knowledge (see Table 1 for an overview).

Table 1. Postulated relationships between characteristics of simulations and knowledge

<table>
<thead>
<tr>
<th>characteristics of discovery simulations</th>
<th>discovery learning</th>
<th>acquisition of intuitive knowledge</th>
<th>features of intuitive knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>perceptually ‘rich’ low transparency</td>
<td>perception-driven partly implicit</td>
<td>in rich dynamic environments extrapolation, inference partly unconscious</td>
<td>perception, visualisation hard to verbalise</td>
</tr>
<tr>
<td>active experience involved</td>
<td>action-driven</td>
<td>direct experiential involvement, more verbal learning not sufficient</td>
<td>quickly available</td>
</tr>
</tbody>
</table>

Although several questions remain unanswered, most researchers agree that, whatever the exact nature of the processes involved in the acquisition, and whatever the precise representation of intuitive conceptual knowledge, the processes involved in the manifestation of the intuitive quality of knowledge can be described as ‘the quick perception of anticipated situations.’

The WHAT-IF test and the explicit knowledge tests

Based on the analysis of intuitive knowledge as being characterised by a ‘quick perception of anticipated situations’, the test format that was developed for the present work is presented by following the words in the definition:

Quick: Response times to the items were included as an important indicator of the degree to which conceptual knowledge has an intuitive quality. It is believed that intuitive quality tunes knowledge and reflects a more efficient access to knowledge.

Perception: In the item format, perception is central, and contrasted with the emphasis of many other “traditional” tests of verbalisation. In the items therefore pictures (graphical or diagrammatic representations) are used accompanied by minimal necessary textual information.

Anticipated: It is argued that anticipation is important for intuitive knowledge. The items consist of situations (see below) in which values of variables are given. A value is then changed and a new situation is to be predicted or anticipated.

Situations: The items consist of a question and possible responses. In the question part, a description of a situation is given along with a change in that situation. In the response parts, descriptions of possible predicted situations are given. In other words, an item contains a situation, an action, and possible postsituations (or a condition, an action, and possible predictions). The condition-part is described by variables, which are given a value. In the action-part a value of one variable is changed and in the prediction-part possible new values of one of the variables of the condition-part are displayed. Situations constitute states of a simulated domain and are always made up of several variables.

The items, in which the ‘quick perception of anticipated situations’ is applied, are said to have a WHAT-IF format. Figure 1 displays two exemplary WHAT-IF items.

![Fig. 1. Exemplary WHAT-IF items.](image-url)
In Fig. 1, the left-hand item the condition-part is displayed in a picture of two balls, the action is given in text (an elastic collision) and the predictions are displayed in three pictures to choose from. In the right-hand item the condition-part and the action are displayed in pictures of electrical circuits and the prediction-part is given in text.

An important aspect of the procedure is that learners are not only asked to give a correct response, but they are also required to do so as quickly as possible. In the studies, latency corresponds to the time each item is displayed on the screen. The items have multiple choice format with three response alternatives. The task is computer administered. Learners cannot go back to previously responded items. The moment learners click with the mouse on the alternative of their choice, the item disappears from the screen and the next item pops up. Latency is measured as the time (in seconds) learners need to read and respond to the item. Latency is used as a measure of the extent to which the items are answered in an intuitive way. Quicker (correct) answers are supposed to better reflect intuitive knowledge than slower (correct) answers.

Across the five studies several types of tests were used, but learners across the studies always completed, next to the WHAT-IF pre-test and post-test, a definitional pre-test and post-test. The tests for definitional knowledge concerned declarative conceptual knowledge with the object of measuring discrete items of knowledge and declarative information that is, knowledge that was not connected and principled.

**Methodology**

The main objective of the current work was to investigate the instructional effectiveness of discovery simulations. Interrelated with this objective was the goal of testing the hypothesis that discovery simulations trigger the acquisition of intuitive knowledge. The intuitive knowledge, in turn, was to be measured with tests designed following the WHAT-IF format.

**Table 2. Overview of the studies**

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Level of education</th>
<th>Type of education</th>
<th>No. of WHAT-IF items</th>
<th>No. of factual items</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collisions I</td>
<td>46</td>
<td>first year university</td>
<td>biology &amp; computer science</td>
<td>37</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Oscillations I</td>
<td>28</td>
<td>first year university</td>
<td>social science</td>
<td>34</td>
<td>30</td>
<td>students not familiar with domain of oscillations</td>
</tr>
<tr>
<td>Oscillations II</td>
<td>63</td>
<td>first year university</td>
<td>technical science, physics</td>
<td>24</td>
<td>24</td>
<td>the simulation used has more explanations and feedback to assignments than in Oscillations I; students had completed introductory course on dynamics</td>
</tr>
<tr>
<td>Circuit I</td>
<td>41</td>
<td>middle vocational training</td>
<td>technical training</td>
<td>24</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Collisions II</td>
<td>112</td>
<td>pre-scientific general</td>
<td>general</td>
<td>24</td>
<td>20</td>
<td>57 worked with simulation</td>
</tr>
</tbody>
</table>

A first general prediction across the studies was that interacting with discovery simulations would result in gains in intuitive knowledge, as measured with the WHAT-IF tests, and not (so much) in increased explicit knowledge. The gain in intuitive knowledge would be reflected in increased WHAT-IF correctness scores and decreased WHAT-IF item response times. A second general prediction was that the discovery simulations with more instructional measures would better support discovery learning (compared to simulations with less instructional measures) and have higher WHAT-IF test scores. It was tentatively predicted that assignments would have a positive impact on the acquisition of intuitive knowledge, and that explanations would relate to higher explicit knowledge scores, but not to higher WHAT-IF scores.

A series of five experimental studies was carried out. In the first four studies, Collisions I, Oscillations I and II, and Circuit I, simulation environments were compared that differed with respect to the type and amount of support added. In the fifth study, Collisions II, a simulation was compared with a hypertext environment. Table 2 gives an overview of the studies and detailed can be found in: Swaak & de Jong (1996); Swaak et al., (1998) and Swaak & de Jong (in press).

In the following, the results of the five studies are compared. From the last study, only the results of the simulation condition are included.

The results of the studies

In the five studies, WHAT-IF pre-tests and post-tests were administered and changes in both correctness and item response time were measured. Also, definitional knowledge tests were used and their correctness scores were collected. In Figs 2, 3 and 4 an overview is given of the results of the WHAT-IF tests and the definitional tests. To make sensible comparisons, percentages instead of raw scores are used.

In all studies a significant gain in WHAT-IF correctness scores was found. Moreover, this gain was, in terms of effect sizes, substantial in all but one study (Circuit I). In addition, Fig. 3 shows substantial changes in WHAT-IF item response times for all studies except Oscillations I. Furthermore, in all studies except Collisions II, the effect size of either WHAT-IF correctness or WHAT-IF time change was larger than the effect size in definitional test gain (see Fig. 4).
Discovery simulations and the assessment of intuitive knowledge

With respect to WHAT-IF item response times, it should be noted that no standards of ‘quick’ or ‘slow’ responses or of ‘small’ or ‘large’ time changes were available. While the WHAT-IF format was used in all of the studies, features of the specific items and the particular students resulted in different average response times. A standard gain could have been obtained by having a control group of students completing the pre-test and post-tests without working with a discovery simulation.

The WHAT-IF pre-test correctness scores in Fig. 3 underscore the view that the topic of harmonic oscillations is difficult for the learners. In both Oscillations I and Oscillations II, the pre-test scores are at chance level. In Oscillations I, this low score was not surprising as the subjects were not familiar with the domain of oscillatory motion, and none of the subjects had a major in physics. However, Oscillations II included first year physics students who had just finished an introductory course on dynamics. Apparently, while the percentage correct on the definitional test was 67%, the course had not prepared students for the WHAT-IF test. In the other studies the WHAT-IF pre-test scores were above chance level, and in all but Collisions II the WHAT-IF pre-test scores were lower than the definitional pre-test scores.

Also, the relation between correctness and item response time of the WHAT-IF test was investigated. The main measure consisted of the correlation between WHAT-IF post-test correctness scores and WHAT-IF post-test time scores computed across items, within subjects (see Table 3).

Table 3. Overview of WHAT-IF post-test correctness x WHAT-IF post-test time correlations

<table>
<thead>
<tr>
<th>Study</th>
<th>Collisions I</th>
<th>Oscillations I</th>
<th>Oscillations II</th>
<th>Circuit I</th>
<th>Collisions II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.22</td>
<td>0.00</td>
<td>0.16</td>
<td>0.31*</td>
<td>0.49**</td>
</tr>
</tbody>
</table>

** p < 0.01, * p < 0.05
For Collisions II the results of the simulation condition are taken; for the other studies, the results across conditions are presented.

Table 3 shows that in Circuit I a trade-off was present between correctness and time for the WHAT-IF post-test items. This trade-off indicates that quicker answers had a higher chance of being incorrect, or, inversely, that the slower answers had a higher chance of being correct. The first interpretation may reflect guessing, the second interpretation suggests that the WHAT-IF items were answered in a reflective or thoughtful manner. A trade-off is difficult to reconcile with conceptions of intuitive knowledge, as postulated in this thesis. Latency was used as a measure of the extent to which the items were answered in an intuitive way. Quicker (correct) answers were supposed to better reflect intuitive knowledge than slower (correct) answers. A trade-off between correctness and time was also found in the simulation condition of Collisions II. For the other studies, no such trade-off was found.
In order to investigate whether the WHAT-IF tests measure a type of knowledge other than explicit knowledge, the WHAT-IF correctness and time scores were correlated with scores on the definitional knowledge tests and the hypotheses lists. In the hypotheses lists, subjects were asked to write down relations between given variables. The main correlations are given in Table 4.

Table 4. Overview of WHAT-IF post-test correlations with the explicit knowledge measures.

<table>
<thead>
<tr>
<th>Study</th>
<th>WHAT-IF post-test correctness x definitional test</th>
<th>WHAT-IF post-test correctness x hypotheses lists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collisions I</td>
<td>0.80**</td>
<td>–</td>
</tr>
<tr>
<td>Oscillations I</td>
<td>0.42*</td>
<td>0.01</td>
</tr>
<tr>
<td>Oscillations II</td>
<td>0.49**</td>
<td>0.12</td>
</tr>
<tr>
<td>Circuit I</td>
<td>0.41*</td>
<td>–</td>
</tr>
<tr>
<td>Collisions II</td>
<td>0.68**</td>
<td>–</td>
</tr>
</tbody>
</table>

** p < 0.01, * p < 0.05

For Collisions II the results of the simulation condition are taken, for the other studies the results across conditions are presented.

In all studies, the correctness scores of the WHAT-IF tests correlated substantially with the scores of the definitional tests. It should be noted that no consensus exists on when two measures are considered interchangeable. Usually, two measures with correlations below 0.75 are not treated as being interchangeable. Furthermore, part of the correlation can be explained by the identical test format of the two types of tests (i.e. overlap by variance due to identical test formats instead of identical constructs). Still, the correlations between the correctness scores of the definitional test and the WHAT-IF test for Collisions I and for Collisions II raise the question of whether the tests within these studies measured sufficiently different constructs.

In Oscillations I and Oscillations II the WHAT-IF post-test correctness scores were correlated with the hypotheses lists scores. No significant correlations between the WHAT-IF scores and the number and precision of correct hypotheses were found. This low correlation may indicate that the tests measured different constructs. This is interesting and seems to be in line with the hypothesis that intuitive knowledge is acquired directly, that is, without a declarative stage. However, the variance of the hypotheses lists scores was low. In other words the low correlation may be a result of this low variance (instead of the supposed differences in measured constructs).

To gain insight on whether the instructional measures we used influenced the acquisition of intuitive knowledge scores on the WHAT-IF test were correlated with information from logfiles. The correlations between the WHAT-IF tests and the use of assignments, explanations and number of runs with the simulation resulted in one consistent pattern across the studies. The pattern indicates that the more assignments were performed, the quicker the WHAT-IF response times. The other correlations were not consistent across the studies.

Conclusion

One conclusion is that the WHAT-IF tests seem able to measure the results of learning with discovery simulations. Another conclusion from these experimental studies is that assignments contribute most clearly to the instructional effectiveness of simulations. In other words, instructional support that enhances the rich, interactive, low transparent character of simulations appears to do best. This work started by
Discovery simulations and the assessment of intuitive knowledge

referring to a review study by Thomas & Hooper (1991), who concluded that results of simulations are not revealed by ‘tests of knowledge’ (p. 479), and that simulations can best give students the opportunity ‘to gain an intuitive understanding of the learning goal’ (p. 499). Perhaps they were right. However, it should be clear by this time that research on effects of learning are to guided by theories and constructs of learning. In other words a theory-driven or ‘construct-centered’ (see Nichols & Sugrue, 1997) approach is to be chosen, to make comparison of research results a sensible activity.

There is an apparent challenge in translating theories of learning into procedures to develop tests, and in validating constructs (Snow, 1989, 1990; Glaser, 1990; Pellegrino, 1992; Snow & Lohman, 1989, 1993). New conceptions about the psychological structures and processes involved in learning should go hand in hand with research on approaches to assessment. Together they might support validation and use of these concepts in applied instructional settings. Unless this approach is taken, assessment of learning outcomes will be superficial and may severely under-represent human cognitive abilities.

References


---

**Modelling software update:**

*Modellus* was reviewed in *JCAL* (Issue 16:1 - March 2000). The final beta version (release 2.5) is available on request by email to:

modellus@mail.fct.unl.pt.

The final 2.5 release will be available on CD and on the web in September.