The effects of scaffolded computerized science problem-solving on achievement outcomes: a comparative study of support programs

Z. Fund
Bar-Ilan University, School of Education, Ramat-Gan, Israel

Abstract
This study examined the effect of scaffolding learning components in a computerized environment, for students solving qualitative science problems in a simulation of laboratory experiments. Four scaffolding components were identified (structural, reflective, subject-matter and enrichment) and used in different configurations to construct four unique cognitive and meta-cognitive support programs based on human teaching. These ranged from low (Enrichment) to full support (Integrated). We compared the scaffolded groups with one another and with a non-scaffolded control group. A ‘mathematics and reading comprehension’ questionnaire was used to divide the participants \((n = 473)\) into ability levels. At different points of time, achievement outcomes were measured by three open-ended subject-matter questionnaires, tapping knowledge and understanding. The findings indicate differential effects of the support programs mostly in the following order: Integrated > Strategic > Operative > Enrichment > Control. The structural component seems to be a sine qua non for success and has a consistent and powerful influence. The combination of reflection and structural components, however, are needed for superior achievement. Both reflection and subject-matter components work cumulatively over time.

Keywords
achievement outcomes, cognitive support, computerized problem-solving, reflective support, structural support, subject-matter support.

Computer learning environments, if appropriately designed, can support constructivist and exploratory learning, giving learners more agency in the learning process (Scardamalia & Bereiter 1991). Some argue that computerized simulations provide the proper environment together with effective learning activities and concrete experiences to enhance science problem-solving skills (Rivers & Vockell 1987). As learners are required to actively infer knowledge, they will also develop insights in science concepts (Swaak et al. 1998; Swaak et al. 2004).

Yet, learning within such environments pose challenges for most students, mainly because of the cognitive and meta-cognitive complexity of these experiences (Azevedo 2005), implying that elements of guiding, coaching and scaffolding are required (Swaak et al. 1998; Davis & Linn 2000). Since Rieber and Parmley (1995) demonstrated that pure simulations are not very fruitful, a number of studies have been conducted to integrate simulation supported activity with class activities (Hennessy et al. 2006), or to help learners accomplish specific tasks in simulation-based discovery learning processes (Rivers & Vockell 1987; Shute & Glaser 1990; Njoo & de Jong 1993; de Jong & van Joolingen 1998).

So far, most such studies have adopted an ad hoc support strategies-oriented approach (Zhang et al. 2004).
focused on the impact of specific support strategies, most of which are directed towards systematic and logical experiment and discovery activities (e.g. Rivers & Vockell 1987; Shute & Glaser 1990; Njoo & de Jong 1993). Zhang et al. (2004) hypothesized three interrelated conditions for discovery learning. Accordingly, they designed a simulation environment and embedded three types of support systems: interpretative, experimental and reflective. They examined effects on intuitive understanding, flexible application and knowledge integration (Reid et al. 2003). Because of the importance of the teacher’s role in optimizing use of simulation (Hennessy et al. 2006), we took this challenge even further and added a human-teacher-like support system. Scardamalia and Bereiter (1991) describe three idealized models of teaching upon which, the scaffolding programs of the current research are constructed.

The Teacher A model is a ‘task model’, with emphasis on activities and work, without concern for the cognitive processes involved in the activities. Learning is assumed as a by-product. In this model, the student’s role as worker or doer is intertwined with the requirement that good work be performed while new knowledge is constructed. The Teacher B model is ‘knowledge-based’, with the focus on understanding and responsiveness to the information needs of students (by setting goals, activating prior knowledge, asking leading questions, directing inquiry, etc.). This very effort, however, retains teacher control of the high-level cognitive processes. The Teacher C model turns control of proximal development to the students, thus promotes ‘growing out of dependency’ (ibid., p. 41). Scardamalia and Bereiter (1991) describe Teacher B as the prevailing model of good teaching, and warn that the Teacher C model could easily lead to overly romantic ideas, as the students are left to formulate goals, activate prior knowledge, ask questions, direct inquiry and monitor comprehension – on their own. Successful learning, according to them, seems to depend on more rather than less intense involvement of the teacher.

The models were implemented, in the current study, by appropriate scaffolding (support) programs – Operative, Integrated and Strategic, respectively (see Method section below), using four support components – in different configurations – that were found to be effective in computerized learning environments – structure, reflection, subject-matter and enrichment. The programs were provided appropriate worksheets (for each student, for each problem), found to be a suitable way of scaffolding students both in the overall structure and in the specific reasoning steps (Njoo & De Jong 1993; Vreman-de Olde & De Jong 2006). This enabled regulation of the problems to be solved at each lesson. Such scaffolding programs were expected to improve cognitive and meta-cognitive skills as well as working patterns and achievement outcomes (see Fund 2002, 2003). Design and implementation were followed by extensive efforts to find the most effective program.

Scaffolded learning environments (De Corte 2000), specifically simulation environments (Zhang et al. 2004), should be adapted to the levels of the learners to maximize support benefits. Less successful learners have more difficulties with simulation environments (Shute & Glaser 1990; Swaak et al. 1998) and in science problem-solving (e.g. Reif et al. 1976); therefore, in this study, we examine the scaffolding effects as a function of students’ academic level.

We will attempt to reach the following goals: (a) presentation of four scaffolding programs; (b) comparison of achievement outcomes for different academic levels; and (c) examination of the achievement outcomes as a function of time of exposure to the learning environment.

Method

The computerized learning environment

The computerized environment called Inquire and Solve (Educational Technology Center, Israel) is a microworld that combines a problem-solving environment with a simulation of laboratory experiments. It consists of 60 qualitative science problems, 42 of which are mapped to the science curriculum and textbook of the present research population (seventh grade). All problems enable application to familiar subject-matter. Each problem presents a question represented by textual and graphical components as exemplified in Fig 1a below (for more details see Appendix A).

In each problem, the student is required to ‘perform’ the simulated experiment, identify and ‘collect’ missing data using computerized tools and then apply appropriate reasoning processes to deduce the answer. In the problem portrayed, the student should use the ‘magnifying glass’ and ‘data pages’ tools (Fig 1b,c) to identify the given metals and find the falling order of the sticks as
well as their, respective, thermal conductivity scales. Sorting the materials and rods in ascending order of thermal conductivity would lead to the appropriate solution.

**The support components**

Four support components, described below, were found to be effective in computerized learning environments and served to construct the scaffolding programs described next. (Describing the support components and scaffolding programs is referred to Appendix A).

1. **A structural component** – providing a general framework that guides the student with the necessary cognitive skills (see Guzdial 1994; Rieber & Parmley 1995; De Corte 2000; the ‘experimental support’ of Zhang et al. 2004) to affect cognitive and work patterns.

2. **A reflection component** – providing a general framework to stimulate and activate meta-cognitive skills such as monitoring and control, self-assessment, and self-regulation (Zhang et al. 2004; Quintana et al. 2005). Such skills are necessary to learn rich domains using computerized environments (Azevedo 2005).
In the current paper, the reflection component is provided to prompt appropriate prediction and assessment of possible solutions, as well as explaining difficulties and mistakes.

3 A subject-matter component – clarifying ideas and concepts relevant to each problem, and addressing general domain-specific guidance. It also provides short guiding questions (specific instructions) to guide the solving process (similar to ‘interpretative support’ of Zhang et al. 2004; see also Swaak et al. 1998; De Corte 2000), and is expected to affect knowledge acquisition and improve understanding. This component was provided in two modes: a hierarchical mode (textual explanation and specific instructions), or a linear mode (specific instructions only), implementing Teacher B and Teacher A models, respectively (see below).

4 An enrichment component – introduced in accordance with the ‘infusion approach’ of Swartz and Parks (1992), including three to seven specific assignment questions (different for each problem) that relate the current problem to other relevant subjects or conditions, hence was assumed to improve understanding.

The structure and reflection components, when provided, were the same for all problems, while the other two components were, by definition, specific for each problem.

The support programs

Worksheets were prepared to implement the described teaching models, using the above components as building blocks as outlined in Table 1.

The ‘knowledge-based’ Teacher B model was implemented by the Integrated (INTEG) full support, i.e. all components, including: hierarchical subject-matter – general guidance (to activate prior knowledge and appropriate cognitive goals), and specific instructions to serve as ‘stimulating and leading questions to direct inquiry’ (Scardamalia & Bereiter 1991, p. 39). The structural component structures the main steps towards solution of the problem while the reflection component reminds the student to ‘monitor comprehension’ (ibid., p. 39) and reflect on the solving process. The Teacher C model was implemented by the Strategic (STRAT) support program, including the structural and reflection components (without specific instructions). This handing over of the regulating role – while searching the simulated experiment and solving the problem – to the learners enables ‘a higher level of agency in the knowledge-building processes’ (ibid. p. 40). Such placement of regulation makes the ‘difference [between Teacher B and C] in the control structure of activities in the zone of proximal development’ (ibid., p. 41).

The Teacher A model places the emphasis on doing the assigned work, assuming the activities result in learning (even without reflection). It was implemented by the Operative (OPER) support, including the structural component with specific instructions (the linear subject-matter component) to direct the inquiry. All three of these structural programs were enriched by assignment questions (enrichment component).

A low-support program – Enrichment (ENR) support as well as a control condition (CONTROL) were added for methodological purposes. The control condition gave no cognitive support at all. Instead, the students were directed to keep full notes and solve all problems for each specific subject, thus spending an equivalent amount of time on the program.

Based on Scardamalia and Bereiter (1991), the achievement outcomes were anticipated to be in the order: INTEG>OPER??/STRAT>ENR>CONTROL. The question mark denotes some uncertainty because of the unclear efficiency of Teacher C model.

Table 1. The components in the four support programs.

<table>
<thead>
<tr>
<th>Teacher B model</th>
<th>Teacher C model</th>
<th>Teacher A model</th>
<th>Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Integrated’</td>
<td>‘Strategic’</td>
<td>‘Operative’</td>
<td></td>
</tr>
<tr>
<td>Structure</td>
<td>Structure</td>
<td>Structure</td>
<td></td>
</tr>
<tr>
<td>Reflection</td>
<td>Reflection</td>
<td>Reflection</td>
<td></td>
</tr>
<tr>
<td>Subject-matter (hierarchical)</td>
<td>Enrichment questions</td>
<td>Subject-matter (linear)</td>
<td>Enrichment questions</td>
</tr>
</tbody>
</table>

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd
Each support program included 42 specific work-sheets composed of two parts: a scaffolding part (some or all of the a–h prompts in Appendix A, according to the specific support program), to be completed while solving the problem, plus enrichment questions to be answered immediately afterwards.

Three academic levels (low, medium, high) were compared, as well as treatment effects and control conditions.

**Sample**

In total, 473 junior high school (seventh grade) students (231 boys and 242 girls) from 16 classes in three schools – similar in size and level of socio-economic status – were studied. The classes were randomly assigned to the five treatment groups. All groups were represented in each school. Based upon the mathematics and reading comprehension (MRC) test, no significant differences in average ability level exist between any of the sample conditions in either mathematics ($F_{4,441} = 1.20$, $P > 0.05$) or reading comprehension scores ($F_{4,441} = 1.12$, $P > 0.05$).

**Instruments**

The content validity of the MRC was supported by the assessment of three experts. Twenty questions each, MRC (multiple-choice), were chosen from an appropriate computerized questions bank (Educational Technology Center, Israel). The questions were designed to distinguish among the three academic levels. Scores (in percentages) ranged from 0 to 20 for each test. Details concerning number of items, scoring ranges, reliability and correlation coefficients of MRC measures, as well as of questionnaires Q1–Q3 (described below), are presented in Table 2.

Three content-specific achievement questionnaires were designed, for different stages of the experiment – after 2, 4 and 6 months (Q1, Q2, Q3, respectively). Q1 and Q2 were open-ended and examined inferred knowledge, understanding and application. Q3 tapped knowledge inferred from the computerized problems as well as understanding.

Following a methodology similar to that of Linn and Songer (1991), two complementary indices were derived from each questionnaire – one ‘lenient’ (Measure A – *surface knowledge*) and one ‘stringent’
(Measure B – deep understanding). Measure A uses a scoring system that measures the low level of surface knowledge or understanding, crediting even fragmented knowledge. Measure B assesses deep understanding and integrated knowledge. On Measure A, each item is scored wrong (0) or right (1) – if at least one element of the response is correct. Additionally, its explanation was scored: missing or wrong = 0, partial or full = 1. On Measure B, the correctness and quality of the answer and explanation were assessed, generally yielding a scoring range of 0–3, and (occasionally) less items for Measure B (see sample questions and keys in Appendices B,C).

Three external judges formulated criteria for content validity and constructed keys for each measure. After content validity processing (95% agreement among judges), two judges scored each student’s responses (inter-judge reliability coefficients of 0.92, 0.93 and 0.95 were found for Q1, Q2 and Q3, respectively).

Q1 contained four open-ended questions, with a total of 12 subsections (see Appendix B), that included 18 items for A and seven four-point scale items for B.

Q2 included three open-ended questions with a total of ten subsections referring to the knowledge inferred from previously solved problems, and required deep understanding and application of the related subjects. Mistaken answers to the third set of questions were sorted for severe and moderate misunderstanding. Measures A and B had 20 and 15 items, respectively, with scores ranging from 0 to 20 and 0 to 45 (in percentages from the maximum available score) (see Appendix C).

Q3 included 20 multiple-choice items with a data page appendix attached. Seven were specially constructed for this research purpose, while the others were adapted from existing tests. Ten items required explanations and were also scored for Measure B. One such item is described, below:

Which roof best protects the house from the heat?
a. A metal roof (e.g. aluminum);
b. A cork roof;
c. A glass roof.

Explain your answer!

In order to answer this question, the student should use the heat conductivity scales in the supplied data page, indicating that cork has the lowest heat conductivity; hence, it is the best insulator.

Procedure

The five groups used the same textbook and worked within the Inquire and Solve computerized environment, once every 2 weeks. The experimental groups were assigned worksheets appropriate to their support programs. The treatments were conducted over a period of approximately 6 months, as part of the regular class program. The MRC questionnaire was administered before the research began, while the three subject-matter questionnaires (Q1, Q2 and Q3) were distributed later. Q1 was assigned after 2 months, when the students had completed at least the minimum number of compulsory problems (12); after 4 months (when 10–16 additional problems were solved); and after 6 months (when six to nine additional problems were resolved). At each stage, students who had successfully solved the compulsory problems had additional problems available.

Results

Differential effects of programs on achievement outcomes

A $5 \times 3$ (treatments $\times$ academic levels) multivariate analysis of variance (MANOVA) was performed for both: Measures A (surface knowledge) and B (deep understanding). The results indicate significant differences among the five treatment groups on both measures at all three post-learning intervals, (Q1, Q2, Q3), as follows: $Q1 - F_{8,822} = 10.66, P < 0.001$; $Q2 - F_{8,758} = 19.85, P < 0.001$; and $Q3 - F_{8,768} = 15.45, P < 0.001$. The three academic levels, obviously, formed a progression over the two measures on all three questionnaires: $Q1 - F_{8,822} = 35.21, P < 0.001$; $Q2 - F_{4,758} = 48.91, P < 0.001$; and $Q3 - F_{4,768} = 43.56, P < 0.001$. Significant interaction of academic level and treatment, however, was found only for Q1 with $F_{16,822} = 1.5, P < 0.05$.

A subsequent $5 \times 3$ one-way analysis of variance (ANOVA) was performed for each measure at each interval. The mean scores and standard deviation of Measures A and B for all groups are presented in Table 3, as well as ANOVA results for each measure for all questionnaires for students who had participated in all measures.

As seen in Table 3, significant differences were found among the groups on both measures, in the following order: INTEG>$STRAT>$OPER>$ENR>$CONTROL,
except for Measure A in Q1 and Q3, where OPER slightly (but insignificantly) outperformed STRAT. The scores of Measures A and B for the five treatments were subjected to contrast analyses (see below for details and results), while the scores of the academic levels were subjected to Scheffe’s post hoc analyses, to find the source of the differences among the five treatment groups and among the three academic levels, respectively. Higher achievement for higher academic level was found (Scheffe’s post hoc analyses) in the expected order. The ANOVA analysis indicates only one significant interaction of treatment and academic level: Measure A for Q1 ($F_{4,414} = 2.18$, $P < 0.05$). Figures 2 and 3 portray the treatment groups’ scores of Measures A (including the interaction) and B, respectively, divided into academic levels for all questionnaires.

Simple main effect analyses were performed to find the source of the above-mentioned interaction. Significant differences at all academic levels were found among the treatments, as follows: low ($F_{4,132} = 5.41$, $P < 0.001$), medium ($F_{4,133} = 12.20$, $P < 0.001$) and high ($F_{4,149} = 6.77$, $P < 0.001$). Scheffe’s post hoc pair-wise comparisons revealed INTEG and STRAT outperformed CONTROL and STRAT outperformed ENR in the low level, INTEG and OPER outperformed ENR and CONTROL in the medium level, and the three structural groups outperformed CONTROL in the high

---

**Table 3. F-scores, means, standard deviations and ANOVA results for all programs ($n = 395$).**

<table>
<thead>
<tr>
<th>Groups</th>
<th>n</th>
<th>Q1 Measure A</th>
<th>Q1 Measure B</th>
<th>Q2 Measure A</th>
<th>Q2 Measure B</th>
<th>Q3 Measure A</th>
<th>Q3 Measure B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M (sd)</td>
<td>M (sd)</td>
<td>M (sd)</td>
<td>M (sd)</td>
<td>M (sd)</td>
<td>M (sd)</td>
</tr>
<tr>
<td>Integrated</td>
<td>80</td>
<td>56.20 (22.58)</td>
<td>66.23 (22.96)</td>
<td>59.08 (26.18)</td>
<td>50.77 (25.86)</td>
<td>63.66 (17.58)</td>
<td>59.20 (22.37)</td>
</tr>
<tr>
<td>Strategic</td>
<td>73</td>
<td>52.05 (25.05)</td>
<td>61.63 (25.14)</td>
<td>52.05 (23.54)</td>
<td>42.47 (21.96)</td>
<td>59.38 (16.91)</td>
<td>55.15 (23.60)</td>
</tr>
<tr>
<td>Operative</td>
<td>77</td>
<td>53.55 (27.66)</td>
<td>59.45 (21.56)</td>
<td>43.49 (23.54)</td>
<td>33.70 (21.97)</td>
<td>62.95 (19.23)</td>
<td>49.39 (23.12)</td>
</tr>
<tr>
<td>Enrichment</td>
<td>91</td>
<td>41.27 (26.77)</td>
<td>52.49 (25.69)</td>
<td>35.00 (24.43)</td>
<td>29.86 (21.10)</td>
<td>39.94 (19.17)</td>
<td>29.67 (22.32)</td>
</tr>
<tr>
<td>Control</td>
<td>74</td>
<td>35.64 (24.71)</td>
<td>48.99 (24.16)</td>
<td>29.86 (22.10)</td>
<td>19.30 (19.07)</td>
<td>50.08 (17.25)</td>
<td>29.67 (22.32)</td>
</tr>
<tr>
<td>F-scores (4,380)</td>
<td></td>
<td>20.41***</td>
<td>13.02***</td>
<td>32.94***</td>
<td>42.06***</td>
<td>11.14***</td>
<td>31.43***</td>
</tr>
</tbody>
</table>

*** $P < 0.001$.  

**Fig 2** Measure A scores of the scaffolded programs for the three academic levels. Enr, Enrichment; H, high; Integ, Integrated; L, low; M, medium; Oper, Operative; Strat, Strategic.
academic level. These findings may indicate that for surface knowledge (Measure A), low-level academics benefit most from reflection, while mediums benefit most from subject-matter. High-level academics benefit from both. For deep understanding (Measure B), there were no interaction effects.

Contrast analysis results

Based on theoretical considerations and anticipated differences, the scores of Measures A and B (see Table 3) were subjected to contrast analyses to find the source of differences among groups. We contrasted structural vs. non-structural groups (Contrast 1), structural with reflection vs. without (Contrast 2), structural with a subject-matter component vs. without (Contrast 3). We also contrasted OPER vs. ENR and ENR vs. CONTROL. An outline of these contrasts is presented in Fig 4. Specific results are presented next as well as a summary in Table 4.

Specific findings

Q1: The structural groups differed from the non-structural on Measures A ($t = 6.21, P < 0.001$) and B ($t = 5.02, P < 0.001$), and OPER outperformed ENR on Measure A ($t = 3.16, P < 0.01$). Both findings imply a positive effect of the structural component.

Q2: Here too, the structural groups differed from the non-structural on Measures A ($t = 7.76, P < 0.001$) and B ($t = 8.74, P < 0.001$). Additional significant contrasts were found between INTEG and STRAT in Measure B ($t = 2.30, P < 0.05$), and between INTEG and OPER on both Measures A ($t = 4.06, P < 0.001$) and B ($t = 4.79, P < 0.001$). Additionally, we found differences between OPER and ENR on both Measures A ($t = 2.28, P < 0.05$) and B ($t = 2.40, P < 0.05$).

Summing up – there exist differential effects among groups INTEG>STRAT>OPER>ENR, CONTROL for Measure A, and even greater differentiation for Measure B with INTEG>STRAT>OPER>ENR, CONTROL. In
addition to the positive effect of the structural component, adding the reflection component (included in INTEG and STRAT, but not in OPER) provides the strongest support system. The subject-matter component (included in INTEG and OPER) had only a small effect – mainly in Measure B.

Q3: Here too, the structural groups differed from the non-structural groups on Measures A ($t = 6.17$, $P < 0.001$) and B ($t = 10.06$, $P < 0.001$). Additional significant contrasts were found for Measure B between INTEG and OPER ($t = 2.58$, $P < 0.01$), and between ENR and CONTROL ($t = 2.93$, $P < 0.01$).

Summing up – the differentiation among groups is INTEG, STRAT, OPER $>$ ENR, CONTROL for Measure A, and even more for Measure B – with INTEG, STRAT $>$ OPER $>$ ENR $>$ CONTROL. This is the only case where the ENR outperformed CONTROL. The substantial effects of the structural and reflection components on achievement outcomes may be deduced from these results.

### Analysis and frequencies of mistakes

For further analysis of deep understanding (Measure B) among the treatment groups, mistakes and misconceptions in the third set of questions in Q2 (see Appendix C) were counted and sorted into severe and moderate misunderstandings, yielding frequencies (in percentages) for INTEG, STRAT, OPER, ENR and CONTROL: of 30.3, 49.4, 61.0, 66.3, 64.5, for at least one mistake, and frequencies of 22.5, 35.1, 52.4, 56.5, 51.3, for severe mistakes. The majority of reflection group (INTEG and STRAT) students made no mistakes (and less severe ones when they did err). Statistical significance was reached on the differences among frequency distributions ($\chi^2 = 31.31$, d.f. = 4; $P < 0.001$).

### Discussion

Our predominant finding was that students in INTEG and STRAT did better than those in OPER. This finding was somehow surprising, as the subject-matter component, by definition, should affect achievement. INTEG and STRAT differ with regard to the subject-matter component, yet they attain quite similar achievement outcomes, while OPER though including this component has lower achievement. Reflecting on Scardamalia and Bereiter’s (1991) models, we imply that Teacher C model (STRAT) scaffolds almost as well as Teacher B (INTEG) (and both better than Teacher A – OPER), far from being the ‘romantic idea’ of which we were forewarned, at least when scaffolding – not teaching – is concerned. It implies that the subject-matter component has a limited contribution to achievements – on the one hand, while the combination of structure and reflection components has a uniquely strong effect – on the other hand. Both issues are further elaborated upon next.

The structure component, which supplies a general framework for solving problems, has a consistent and powerful influence (the three Teachers’ models which include this component outperformed the non-structural groups at all three intervals), and seems to lead to effective work patterns (see Fund 2002). Explaining the solution, one of its sub-components, creates links between previous and newly acquired knowledge (Chi et al. 1994), thereby improving knowledge construction and understanding. Further improvement is attained by the internal dialogue engendered when obliged to commit and explain the solution in writing (Scardamalia & Bereiter 1991). Yet, structure alone is insufficient for maximum benefit. The reflection component (included in Teachers B and C), seemingly, encourages meta-cognitive processes, which play a crucial part in knowledge construction. This, in turn,
effects understanding, activates internal changes, enhances self-responsibility and leads to more effective solving strategies (Fund 2002, 2003). The combination of reflection and structural components creates somehow, a ‘constructivist-oriented learning environment’ which requires (among other features) ‘drastic changes in the role of the teacher’ (De Corte 2000, p. 264). The scaffolding of Teachers B and C induces, apparently, an ‘intellectually stimulating climate’, as it ‘models learning and problem-solving activities (prompts structure and reflection components), . . . provides support to learners through coaching and guidance (both scaffolding), and fosters students’ agency over and responsibility for their own learning (mainly in Teacher C, to a lesser extent in Teacher B)’ (ibid., p. 264). Hence, the ‘winning combination’ of structural and reflective components (Teachers B and C) leads to constructivist knowledge acquisition and deeper understanding. In our case, i.e. computerized science problem-solving and discovery learning, this combination encourages the whole cycle of predict-observe-compare-explain to occur. The student ‘observes’ and examines the simulated experiment problem (structure), ‘predicts’ an answer (reflection), ‘compares’ the predicted answer with the correct one (reflection) and ‘explains’ the solution (structure) or difficulties and mistakes (reflection).

The limited contribution (though somewhat incremental) of the subject-matter component might be attributed to its specific instructions sub-component, a sort of expository instruction (Swaak et al. 2004) that reduces the active and constructive acquisition processes, and makes the learner merely a passive recipient of information (De Corte 2000). Our previous studies show that specific instructions develop some dependency on the support and reduce the student’s ability to conduct an independent inquiry. Its absence in the STRAT support induces more effort, more demanding problem-solving process, increasing challenge and motivation (Fund 2002).

Comparing Teacher B with Teacher C indicates that the additional subject-matter component results in higher achievements (but mainly insignificant, except for deep understanding in the second interval). Yet, comparing Teacher B with A (both include subject-matter component) indicates significantly higher achievement for Teacher B (with reflection), at most intervals. Thus, we can assume that the reflection component adjusts the ‘balance’ between the powerful constructivist environment and ‘regular’ learning environment, or between self-regulation and external regulation. The reflection component, as above mentioned, increases the responsibility of the students for their own learning, which probably affects the quality of knowledge construction, thus resulting in higher achievements. It also seems to reduce in INTEG – Teacher B some of the disadvantage of ‘taking the learner by the hand’ with specific instructions. At the same time, it compensates for the reduced guidance in STRAT – Teacher C, in which the teacher’s goal is to turn over to the students some of the high-level knowledge-building processes.

The interaction of time and scaffolded environment is assumed to affect learning efficiency (Swaak et al. 2004). Examination of differential achievement outcomes as a function of time of exposure to the scaffolded learning environment (the third goal of the paper), based on the differences among the treatments, is discussed next (see Table 4 for a summary of these findings).

Even after only a short exposure to the support programs (Q1), differences appeared between the structural and non-structural groups, favouring the structural groups (with no differences among them). This seems to indicate that as the three teachers begin their scaffolding, each succeeds in his own way. It appears to take time to get used to a new teaching style; thus, the unique effects of each teacher’s support system are not yet evident. Thus, effects are similar. This accords with previous results, based on interviews and observations that indicate that adaptation time to a new environment was about 2 months (Fund 2002). Thus, the structure component seems to have an early and dominant influence, while the influence of the other two components is still latent. The above-mentioned contradicting effects of subject-matter component – probably – take time to resolve; the very beginning, and still weak reflection effects (see below), both might explain the similar achievements of the three structural groups in Q1.

The only interaction of treatments and academic levels on surface knowledge at this stage, with positive (but limited) effects of subject-matter (INTEG and OPER) on medium and high levels, might be due to its scaffolding effects for those capable, with discovery task within their zone of proximal development (see also Zhang et al. 2004). The demanding cognitive ability and capacity required for the acquisition of new
knowledge (e.g. De Corte 2000) might explain the higher achievements for higher academic levels in all support programs.

After a longer exposure (i.e. at Q2), increased differentiation was found among the three structural groups. INTEG (Teacher B) outperformed STRAT (Teacher C), but only on Measure B (deeper understanding and fewer mistakes). We may relate this to a moderate effect of the subject-matter component to all students. At the same time, INTEG and STRAT (Teachers B and C) outperformed OPER (Teacher A), in both A and B measures. This seems to imply that the reflection effect increases as a function of elapsed time even more than the subject-matter effect.

After continued exposure to the treatments (Q3), the disparity among groups was somewhat diminished. Now, no differences were found among the reflection groups (Teacher B with subject-matter vs. C without), implying that over time the reflection component out-weighs the subject-matter effect. By the end of the treatment period, the reflection (Teachers B and C) and OPER (Teacher A) groups differed only in Measure B (having differed in both measures at Q2), suggesting that the subject-matter effect on achievements continues to increase, improving OPER, at least as demonstrated by our lenient measure (surface knowledge).

The long period of exposure to program treatments revealed for the first time some superiority of ENR over CONTROL. This might be due to growing familiarity with the test measures (similar patterns of pen and paper questions in Q3 and in enrichment assignments). This weak and late effect differed from the positive effects of assignments in Swaak’s et al. (1998). This difference should invoke a further examination of the effects of assignments that constitute an integral part of the work, as compared with those answered afterwards and playing an ‘outside’ role.

Even though Teacher B model (INTEG) results in the highest achievement outcomes, reasonable success is attained with Teacher C scaffolding (STRAT), which is easy to prepare and integrate into any computerized learning environment. Thus, a practical implication of the current research is to supply a Teacher B scaffolding for the first stages of working in a computerized learning environment, when the learners still need considerable external support to compensate for voids and weaknesses. In time, the learners become accustomed to the environment, and then the support can be gradually withdrawn, by ‘deleting’ the subject-matter component, yielding Teacher C scaffolding.

A final word about the incremental effect of reflection and subject-matter components. The reflection component works cumulatively over time, probably as the reflective processes become internalized. The subject-matter incremental effect, however, might also be explained on the basis of the Adaptive Character of Thought (ACT-R) theory (e.g. Anderson et al. 1995). This theory suggests that declarative knowledge is converted into production rules in the context of problem-solving activity. It is assumed that both declarative and procedural knowledge acquire strength with practice. As the subject-matter component is involved in these processes, it has a cumulative character, thus strengthening and increasing as students solve more problems and acquire practice.

Appendix A

The Integrated support model

Worksheet of problem 47

Student name: ____________________

☒ a) The problem to be solved

____________________________________________________________________________

☐ b) General guidance: Three rods, with wooden sticks attached to them with wax, are equally heated. When wax attached to each rod melts – due to the conducted heat – the stick falls. The sooner it falls – the better thermal conducting is the rod. You should find the metal each rod is made of, by comparing its conductivity to the conductivity of the other rods, as per the ‘data pages’.
c) The important data

Which metals are the rods composed of?

Why is wax used to connect the sticks? Explain!

From which rod did the stick fall first? ______ second? ______ third? ______

Which rod is the best conductor? ______ medium? ______ worst? ______

Use ‘data pages’ to determine the thermal conductivity of the three metals.

Which metal is the best conductor? ______ medium? ______ worst? ______

d) Proposed answer: __________________________

e) Did you give a correct answer? (Use the flag) yes/no

f) The correct answer: _________________________

g) Explain your answer and how you obtained it ________________________________

h) If you proposed a wrong answer, how does it differ from the correct answer? Explain why you were wrong _______________________________

Enrichment questions

1. Two rods are heated by non-equal burners, and the stick falls first from rod 2. It can be deduced from this experiment that: (a) thermal conductivity of rod 2 is higher than rod 1; (b) thermal conductivity of rod 2 is lower than rod 1; or (c) because of the different conditions, no conclusion can be deduced.

   Explain: ________________________________

2. Five sticks are gummed with wax along a heated rod. The stick that falls first is located: (a) farthest from the heater; (b) in the middle of the rod; (c) nearest the heater

   Explain: ________________________________

Notes:

1. The symbols denote:
   Subject-matter component:
   Specific instructions (in prompt c)
   Hierarchical mode (prompts b + c);
   Linear mode (prompt c)
   Structural component [prompts a, c (without specific instruction), f, g]
   Reflection component (prompts d, e, h)

2. The Strategic support (STRAT) included prompts: a, c (without specific instructions), d, e, f, g, h and Enrichment questions.

3. The Operative support (OPER) included prompts: a, c (with specific instructions), f, g and Enrichment questions.

Appendix B

Representative questions from Q1.

Three sets of questions refer to the problem presented in Fig A1 below.

The questions require the student to describe the experimental system, the missing information and how to collect it. The students submit their answers (twice) before receiving additional screens and more questions, such as:

• the stage at which the collected information enables reaching the proper answer;
• possible reasons for mistaken answers;
• the correct answer (see sample keys below);
• the answer if the same experiment shows different results.

Sample keys (rubrics) for Measures A and B of the presented problem (see Fig A1 above)

Two fragmented factors – the answer and the explanation (yielding two items) are scored on Measure A: The correct answer (the gas amount in Vessel 2 < Vessel 1) is scored 1, and 0 otherwise. The explanation is scored 1 for a partial (‘the water went higher’) or full explanation (‘the water replaced the missing amount of air’), and 0 when there is no explanation or a wrong explanation.

On Measure B, the correctness and quality of the answer and explanation are scored differentially, as follows: wrong answer and wrong or missing explanation = 0; correct answer with wrong or missing explanation = 1; correct answer and partial explanation = 2; correct answer and full explanation = 3.

Appendix C

Sample from Q2 (Question 3a), and sample keys for Measures A and B

A screen of a previously solved problem is presented (Fig A2). The student must explain why the heights of the oil drops have changed.

Sample keys (rubrics) for Measures A and B

Two fragmented factors (yielding two items) of the explanations are scored on Measure A:

• The effect of heat on gas volume in the vessels when mentioned = 1, when not mentioned = 0;
• The temperature differences among the containers when mentioned = 1, when not mentioned = 0.

On Measure B:

• The answer is scored 0 when severely wrong (e.g. the oil drops get hot) or a tautological explanation is offered;
• 1 point for a good explanation of only one factor, and possibly a moderate mistake;
• 2 points for a good explanation of two factors, and possibly a moderate mistake;
• 3 points for a good explanation of the two factors, with no mistake.
The problem concerns the relationship between mass, volume and temperature of gas. In the second episode, the vessels are immersed in the water containers. In the third episode, the oil drops’ heights would be changed.

Information from ‘magnifying glass’ in the first episode.

Fig A2 The problem screen presented for the third set of questions in Q2.

References

Rieber L.P. & Parmley M.W. (1995) To teach or not to teach? Comparing the use of computer-based simulations in...


