Effect of Similarity-Based Guided Discovery Learning on Conceptual Performance

Pierre-A Mandrin
Daniel Preckel
ETH Zurich

Analogies are known to foster concept learning, whereas discovery learning is effective for transfer. By combining discovery learning and analogies or similarities of concepts, attractive new arrangements emerge, but do they maintain both concept and transfer effects? Unfortunately, there is a lack of data confirming such combined effectiveness. This experimental study involving 280 young students in the domain of physics showed that adequately structured similarities between mechanics and geometry improves conceptual performance (perception of functional relations) by as much for discovery learning as for conventional teaching texts with questions. Adequate structures were provided following Glynn’s teaching with analogy model. The learning form had no significant impact on concept performance. The effect of similarity increased when the level of difficulty of the treatment was raised. These results were found using a 2×2-factorial design. A qualitative questionnaire provided individual information about the usefulness of similarities and about learning strategies of the participants.

One of the major concerns in science education is how to foster thinking with similarities (or analogies), in order to improve the understanding of general science concepts. Indeed, quite a lot of concepts of usual curricula in secondary education yield similar features. The first author has started to integrate similarity of concepts in various approaches of guided discovery learning during school programs. He aimed at a facilitated transfer of concepts by similarity. The still unresolved question about the effectiveness of similarity-based discovery learning was the starting point of this study.

Discovery Learning

Various definitions of the term discovery learning exist within instructional research. As Anderson (2002) points out, the term inquiry teaching is often employed in US to design guided discovery learning as well as open inquiry learning. However, the conception used in this study is much more restrictive. Therefore, it shall first be delimited from other conceptions. A taxonomy of laboratory instruction styles has been proposed by Domin (1999) and may additionally prove helpful for more general purposes.

Guided Discovery Learning (or discovery learning) denotes, according to Domin’s taxonomy, an instruction style by which the learners are led to discover a predetermined outcome. The predetermined goal usually consists of finding some general principle. The learners commonly develop this general principle by studying specific situations (that is, by induction).

Open Inquiry Learning or inquiry learning, sometimes also called open discovery learning (e.g. Mayer, 2004) is described in detail by Hameyer (1999). The learners explore a new field in order to acquire a better understanding in this field. However, the teacher has not determined in advance any knowledge to be acquired (e.g., Domin, 1999). The students will rather determine by themselves, which question to investigate and how to proceed. This process is usually inductive.

The first conception of guided discovery learning appeared to be particularly suitable for use in the experimental design of this study. Thus, this type of discovery learning was focused on.

Learning aids are usually included and may be offered in various amounts. In the physics course, for example, if students are to discover the concept of heat flow, the teacher may distribute written information showing useful formal or graphical representations. By performing experiments, the students may discover, for example, that metal rods are better thermal conductors than wooden sticks. The instruction notes describe how the students should proceed and which experimental data they should collect. After completion of the experiments, the students should make conclusions on the properties of thermal conductivity.

Output studies in the past show, in average, an improvement of transfer effects when using guided discovery or inquiry learning rather than direct teaching, such as lectures, for instance. Mayer (2004) briefly summarizes the main findings. Positive effects in respect to school performance have been confirmed by meta-analyses, with a mean effect size of 0.4 or a corresponding correlation of $r = .2$ (Anderson, 2002; Fraser, Walberg, Welch, & Hattie, 1987; Neber, 1981). An effect has also been reported in conjunction with
computer simulations (de Jong & van Joolingen, 1998). The meta-analyses report that there are several moderating factors influencing the effect of discovery or inquiry learning, for example guidance and explicit vs. implicit approach. Often, successful learning is only made possible by appropriate and sometimes detailed guiding hints (Hameyer, 1999; Mayer, 2004). For example, the teacher instructs the students to examine the precise shape of an experimentally obtained graph of the thermal conductivity of copper. Moreover, inquiry oriented instruction should always precisely foster those competencies which match the immediate goal related to the desired transfer effect (Khishfe & Abd-El-Khalick, 2002). The students shall understand basic general principles, for example of thermal conductivity. The students shall then be able to explain, for example, why a metal plate is felt as being colder than a wooden plate. According to the findings of Khishfe and Abd-El-Khalick, a transfer effect is to be expected if the students themselves discover this general principle and explicitly formulate this principle (explicit-reflective discovery learning). Likewise, the students in the paleontology course shall raise a hypothesis on fossils, - about what a given living being may have looked like. The general principle resides in to what extent scientists are able to reconstruct properties of extinguished living beings and to what extent they are not.

In contrast to the explicit-reflective learning mentioned above, implicit learning occurs if transfer goals are not explicitly referenced (Khishfe & Abd-El-Khalick, 2002). In the last example of paleontology, no emphasis would be made on the reconstruction process of extinguished living beings. Implicit learning as such is usually employed, but may hardly allow the intended transfer goals to be reached.

Although open inquiry learning differs somewhat from guided discovery learning, transfer goals may be applied in both cases.

**Analogies and Similarities**

According to the definition of Duit (1991) and Duit and Glynn (1995), an “analogy” denotes “similarities between two domains with respect to specific properties” (Duit & Glynn, p. 44). Similarly, Gentner (1989, p. 201) describes an analogy as a “mapping of knowledge from one domain (the base) into another (the target)”, thus yielding “relational commonalities” between these domains. One frequently wishes to explain some new concept (target) by referring to a concept already known or understood (base or analog). Thus one considers an analogy relation between the analog and the target.

Example: There is an analogy between an electrical circuit and a water circuit. The electrical circuit is usually introduced as a target (new concept), and the water circuit is used as an analog.

Furthermore, similarities of concepts may link together different contexts or school subjects. In this way, geometry and physics concepts may show many corresponding features. However, students may not fully realize these correspondences by themselves, especially if these concepts are taught by different teachers.

Examples: There is a similarity between the geometric concept of similar triangles (used as an “analogy”) and the physical behavior of a bicycle on the inclined plane (used as a “target”). Or: The geometric concept of equal surface for rectangles shows similar features as the concept of equal torque for levers. These examples are further considered in the course of this article.

As could be shown in experimental studies, the use of analogies can foster the understanding of concepts. Performance in problem solving and comprehension tests can be improved, and experimental data exist for many different learning conditions. Mayer (1989) summarizes 20 studies; Hammond, Seifert, and Gray (1991) focus on analogical transfer; McVey (1993) combines analogies with meta-cognition; Schwartz (1993) treats symbolic visualizations; Newton and Newton (1995) report on children’s understanding; Lin, Shiau, and Lawrenz (1996) use pictorial analogies; Markmann and Gentner (1997) relate analogies with memory effects; Loewenstein, Thompson, and Gentner (1999) deal with multiple analogies; Antonietti (2001) treats and investigates basic learning mechanisms systematically. Several studies yield an effect of more than one standard deviation on performance. However, the use of analogies or similarities also may prove detrimental in certain cases. The learners may develop misconceptions (Duit, Roth, Komorek, & Wilbers, 2001).

For example, the above-mentioned analogy between the electrical circuit and the water circuit may give rise to the impression that an electric conductor could be "emptied" like a water pipe. Or in the example of the inclined plane, learners may mistake forces for lengths (such an example is reported in this study). Such misconceptions may lead to a reduced learning performance. Klauer (1991) systematically demonstrates problems which may arise when procedural support is lacking in the education of thinking with analogies.
Without the explicit approach of an analogy, the analogy learning may fail (Brown & Clement, 1989). Therefore, analogy learning generally requires a high level of guidance and learning hints.

Duit (1991) describes how situations of possible misconceptions during instruction with analogies may be confronted, namely by using the following two structuring techniques: “Teaching-with-analogy”-model or TWA-model (Glynn, 1989), and if necessary “Bridging Analogy” (Brown & Clement, 1989; Clement, 1993).

The TWA-model comprises the following steps:
1. The target concept has to be introduced first (e.g. electrical circuit).
2. The learners are then required to recall the analog concept (e.g. water pipe).
3. The learners shall identify similar features of both concepts (e.g. current).
4. These similar features are mapped (e.g. by drawing a table).
5. Conclusions are drawn about the concepts.
6. It is indicated where the analogy breaks down (e.g. behavior of open circuit).

Bridging Analogy: Analogy relations may sometimes fail to be recognized by learners. In this case, intermediate situations may be introduced between the analog and target concept (so called bridging analogy).

The authors of this study consider that the findings about analogies may also be applied to similarities of concepts (for example in the context of forces and geometry). In the Discussion, we support this view further as we compare our results with former research about analogies.

Application of Similarities (Including Analogies) in Combination with Discovery Learning

The TWA-model may be considered as an appropriate means for explicit-reflective inquiry (or discovery) learning along the lines of Khishfe and Abd-El-Khalick (2002). A similarity (with its correspondences and differences) is a form of general principle. Therefore, transfer is expected to occur with respect to this “similarity thinking”, and the “explicit approach” relies on the TWA-model.

So far, there has been a lack of experimental studies concerning similarity-based discovery learning. Though Flick (1991) did explicitly use analogies in discovery learning, this was not implemented as an experimental study. Lin et al. (1996) employed teacher-based pictorial analogies according to the TWA-model. Different social forms were run during the examined lessons: presentations by the teacher, questions addressed to the class, and discussions in small groups. Regarding conceptual performance, an improvement of up to three standard deviations was measured. However, the experimental and the control group were not taught by the same teachers, and the influence of the applied learning forms on the effect was not investigated.

**Objective and Hypotheses**

Until now there has been a lack of objective comparison between different learning forms for the use of similarities. In particular, it has remained unexplored as to which extent similarities, in combination with guided discovery learning, may foster conceptual performance when presented in a written assignment for individual learning (similarity-based guided discovery learning). In fact, there was no guarantee that similarities (including analogies) act independently of the employed learning form. So if the choice of the form came to interfere with the effect of similarities, concept learning could well be affected. Especially the time required for thinking might have an impact on how effectively concepts are acquired.

This study therefore aims to examine the combined effect of similarity and learning forms (guided discovery learning vs. expository learning). It is designed as an experimental and double blind study.

In this study, guided discovery learning was conceptualized according to the taxonomy of Domin (1999). The following hypotheses were investigated in this study:

**Hypothesis 1:** Similarities are as effective in the case of guided discovery learning as they are in the case of direct teaching, even if the time available for learning is not any longer.

**Hypothesis 2:** The effect of a similarity on concept learning is particularly large if study material challenges the learners (difficulty level as a parameter). Conversely, we would expect no effect in case of extremely easy material.

**Hypothesis 3:** The use of similarities affects concept learning stronger than the learning arrangement applied (here: expository or guided discovery learning). Discovery learning alone is not sufficient to simultaneously foster transfer and concept understanding.

**Method**

**Sample**

There were 298 young students (8th- and 9th-grade students) from eight Swiss-German schools participat-
ing in this study. The schools were partly in urban, partly in rural environments, and were located in three different cantons. The experiment was conducted by 12 teachers who taught the classes on a regular basis. For the final analysis, 280 students who met the following criteria could be included.

The authors and evaluators were excluded from participation as teachers; teachers and students were not informed about the goals of the study (double blind study). The experiment was integrated into the regular lessons and teaching environment. Data collection took place in July 2004 and between April and July 2005. The study design was experimental, i.e. the students of the 14 participating classes were randomly distributed to the different groups of the experiment, one of which served as control group (2005: four groups; 2004: two groups). In each class, the participants (who were assigned to different groups) worked simultaneously, silently and without any interaction, during the intervention and the tests. The intervention was in form of written assignments. The role of the teacher was to care for correct organization and timing. Explanations of the teacher were restricted to the introduction before the actual investigation.

Material

For this study the first author developed four different variations of written material for the learners. All documents were written in German. We first give a brief description and then present the more detailed content of each of the four variations. The first variation combined similarity and discovery learning (Variation SD). In this variation, an assignment for guided discovery learning was offered with a high degree of guidance due to goal-related questions and especially due to helpful figures. The assignment guided the learners to a predetermined similarity of concepts (proportionality) by using questions and pictures, and in accordance to the TWA-model. The similarity was between the geometry of similar triangles and the physical behavior of a bicycle on the inclined plane. The second variation combined similarity and teaching text with interspersed questions (Variation ST). By direct mediation, the Variation ST learners were confronted with a given description and interpretation of the before-mentioned similarity. The third variation showed no similarity but discovery learning (Variation ND). The learners received the same material as for Variation SD, but without reference to any similarity. Instead of the explicit question concerning the similarity, an alternative problem was posed on the same topic (“target” concept). The fourth variation showed no similarity but teaching text (Variation NT). As in Variation ND, the similarity was replaced by alternative information. Moreover, the learning form was the same as in Variation ST.

In 2004 the variants SD and ND were presented to the participants, in 2005 all four variants. Table 1 shows the study design (a 2x2 factorial design).

All four variations of assignment were comparable in terms of amount, structure and number and kind of figures. The assignment and all other materials for the students were examined in respect to physical correctness and adequate domain specific logic structure (physics / didactics context) by two didacts of physics from the ETH Zurich (experts). On average, the students worked on the assignment for 17 minutes (± 3 minutes). The time spent on the assignments did not differ significantly between the four groups.

During the experiment a construction figure of the target concept was projected (and was used as first step according to the TWA-model).

Content of Variation SD (first experimental group).

Two pictures represent similar objects (Figure 1: Target – inclined plane; Figure 2: “Analog” – similar triangles). The learners have to find out how the labeled sides of the triangles relate to each other (second step of the TWA-model). In this way, the learners are led to the concept of proportionality (or similarity). Subsequently, they have to search for the intended “similarity relation” (which does not appear in proper words in the instructions, however). The learners note correspondences and differences (third, fourth and sixth step of

<table>
<thead>
<tr>
<th>Variables</th>
<th>Discovery</th>
<th>Teaching Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>With similarity</td>
<td>Variation SD (experimental group) ( n = 77 )</td>
<td>Variation ST (experimental group) ( n = 65 )</td>
</tr>
<tr>
<td>Without similarity</td>
<td>Variation ND (experimental group) ( n = 83 )</td>
<td>Variation NT (control group) ( n = 55 )</td>
</tr>
</tbody>
</table>

Table 1. Study Design with Four Groups
the TWA-model, respectively). Noting correspondences makes the learners aware of the usefulness of the similar concepts. Consequently, they have to find the relation between the forces and lengths for the inclined plane (fifth step of the TWA-model). This relation corresponds to the concept of proportionality, yet in a different situation. All requirements of the TWA-model are fulfilled at this point. Subsequently, the learners check their discovery (target concept) with examples of their own choice. These freely chosen examples should allow for a deeper insight into the similar concepts. Finally, the learners resume their discovery in an abstract form (proportionality / formula). The goal of this activity is that participants fully understand and explain the meaning of the proportionality relation in the context of the inclined plane. On average, if learners encountered increasing difficulty during discovery learning, they had a lower chance to find the correct proportionality relation, expressed as a formula. For this reason, the rate of finding the correct formula appeared to be helpful to compare the difficulty of different versions of assignments.

The content of the assignment is as follows (texts are in quotation marks and translated from German; the answers given by participant number 286 are shown as an example in quotation marks):

Part 1: “Consider the following triangles” (Figure 2).
“Complete the following calculation, so that it corresponds to the figure: 1.5 : ... = 3 : ...”
Answer by participant: “1.5 : 4 = 3 : 8”.
Part 2: “The steep street” (Figure 1).
“What are the correspondences between this steep street and the triangles of Part 1?”
Answer by participant: “Right angle triangle; the smaller one is the triangle which determines $F_H$, the larger one is the street.”
“What is a difference?”
No answer by participant.
“Is there a similar calculation as in Part 1? What is it (with numbers)?”

---

**Figure 1.** Target - inclined plane with sloping site and parallelogram of forces (Version 2005).

**Figure 2.** "Analog" - similar triangles (Version 2005).
Answer by participant: “4.5 : 1.2 = 4 : 1.5”.

Part 3: “Draw a street yourself – make it either steeper or flatter! Take the same bicyclist as in Part 2!”

Participant draws a steeper inclined plane.

“What is the difference of your street compared to Part 2?”

Answer by participant: “Mine is steeper”.

“Draw the forces again!”

Participant draws a small triangle with the forces $F_G$ and $F_H$.

“How large are here: $l = \ldots$, $h = \ldots$, $F_G = \ldots$, $F_H = \ldots$”

Answer by participant: “$l = 10$ m, $h = 7.5$ m, $F_G = 3.5$ m, $F_H = 2.6$ m”.

“What is the calculation here?”

Answer by participant: “$7.5 : 10 = 3.5 : 2.6$”.

“...Replace all numbers by the corresponding symbols $l$, $h$, $F_G$, $F_H$;”

Answer by participant: “$h : l = F_G : F_H$”.

Unfortunately, participant number 286 did not (fully) succeed in finding the correct proportions, although the participant could partly take advantage from the similarity (the participant also confirmed this by the qualitative questionnaire). Moreover, the participant mistook the forces for lengths. This problem might have been avoided, if the participant had stated at least one difference between Figures 1 and 2 in Part 2.

**Content of Variation ND (second experimental group).**

Variation ND differs from Variation SD in a substitute depiction with a different accompanying text (see Figure 3) instead of Figure 2. The substitute depiction represents an inclined plane in which the subjects have to identify and measure two identical acute angles. Variation ND does not explicitly lead to any similar features, but has the same extent as Variation SD.

The content of the assignment is as follows: Part 1: “In this figure (Figure 3), the resulting force $F_H$ is composed of the components $F_G$ and $F_{\perp}$. Measure the angle $\alpha: \alpha = \ldots$. Find a further angle of equal size as $\alpha$. Label the angle as $\beta$. Check by measuring the angle: $\beta = \ldots$” Part 2: “The steep street” (Figure 1). “Complete the following calculation, so that it corresponds to the figure: $150 : \ldots = 4.5 : \ldots$” Part 3: This part is identical to Part 3 of Variation SD.

**Content of Variation ST (third experimental group).**

Variation ST corresponds to a standard teaching text with interspersed questions. The content is exactly the same as for Variation SD, except that learners do not have to discover anything in Variation ST. Instead, similar and dissimilar features of “analog” and target are explained, and the proportionality relation is shown.

The content of the assignment is as follows: Part 1: “Introduction: Consider the following triangles (Figure 2). The numbers fulfill a simple relation: $1.5 : 4 = 3 : 8$. Check the side lengths with a ruler. Label the measured lengths near the sides.” Part 2: “The steep street” (Figure 1). “In this steep street, there are similar triangles as in Part 1: one larger and one smaller triangle. Draw the right angles. Label in Part 1 the sides with the symbols $l$, $h$, ... corresponding to the sides of Part 2. Yet there is also an important difference: The paral-

---

**Figure 3.** Substitute representation instead of Figure 2 (Version 2005).
lelogram is made of arrows of forces. Forces are something else than the lengths of the triangle sides of Part 1. For this steep street, the calculation is: 150 N : 400 N = 4.5 m : 12 m. Does this calculation correspond to the one in Part 1? Look for the side “150 N”, then the corresponding side in Part 1, underline its value in the calculation. Etc.”

Part 3: This part is identical to Part 3 of Variation SD.

Content of Variation NT (control group).

Variation NT corresponds to a standard teaching text with interspersed questions. In respect to learning content, ND and NT correspond. Of all four assignments, Variation NT resembles most to the average usual teaching practice (for a characterization of regular classes see e.g. Hage et al., 1985, p. 76–87).

The content of the assignment is as follows: Part 1: “In this figure (Figure 3), the resulting force $F_H$ is composed of the components $F_G$ and $F_\perp$. Measure the angle $\alpha$: $\alpha = \ldots$” In the figure, one of the acute angles of the parallelogram has been marked as $\beta$ by the main author. “Measure the angle $\beta$: $\beta = \ldots$ The angle $\beta$ is of equal size as $\alpha$. There is another angle of equal size: Label it.” Part 2: “The steep street” (Figure 1). “Street length: $l = \ldots$, street height: $h = \ldots$, weight force: $F_G = \ldots$, force along the inclined plane: $F_H = \ldots$ In this figure, the numbers fulfill a simple relation: 150 N : 400 N = 4.5 m : 12 m. Check the side lengths with a ruler. Label the measured lengths near the sides. Put your measured values above.” Part 3: This part is identical to Part 3 of Variation SD.

To test the second hypothesis, the difficulty level of the tasks was slightly varied between the years 2004 and 2005. The rate of finding the correct formula was used as an indication of the difficulty. Figures 1 and 2 show the less demanding version of 2005. With this version, 72% of the students came up with a correct formula of the assignment. The year before, one of the four numbers appearing in the depiction was removed (in 2004, there were Variations SD and ND only). Only 51% of the students came up then with the correct formula. Therefore, the version of 2004 was on a higher level of difficulty. Unlike the assignments, the pre-tests and post-tests of 2004 and 2005 were corresponding. Thus it was possible to analyze the influence of changing the level of difficulty on concept performance.

Procedure

Seven days before the experimental intervention with the assignment, subjects underwent a uniform pre-test (only 2 days before for two classes with a total of 45 students, and 4 days in advance for one class with 22 students). During the experimental intervention, the subjects received only written information (assignment). Subjects underwent a uniform post-test after completion of the intervention. The questions of the post-test were preceded by a restatement of the end formula of the intervention assignment, thus ensuring that every participant could theoretically answer the questions. After the post-test, a qualitative questionnaire had to be filled out. Closed and open questionnaire items revealed, to which extent the learners benefited from a similarity. Subjects kept the documents of the experimental intervention until all parts of the experiment were completed (in the course of one lesson). The pre-tests and post-tests however, were collected immediately after the lapse of time (7 minutes each). The role of the teacher was to organize the lesson according to the study design.

Table 2 shows the sequence of the various experimental stages.

<table>
<thead>
<tr>
<th>Point in time</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Before pre-test</td>
<td>Make sure that necessary requirements are met (similarity of triangles; vector addition of forces)</td>
</tr>
<tr>
<td>2. 1 week before intervention</td>
<td>Pre-test</td>
</tr>
<tr>
<td>3. Beginning of intervention.</td>
<td>Intervention (four variations of assignment)</td>
</tr>
<tr>
<td>4. Immediately after intervention.</td>
<td>Post-test</td>
</tr>
<tr>
<td>5. Immediately after post-test.</td>
<td>Qualitative questionnaire</td>
</tr>
</tbody>
</table>

## Table 2

**Sequence of the Various Experimental Stages**

| Stages 3, 4, and 5 took place within one lesson. |          |

Stages 3, 4, and 5 took place within one lesson.
test is (text in quotation marks is translated from German): “Complete the construction of parallelogram of forces in the following figure …”. In the pre-test, the figure represents the inclined plane, the weight force \( F_G \) and the line of force perpendicular to the inclined plane. In the post-test, the figure represents the inclined plane, the force \( F \perp \) perpendicular to the inclined plane and the line of force of the weight force \( F_G \).

The “concept variable” referred to conceptual performance in the sense of complete comprehension of a functional relation (here: proportionality with an existing “analog” in the geometry of triangles). The questions in the pre-test are (translated from German): “The street could also be flatter than in the figure above” (same figure as for the algorithmic variable). “Would the resultant force \( F_H \) be larger, smaller or equal compared to the figure? Give the reasons for your answer in a short sentence and using \( F_H \) etc. How could we make \( F_H \) larger without modifying the steepness?” The questions in the post-test are: “What does the above relation express?” \( \frac{F_H}{F_G} = h : l \) “Complete the following sentences. 1. If the height \( h \) is higher at constant length of the street, the street is steeper and therefore … 2. If the street is longer at constant height, … 3. Also vary something else than \( h \) or \( l \): … 4. Why does this rule apply to any situation? Make a few sentences.”

In a certain sense, learning with or without similar concepts may be considered as a matter of problem solving strategy (the same consideration may also apply to former studies, see for example Lin et al., 1996). But similar concepts may also act on concept understanding. In contrast, problem solving strategies are not necessarily related to comprehension of a concept. That is why some of the test items above are not related to problem solving ability, but characteristic for the concept variable.

To avoid mere repetitions, the questions of the post-test are slightly different from those of the pre-test, although of the same type. The main author divided the expected solutions into equal solution steps, each step corresponding to one scale point for the evaluation.

The pre-test and the post-test were designed in such a manner that, on average, the learners attained about half of the scores (based on pilot tests). Thus a difference between pre-test and post-test cannot actually be interpreted as an “improvement” or “loss” of competence (in case of positive or negative difference, respectively).

In addition, the algorithmic variable was used to test whether differences in algorithmic solving ability resulted from the variations of assignment used in the intervention. If all variations were constructed to be fully equivalent except for the above-mentioned intentional differences, no differential effects should be found for the algorithmic variable. A confirmation is given in the results section.

**Test rating.**

Two raters independently evaluated the algorithmic and the concept variable in the pre- and post-test according to a rating scheme (two physicists holding at least a bachelor degree; the authors, experts and investigating teachers were excluded).

The two independent ratings corresponded well (Cohen’s Kappa: 0.8). Consequently, we calculated the average of the two ratings for all scores and used this for further analyses. Statistical evaluation was then performed on the four groups (SD, ST, ND, NT) using 2×2 ANOVA and multiple group comparison techniques.

**Group comparability in the pre-test.**

Before the main evaluation, equivalence of learning conditions of the four groups was tested. We first checked that the groups were on comparable levels of performance in the pre-test. None of the two variables yielded any significant difference at the .05 level, algorithmic variable: \( F(3, 279) = 1.33, p > .2 \); concept variable: \( F(3, 279) = 1.02, p > .3 \). Nor did a multiple comparison between the four groups yield any significant differences \( (p > .05) \).

Additionally, averaged grades in physics and mathematics were collected. They correlated significantly with the algorithmic variable in the pre-test scores (grades in physics, \( r = .22 \); grades in mathematics, \( r = .20 \)). The grades in physics also correlated significantly with the concept variable in the pre-test scores \( (r = .16) \).

**Results**

From this point, we accounted for pre- and post-test data and thus introduced the time dimension (repeated measures). We also referred to the 2×2 factorial design shown in Table 1. So we conducted a 2×2 repeated measures ANOVA with similarity (with similarity, without similarity) and discovery learning (with discovery learning, without discovery learning) as the two repeated-measures factors. In some cases, we additionally sought for possible interaction-time-effects. An alpha level of .05 was used for all tests.

**Algorithmic variable**

We first tested whether the variations of assignments
could cause undesired differences in algorithmic solving ability in the post-test. No significant differences could be found for the algorithmic variable between the four groups, $F(3, 279) = 0.90, p > .4$, multiple comparison between groups: $p > .1$. Moreover, no significant interaction-time-effect was observed ($p > .7$). Up to this point, all necessary tests had been completed to indicate equivalent learning conditions for all groups.

Concept variable

We shall now present the main results concerning the concept variable in the post-test.

Before testing our three hypotheses, we compared the scores of the different groups.

We found a significant group effect, $F(3, 279) = 2.74, p < .05$. A multiple comparison between groups yielded differences as follows: Group SD lay 0.45 points higher than Group ND in the score (LSD-test: $p < .01$; Bonferroni-test: $p < .05$); Group ST lay 0.36 points higher than Group ND in the score (LSD-Test: $p < .04$).

This means that both groups with similarity reached better scores than did the discovery learning group without similarity.

No further significant differences were found between the other group pair combinations ($p > .05$).

The result of the univariate analysis was in accordance with the measured interaction-time-effect with respect to the concept variable ($F = 2.73, p < .05$). That means that the scores with similarity also showed more progress with respect to the pre-test scores.

Table 3 shows the averaged scores of the concept variable in the post-test for all four groups (on a scale from 0 to 5). The highest score was obtained by the discovery learning group with similarity (SD). The lowest score was reached by the discovery learning group without similarity (ND).

We also compared the scores with and without similarity. To do this, we joined both groups with similarity to form a “similarity group” and both groups without similarity to form a „non-similarity” group, respectively. We found a significant improvement of 0.31 points in the score in favor of the similarity group, $F(1, 279) = 6.16, p < .02$: The difference was 0.3 standard deviations. Considering the interaction-time-effect, a slightly stronger effect appeared ($F = 7.94, p = .005$).

To conclude our evaluation, we routinely checked for undesired effects on the similarity and non-similarity groups. Regarding the algorithmic variable in the post-test as well as the algorithmic and conceptual variables in the pre-test, all differences were not significant ($p > .2$). There was no observable interaction-time-effect on the algorithmic variable either ($p > .8$).

Test of Hypothesis 2: In this test, we analyzed how the effect of similarity of concepts changed depending on the difficulty of the assignments. In 2004 the difference was 0.72 scale points for the 37 subjects, $F(1, 36) = 6.35, p < .02$ (assignments in 2004 comprised of less numbers in the depictions and were therefore more demanding compared to 2005). The score difference in 2004 was considerably larger then the score difference in 2005. This means that the similar concepts were more effective when applied in conjunction with a higher degree of difficulty (assignment yielding 51% success only). Hypothesis 2 was confirmed.

Test of Hypothesis 1: We did not find any signs indicating that similar concepts would be less effective when presented within discovery learning compared to teaching texts (difference between SD and ND: 0.45; difference between ST and ND: 0.36). Hypothesis 1 was confirmed.

Test of Hypothesis 3: We postulated that similar concepts had a stronger action on concept learning than did the variation of learning form (direct mediation with teaching texts or guided discovery learning). We thus checked whether discovery learning influenced the scores substantially compared to teaching texts. Overall, we found no significant difference between discovery learning and texts, $F(1, 279) = 0.51, p > .4$. Even with the comparison confined to samples with similarity (SD compared to ST) and without similarity (ND compared to NT), there was no significant difference. We again concluded the evaluation with a routine check for undesired effects. Discovery learning did not differ from teaching texts in respect to either the algorithmic variable in the post-test or to the algorithmic and concept variable in the pre-test ($p > .5$). Neither

Table 3

<table>
<thead>
<tr>
<th>Concept variable</th>
<th>Discovery</th>
<th>Teaching text</th>
</tr>
</thead>
<tbody>
<tr>
<td>With similarity</td>
<td>2.05 (1.19)</td>
<td>1.96 (1.00)</td>
</tr>
<tr>
<td>Without similarity</td>
<td>1.60 (1.00)</td>
<td>1.85 (1.03)</td>
</tr>
</tbody>
</table>

School Science and Mathematics 141
did we find any significant interaction-time-effect regarding the algorithmic or concept variable ($p > .3$). In summary, the learning form did not have any noticeable effect on concept learning. So Hypothesis 3 was confirmed.

**Temporal changes on the concept variable.**

The overall trends of the concept variable can be summarized as follows. We observed a general positive trend for the groups with similarity, compared to the groups without similarity. Within Group SD and Group ST, the means tended to increase from pre-test to post-test (positive trend, with similarity). Within Group ND and Group NT, a diminution of means could be observed from pre-test to post-test (negative trend, without similarity). Figure 4 illustrates this observation. The observed trends correspond to the significant interaction-effects which we already have quantified in the main results (group effect and effect of similarity group).

**Discussion**

In accordance to former empirical data about analogy learning, our study showed that the similarity of concepts allows improved conceptual performance (similarity structured according to the TWA-model). The effect was still observed when the similarity of concepts was included into guided discovery learning. In addition, we were able to show that the effect is more important when the level of difficulty in the intervention is raised.

As discovery learning is adequate to foster transfer, and as similarity of concepts remains powerful when combined with discovery learning, similarity based discovery learning should be recommended for practice. Furthermore, analogies and similar concepts including the TWA-model make the intended transfer goal explicit.

For us, the results of this study indicate the need for an appropriate cognitive structuring aid for discovery learning (see also Mayer, 2004). In our case, similar features served as a structuring aid. We are, however, of the opinion that it is not sufficient to briefly mention similar features, metaphors or analogies (Glynn, 1989). Rather, guiding aids have to be integrated explicitly. The TWA-model was already used in the past as a guiding aid. The assignments of this study were constructed on this basis. In our view, the TWA-model represents more than a problem solving strategy, as it fosters concept comprehension. In particular, the TWA-model may be necessary for successful analogy mediation, regardless of the chosen learning form.

In this study, two very different learning arrangements were used (guided discovery learning and teaching texts with interspersed questions). However, we could not find any indication that the choice of specific learning arrangements could influence the effect of
similar concepts significantly. This means that no interference between the employed learning form and the learning with similarities is apparent.

The results presented here are concerned with the concept of proportionality in physics. Further concepts could be applied in a similar way, although they were not subject of this investigation. For example, the inverse proportionality could be implemented in connection with the law of the lever (Figure 5): Instead of the similar triangles depicted above, rectangles of an equal surface would be used in the “analog” (Figure 6). The forces \( F_1, F_2 \) correspond to the vertical sides of the rectangles, the lever arms \( l_1, l_2 \) to the horizontal sides of the rectangles. The delineated construction of assignments can then be immediately transferred. The learners would have to find an equation for the inverse proportionality from Figure 6, compare Figure 5 and 6 (similarities and differences), induce the relation between forces and lever arms according to Figure 5, generate a new example of their choice and summarize their findings. As for the inclined plane, more demanding assignments can arise by presenting less than the four numbers shown in the depictions. With assignments of this kind, we would expect a similar effect as in our study.

In a similar manner, the analogy of two physical concepts may also be introduced. For example, a water circuit might be depicted instead of the similar triangles of our study, and an electrical circuit would appear instead of the inclined plane. The learners would have to

*Figure 5. Target - lever in balance.*

*Figure 6. “Analog” - rectangles of equal surface.*
find differences and correspondences for concepts like current, tension / pressure and resistance. They would have to draw circuits with large and narrow water pipes (and thus low and high resistance). They should then be able to explain, for example, how the electric tension and the current affect the resistance.

As former studies already found a significant effect of various analogies on concept performance, we expect those findings to apply more generally, when discovery learning is associated to analog or similar concepts. However, confirmation of this generalization would need additional experiments.

Due to the results we assume that the effect produced by analogy learning might be fundamentally independent of the teaching methods applied. To ensure that our findings can be generalized on a wider range of learning forms or methods, one would have to test additional methods or variations systematically (such as, for example, problem-based learning or more open discovery or inquiry learning).

References


**Author Note**

We are much obliged to the following people for their support in this study: counseling for the statistical analysis: Prof. Dr. Joachim Fischer; counseling and support during the early project phases: Prof. Dr. K. Frey; counseling and didactic expertise of teaching materials: Prof. Dr. R. Duit, W. Vogel and Dr. M. Lieberherr; counseling concerning theoretical concepts: Dr. J. Wilbers; execution of the experiment in classes (pilot and main phases, 2004 / 2005): T. Rösch, S. Büchi, Dr. N. Emmenegger, T. Fröhlich, S. Gamper, Dr. W. Grentz, P. Läßer, Dr. A. Prieur, H. Rubin, S. Schäublin, Dr. M. Schmid (execution and internal school coordination), M. Vogel-Stalder.