Effects of cueing and collaboration on the acquisition of complex legal skills

Hans G. K. Hummel*, Fred Paas and Rob Koper
Open University of The Netherlands

Background. To overcome the teacher bandwidth problem in supporting large groups of students, both automated process support (cueing) and face-to-face feedback by peers during small group work (collaboration) can be provided to students.

Aim. The purpose of this experimental study was to examine whether a multimedia practical containing cueing could be effectively combined with peer feedback to support the acquisition of the complex skill of preparing a plea in court.

Sample. In the context of a regular court practical, 50 junior law students at a Dutch university individually studied a multimedia practical and participated in small group discussions about intermediate learning outcomes of the practical.

Method. The study examined the effects of cueing and collaboration on training outcomes and transfer pleas, and on cognitive activity during collaboration, by combining the multimedia practical and small group collaboration to support the complex task of preparing a plea in court.

Results. Both cueing and collaboration positively influence training outcomes, with participants without cueing benefiting most from additional collaboration. Transfer plea scores reveal a positive effect of collaboration but a negative effect of cueing. Analysis of discussions during small group collaboration reveals a negative effect of cueing on the level of cognitive activity.

Conclusion. Peer feedback during small group work indeed appears to be a feasible option to be combined with (or partially) substitute individualized cueing when training complex learning tasks.

Distance education and lifelong learning call for individualized learning to support large and heterogeneous groups of students, especially for training complex tasks. Direct teacher-student interaction is not considered an economically feasible option in upscaled learning environments. As a consequence, automated support via intelligent instructional techniques has long been regarded as the only viable solution. For instance, a considerable amount of energy and money has been used to develop multimedia practicals with techniques to support task execution to overcome this so-

*Correspondence should be addressed to Hans G. K. Hummel, Educational Technology Expertise Centre, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands (e-mail: hans.hummel@ou.nl).

DOI:10.1348/000709905X85629
called teacher bandwidth problem (Wiley & Edwards, 2003, p. 2). Cueing is one of these instructional techniques, defined by Hummel and Nadolski (2002) as a technique that facilitates cognitive processes that enable problem-solving transfer; that is, the interpretation and construction of problem schemas. They studied how automated cueing could be provided to learners by focusing on two cueing formats: worked out examples (WOEs; see Renkl, 2002) and process worksheets with leading questions (see Land, 2000). The results of their pilot study show that a combination of examples to stimulate near transfer on similar tasks, and worksheets to stimulate far transfer on different tasks is perceived by students to guide and promote problem solving. Experimental studies with these formats have revealed encouraging results, such as positive effects on the training and transfer of complex problem-solving skills (Hummel, Paas, & Koper, 2004, 2006).

Although multimedia practicals with cueing offer powerful individualized learning environments and decrease the working expenses of education, they are also expensive to develop and suffer from a number of weaknesses during operation. The laborious and costly support of individual students by either teachers or automated systems represents a serious problem to all educational institutes. Although cueing may be effective, alternative ways to provide support for training complex skills need to be further explored. This study will explore possibilities to combine (pre-designed) automated cueing, using examples and worksheets, with collaboration, using peer feedback during group discussion. Among others, Wiley and Edwards (2003) identified collaboration as a promising solution to the high costs of support. This introduction will now address the relation between collaboration and peer feedback, the extent to which collaboration should be structured in advance and the relation between structure and cognitive activity during group discussion.

Collaboration and peer feedback

The potential of teamwork or other types of face-to-face collaboration for learning has been demonstrated by various studies in a variety of domains (see Barlow, Phelan, Harasym, & Myrick, 2004; Pawar & Sharifi, 1997; Pearce & Ravlin, 1987), and for computer-supported collaborative learning (CSCL) environments (e.g. Gunawardena, Carabajal, & Lowe, 2001; Gunawardena, Lowe, & Anderson, 1997). The interaction between learners in CSCL can lead to further elaboration and refinement of individually constructed schemas, since it incites learners to explicate the actual level of schema development and demands them to explicitly compare their own schemas with schemas of others as to defend or criticize (Jeong & Chi, 2000). Wiley and Edwards (2003) investigated the potential of on-line self-organizing social systems (OSOSS) without any central guiding authority where users provide each other with peer feedback or ‘real-time peer review’ to accomplish any significant purpose. For collaborative problem solving (CPS), they found that learners were creative in finding ways to support each other’s learning on the fly. The only thing needed for CPS, according to Nelson (1999), is a learning environment that enables collaboration and stimulates the exchange of ideas and information. Wiley and Edwards focus their research on web-based CSCL infrastructures from which OSOSS is expected to ‘simply’ emerge without centrally adding any content, commentary, structure or user support in advance. This study explores whether cueing can be fruitfully combined and balanced with face-to-face,
unstructured, small group collaboration in order to further improve learning outcomes.

Collaboration and structure

Researchers also state that for effective problem solving during collaboration, there ‘...seems to be a need to structure the learning in small group interaction in advance in a way that will prompt students to elaborate the problem, reflect on the solution process, and really construct relationships between prior and new knowledge’ (Mevarech & Kramarski, 2003, p. 450). However, the means by, and extent to, which collaboration should be structured in advance, whether this should be face-to-face or computer-supported, how individual and group support could be balanced and what ‘collaborative tools’ could be applied in collaboration remain largely unresolved issues.

The structure of collaboration can be operationalized in various ways: the collaboration process can be structured by assigning functional roles to students in advance (Strijbos & Martens, 2001), by setting clear boundaries in terms of time and number of contributions (Owen, 2000), by providing a tool to support the explicit formulation, representation and testing of hypotheses (Van Bruggen, Kirschner, & Jochems, 2002) and by providing a negotiation tool to support the process of finding common ground in problem-solving groups (Beers, Boshuizen, & Kirschner, 2003). De Wever, Valcke, and Van Winckel (2003) found that adding structure to the discussions led to higher levels of knowledge construction, as measured by the levels of analysis by Gunawardena et al. (1997). Providing cueing to students in advance might also indirectly structure and influence collaboration. For a first indication in this direction, Mevarech and Kramarski (2003) compared WOEs and meta-cognitive questions (MCQ) in written material as instructional techniques to support mathematical problem solving and knowledge construction both during individual study and during small, face-to-face group discussions. They found the complexity of the task and the instructional technique to be important variables in mathematical communication and achievement. During small group discussions about a complex mathematical task, students that had individually received MCQ demonstrated more meta-cognitive questioning and higher-levelled discourse; for a simple task, WOEs yielded better group discussion. This study examines whether automated cueing in a multimedia practical will structure activity and influence the level of small group discussions.

Collaboration and cognitive activity

It has become apparent that characteristics of the task environment influence collaborative knowledge construction activities (e.g. Henri, 1992, 1994), and some researchers have mentioned structure as the key variable to invoke more focused and higher-level cognitive activity. In order to measure increase of the level of cognitive activity by cueing, for example, because leading questions can structure problem solving during small group discussion, ways to analyse cognitive activity must be found first. Concurrent protocols predominantly contain information on actions and concrete products, and to a lesser degree, information on discussions about strategies and tactics, on rules and principles that govern the problem-solving process, and on the monitoring or reflection on the task execution itself (e.g. Carletta et al., 1997). Henri distinguished two types of cognitive activity: (a) implicit interactions ('independent interventions' or 'comments to' pertaining to information that learners put in independent from others,
reflecting low levels of schema elaboration) that reflect a lower level of cognitive activity; and (b) explicit interactions (‘interactive interventions’ or ‘answers to’ pertaining to input from learners that entails the actual comparison of schemas, reflecting high levels of schema elaboration) that reflect a higher level of cognitive activity.

Our hypotheses are that: (1) cueing will increase training and transfer task outcomes, (2) collaboration will further increase training and transfer task outcomes and (3) cueing will indirectly structure and increase the level of cognitive activity during collaboration.

Method

Participants

Participants were 50 junior law students at a Dutch university who volunteered to take part in the experiment, which was organized in the context of the regular court practical. Students, enrolled for this practical, could receive the equivalent of about $250 US for participating. Participants were assigned to three conditions in a randomized controlled trial. During the experiment, four participants dropped out due to study planning problems. A full dataset was obtained from 46 participants (33 female, 13 male; mean age = 21.80 year, SD = 1.78). Comparability of pleading experience was assured by a prior knowledge questionnaire. The overall prior presentation skills were low (M = 3.80, SD = 3.19, on an 18-point scale) and did not differ as a function of experimental condition, F(2, 43) = 0.39, MSE = 10.49, p = .68, η²p = .02.

Learning materials

Two versions of the multimedia practical Preparing a plea (Wöretshöfer et al., 2000) were produced with cueing for both training tasks (cases Bosmans and Ter Zijde) being either present or absent. The cueing in each step of the training task, provided to support individualized learning, consisted of a combination of evaluation criteria and leading questions contained in process worksheets (PW), and accompanying WOE provided by the (virtual) coach in the program. These WOE were based on the same criteria and questions. The practical requires law students to learn and demonstrate the ‘whole task’ of preparing a plea to be held in court. Figure 1 contains excerpts from concrete examples of PW and WOE.

In the ‘no-cueing’ groups (Conditions 1 and 2), participants received global subtask instruction without further cueing. In the cueing group (Condition 3), participants could access available PW and WOE for all steps and cases at any time; however, the filled-in PW (reports) could only be sent in for feedback within the appropriate step. Besides cueing, both versions presented identical support tools, such as a ‘plea checker’ to analyse pleas, discussions of ethical issues in pleading, numerous files and documents and two non-compulsory practice dossiers. The program has an average study load of about 40 hours, and had to be studied as part of the court practical of about 150 hours.

This practical starts with familiarizing its operation and the stepwise procedure. Then students receive two compulsory training tasks (the civil law case Bosmans and the criminal law case Ter Zijde) and two additional cases for extra practice. Training tasks consist of nine steps, but allow students maximal freedom to work through them. The following constituent skills for holding a plea are imparted and combined in these steps: (1) ordering the file of the case, (2) getting acquainted with the file, (3) studying
the file, (4) analysing the pleading situation, (5) determining the strategy for the pleading note (PN) and plea making, (6) writing a PN, (7) transforming the PN into a plea, (8) practising the plea and (9) actually carrying out the plea. At the end of each of the four steps (3) to (6), students are required to send in a report to their (virtual) coach. After approval, they are allowed to proceed to the next step. The last steps are carried out outside the program. For two consecutive steps, the latter always includes cognitive feedback on the former as well as a new task instruction. Each consecutive report thus builds on the previous one. In the Bosmans task, the two subtasks under study were: (a) the construction of a pleading inventory (PI; outcome of Step 3), which is a (more) process-oriented subtask aimed at the selection of juridical arguments for the oral plea; and (b) the construction of a PN (outcome of Step 6), which is a (more) product-oriented subtask aimed at finalizing the written PN.

**Experimental procedure**

At the start of the experiment, participants were informed by a recruitment text, a written instruction and program manual about the study load of the program, required prior knowledge and ICT skills, possible meeting dates and overall planning. They were randomly assigned to one of three conditions and one meeting, and invitations for meetings were sent at least 3 weeks in advance. Learning materials (including the instruction, manual and prior knowledge questionnaire) were sent to the participants’ home addresses. The questionnaire had to be filled in and returned before starting to work on the program.

Participants were allowed 5 weeks to study the practical before they had to sent in their individual PI (Subtask 3 for Bosmans case), and another 2 weeks to send in their
individual PN (Subtask 6 for Bosmans case), averaging a total of about 25 study hours. Participants were urged in the instructions work individually on the program (which was confirmed by comparing reports) and not to discuss anything with fellow students or teachers in order to maintain independence. After the individual report had been received, participants were allowed to attend the meeting and collaborate on this report in a triad of peers. All participants sent in required reports for PI and PN and attended the meeting; consequently, there were six trios to discuss the PI and six trios to discuss the PN (see Figure 2).

Besides the practical, students were assigned to Study 1 of six cases on paper and prepare a plea according to the stepwise procedure at the end of the court practical. While other court practicals that use the programme demand students to carry out a transfer plea about a known case provided by the programme, this could not be organized within this law faculty. It was, however, ensured that all six cases pertained to different law domains from the training task (i.e. civil law) in order to establish far transfer. About another 2 months after the experimental period (meetings about the PN), the court practical ended with students holding their transfer pleas in a real courthouse.

Performance on the PI and PN reports were measured as intermediate learning outcomes, while performance on the transfer plea was taken as a measure of transfer. All reports and videotaped discussions were blindly and independently scored by two raters - law students nearing graduation who had received training on the pleading measurement instruments and coding scheme.

![Figure 2. Outline of experimental procedure. PI = pleading inventory report; PN = pleading note report; coll = collaboration during small group discussions.](image-url)
**Procedure for collaboration**

At the start of each meeting, each triad of peers was read the standardized instruction by one of the experimenters, explaining purpose, set-up and 'rules' for collaboration. Group members were given each other’s individual reports in print to read and compare. These reports were also electronically available on the computer for writing the group report. Their version of the program ran on another computer, slightly modified to enable access to information from previous steps. The general assignment was to reach unanimous agreement and write a group report within the time allowed. Participants were advised to first compare individual reports and to start writing the group report at least a quarter of an hour before deadline, but no extra directives were given and no structure was offered. From instruction to deadline, group members were allowed 1.5 hours for reading, discussion and writing. This period of time was videotaped for each group. In addition, 15 minutes after starting and 15 minutes before ending, peers were informed about the remaining time. To evaluate the meeting, participants individually filled in the predetermined recall questionnaire and engaged in an informal discussion before leaving.

Participants receiving cueing/collaboration received an e-mail containing expert’s WOE directly after the meeting (which concluded the subtask). Participants receiving no cueing/no collaboration had to be controlled for confounding time-on-task effects. They received individual reports from other peers by e-mail, with the request to (individually) adjust their report. To control for time-on-task effects, they were instructed to spend the same amount of time as was granted during meetings, and again send in their adjusted PI and PN.

**Questionnaires and pleading measurement instruments**

The prior knowledge questionnaire (Nadolski, Kirschner, & Van Merriënboer, 2005) pertained to commitment to the field of law (reading legal journals, watching legal programmes), prior presentation skills (both writing and speaking in public, membership of debating club) and ICT skills (familiarity with and attitude towards computers). The recall questionnaire pertained to the way participants experienced the meeting, and (only for Condition 3) the role cueing had played during individual and group work on the report. The items of this recall questionnaire with means and standard deviations are displayed in Table 1.

Specific pleading measurement instruments (see also Nadolski et al., 2005; Hummel et al., 2004, 2006) were used to determine the quality of the PI, PN and transfer plea. One teacher scored the transfer pleas using the plea-checker tool from the program, which consists of nine criteria (such as drawing attention, anchoring the message, consistency and legal correctness). The first two instruments were independently scored by two almost graduated law students on an average of 60 items that pertain to both correctness of legal content and adequateness of presentation. Scores were normalized on 100-point scales. Inter-rater reliability and consistency were assessed using intra-class correlations (ICC) and Cronbach’s alphas. The ICC (3, k) two-way mixed model (Shrout & Fleiss, 1979) for the PI and PN instruments revealed significant average measure reliability (AMR) on absolute agreement of .89 and .78, respectively, with ICC > .70 generally considered to be acceptable (Yaffee, 1998). Cronbach’s alphas for internal consistency of these instruments were .91 and .80. The plea checker was reliable in an earlier study by Hummel et al. (2006).
Table 1. Recall questions after collaboration with means and standard deviations (n = 36 for Items 1 to 5; n = 18 for Items 6.1 to 7.3e)

<table>
<thead>
<tr>
<th>Nr</th>
<th>Question</th>
<th>Scale/options</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How much mental effort did you feel during the group discussion?</td>
<td>Very, very little (1) – Very, very much (9)</td>
<td>3.53</td>
<td>1.44</td>
</tr>
<tr>
<td>2</td>
<td>How motivated were you during the group discussion?</td>
<td>Very, very little (1) – Very, very much (9)</td>
<td>6.25</td>
<td>0.97</td>
</tr>
<tr>
<td>3a</td>
<td>Indicate which statements are true, by dividing 10 points over . . . [a–e]</td>
<td>Discussion took place in a positive atmosphere</td>
<td>2.39</td>
<td>0.78</td>
</tr>
<tr>
<td>3b</td>
<td>. . . . . . . . .</td>
<td>Discussion led to new knowledge and improvement of the report</td>
<td>1.89</td>
<td>0.83</td>
</tr>
<tr>
<td>3c</td>
<td>. . . . . . . . .</td>
<td>I made a substantial contribution to the group report</td>
<td>2.22</td>
<td>0.44</td>
</tr>
<tr>
<td>3d</td>
<td>. . . . . . . . .</td>
<td>I was able to clarify my opinions</td>
<td>1.89</td>
<td>0.60</td>
</tr>
<tr>
<td>3e</td>
<td>. . . . . . . . .</td>
<td>There was considerable mutual misunderstanding and conflict</td>
<td>1.44</td>
<td>1.24</td>
</tr>
<tr>
<td>4</td>
<td>Indicate to which extent the discussion led to new knowledge and improvement of the report</td>
<td>Very little (1)–Very much (5)</td>
<td>3.42</td>
<td>0.77</td>
</tr>
<tr>
<td>5</td>
<td>Which improvements will make the meeting more efficient?</td>
<td>Open question</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6.1</td>
<td>Did you make use of the WOE when writing your individual report?</td>
<td>Very little (1)–Very much (5)</td>
<td>3.22</td>
<td>1.66</td>
</tr>
<tr>
<td>6.2</td>
<td>Did you make use of the WOE when writing the group report?</td>
<td>Very little (1)–Very much (5)</td>
<td>2.83</td>
<td>1.15</td>
</tr>
<tr>
<td>6.3a</td>
<td>Indicate the contribution of WOE on the group discussion, by dividing 10 points over . . . [a–e]</td>
<td>Used while orienting on the task</td>
<td>1.61</td>
<td>1.92</td>
</tr>
<tr>
<td>6.3b</td>
<td>. . . . . . . . .</td>
<td>Used while planning the task</td>
<td>1.17</td>
<td>1.30</td>
</tr>
<tr>
<td>6.3c</td>
<td>. . . . . . . . .</td>
<td>Used while executing (process) the task</td>
<td>2.00</td>
<td>1.82</td>
</tr>
<tr>
<td>6.3d</td>
<td>. . . . . . . . .</td>
<td>Used while finalizing (product) the task</td>
<td>2.94</td>
<td>2.51</td>
</tr>
<tr>
<td>6.3e</td>
<td>. . . . . . . . .</td>
<td>Did not use them</td>
<td>2.28</td>
<td>2.89</td>
</tr>
<tr>
<td>7.1</td>
<td>Did you make use of the leading questions (PW) when writing your individual report?</td>
<td>Very little (1)–Very much (5)</td>
<td>3.00</td>
<td>1.57</td>
</tr>
<tr>
<td>7.2</td>
<td>Did you make use of the leading questions (PW) when writing the group report?</td>
<td>Very little (1)–Very much (5)</td>
<td>1.94</td>
<td>1.40</td>
</tr>
<tr>
<td>7.3a</td>
<td>Indicate the contribution of PW on the group discussion, by dividing 10 points over . . . [a–e]</td>
<td>Used while orienting on the task</td>
<td>0.94</td>
<td>1.35</td>
</tr>
<tr>
<td>7.3b</td>
<td>. . . . . . . . .</td>
<td>Used while planning the task</td>
<td>1.11</td>
<td>1.45</td>
</tr>
<tr>
<td>7.3c</td>
<td>. . . . . . . . .</td>
<td>Used while executing (process) the task</td>
<td>3.67</td>
<td>2.50</td>
</tr>
<tr>
<td>7.3d</td>
<td>. . . . . . . . .</td>
<td>Used while finalizing (product) the task</td>
<td>0.72</td>
<td>1.02</td>
</tr>
<tr>
<td>7.3e</td>
<td>. . . . . . . . .</td>
<td>Did not use them</td>
<td>3.56</td>
<td>3.18</td>
</tr>
</tbody>
</table>
Participants were asked to score the perceived amount of mental effort, both during individual study and collaboration on the subtask, on an adapted version of the 9-point scale developed by Paas (1992; see also Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Extra time on task spent outside the programme while constructing the individual report for the subtask (M = 60.54, SD = 47.58, in minutes), together with relevant scores on the prior knowledge questionnaire, was taken to assess motivation (on a 12-point scale).

**Coding scheme**

Complex problem-solving processes are typically hard to observe because they take place ‘in the solver’s head’, and quite frequently, the only external evidence is the final solution reported. The coding scheme for analysing cognitive activity during group work had to meet certain requirements (see e.g. Veldhuis-Diermanse, 2002). First of all, categories had to be based on our theoretical orientation (schema-based learning) and research questions, and therefore represent relevant types of cognitive activity. Second, categories should be based on the subtasks and content domain (i.e. the domain of civil law) that guide this study. They should reflect the message content and contain prototype examples from these subtasks for each category of the coding grid. Last but not least, categories must be semantically meaningful, mutually exclusive, all encompassing and scored reliably.

Taylor and Dionne (2000) stressed that content analysis should also access the strategies used in the problem-solving process, as well as the principles and conditions under which a strategy is useful. Recently, Van Gog, Paas, Van Merriënboer, and Witte (2005), studying trouble shooting tasks with malfunctioning electrical circuits, constructed a coding scheme based on four main types of cognitive activity: action, how, why and meta, which are inspired by this new approach. Apart from actions, they distinguish strategic discussions that result in actions (‘how’ information), principled discussions behind the strategies (‘why’ information) and monitoring of the problem-solving process (‘meta’ information).

We adopted these four main categories and extended each with a process-oriented and product-oriented subcategory to fit our research objective. Cognitive activity is characterized as more process-oriented when aimed at orientation, investigating, clarifying possible solutions to the problem (e.g. what information could be used in our report or which arguments are valid for this case). Cognitive activity is characterized as more product-oriented when aimed at finalizing or refining chosen solutions (e.g. how are we going to use this argument in our report or which steps are yet to be taken to draw up the report). Besides these task-valid subcategories, four task-irrelevant subcategories were added. For each (sub)category, a description of typical activities was added. After the video-recordings had become available, some prototypical examples from discussions on the Bosmans case were added to facilitate raters in using the coding scheme, which is presented in Table 2.

The portion of discussion was taken as measure for the level of cognitive activity, in line with the approach taken by Garafalo and Lester (1985). Cognitive behaviour is defined as information-processing actions, when it deals with activities as reading, writing or giving final solutions. Only when students are really engaged in discussions about the problem, and their comments could be heard, is behaviour considered to indicate meta-cognitive activity.
Table 2. Coding scheme for cognitive activity during collaboration

<table>
<thead>
<tr>
<th>Nr</th>
<th>Main category</th>
<th>Subcategory</th>
<th>Activity related to</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Product</td>
<td>Action</td>
<td>Executing actions: applying information, writing, dictating, editing...</td>
<td>How are we going to phrase this argument? We should place the most important argument first in the list. Let's delete that sentence anyway.</td>
</tr>
<tr>
<td>2</td>
<td>Process</td>
<td>Action</td>
<td>Preparatory actions: searching information, reading aloud, selecting usable information...</td>
<td>Reading aloud the exact text of the demanding party (what exactly is claimed here). What was the story behind the insurance?</td>
</tr>
<tr>
<td>3</td>
<td>Product</td>
<td>How</td>
<td>Discussing the chosen strategies or tactics, for example, how to apply the solution of WOBs in the report...</td>
<td>Are we claiming or disputing the contract? What is primary, subsidiary? Are we going to use liability?</td>
</tr>
<tr>
<td>4</td>
<td>Process</td>
<td>How</td>
<td>Discussing possible approaches or heuristics for report, for example, by examining case law, consulting experts, or applying reading questions and criteria...</td>
<td>Can we use Article 717 Sub 4 as an exception to non-conformity? Does plaintiff claim miscarriage? Should we include the meaning of opposing party in this argumentation?</td>
</tr>
</tbody>
</table>

Prototype examples Bosmans case:
- How are we going to phrase this argument?
- We should place the most important argument first in the list. Let's delete that sentence anyway.
- Reading aloud the exact text of the demanding party (what exactly is claimed here).
- What was the story behind the insurance?
- Are we claiming or disputing the contract?
- What is primary, subsidiary?
- Are we going to use liability?
- Are we going to charge the process costs?
- We better combine a neutral plea with emotions, but only when relevant.
- Can we use Article 717 Sub 4 as an exception to non-conformity?
- Does plaintiff claim miscarriage?
- Should we include the meaning of opposing party in this argumentation?
- Should we speak about mutual miscarriage?
<table>
<thead>
<tr>
<th>Nr</th>
<th>Main category</th>
<th>Subcategory</th>
<th>Activity related to . . .</th>
<th>Prototype examples Bosmans case</th>
</tr>
</thead>
</table>
| 5  | Product       | Why         | Discussing juridical principles, rules and facts behind the *chosen* solution | – Mentioning default is redundant here.  
– If Article 218 Sub c, then we refer to 6.230  
– Are we addressing this issue in a relational or more objective tone?  
– We should restrict to sub c, because . . .  
– Which facts are still missing? |
| 6  | Process       | Why         | Discussing juridical principles, rules and facts behind *possible* solutions | – Does default apply here?  
– Now, what exactly is the juridical question?  
– Is there a principal difference between making one or two test drives?  
– What is the technical state of the Honda?  
– Does a duty of giving notice apply here? |
| 7  | Product       | Meta        | Orientation, monitoring and evaluating chosen solution | – Let’s leave those headings bold-faced . . .  
– What should happen with this report?  
– Do you still think this sums it up well? |
| 8  | Process       | Meta        | Orientation, monitoring and evaluating the collaboration | – Is everybody satisfied?  
– We should start dividing tasks.  
– Let’s first have a look at what everybody has as extras. |
| 9–12 | Task-irrelevant | [various] | Praise/complaints about program or meeting. Reading or writing individually. Fragments that cannot be scored | – How irritating that you cannot scroll through or print those documents  
– Replacing the computer or flap-over.  
– Audio fragment is not audible (bad quality of recording). |
For the actual coding of the discussions on the predetermined categories, a method of time sampling was applied, scoring the type of cognitive activity on every exact minute. Videotapes displayed a uniform time code in the upper left hand corner of the screen. Inter-rater reliability of the (first time use) coding scheme was assessed (with \( k = 2 \)) and appeared to be (very) satisfactory both on the level of the five main categories (\( K = .87, N = 1.080 \)) and the 12 subcategories (\( K = .85, N = 1.080 \)). Leaving out the proportion (27.5%) of task-irrelevant behaviour (Subcategories 9 to 12), these measures were even a little higher both on the level of the four main categories (\( K = .89, N = 758 \)) and eight subcategories (\( K = .89, N = 758 \)). In qualitative analysis, a Cohen’s kappa between .81 and 1.00 is considered ‘almost perfect’ (Heuvelmans & Sanders, 1993, p. 450).

**Experimental design**

Participants in the cueing/collaboration condition \( (N = 18) \) received individual training through a version of *Preparing a plea* with cueing, and additional collaboration on one of the subtasks under study. In the no cueing/collaboration condition, participants \( (N = 18) \) received a version of the program without cueing, but with the additional collaboration. In the third no cueing/no collaboration (control) condition, participants \( (N = 10) \) received neither cueing nor collaboration.

We applied a between-groups design, inviting half of the participants \( (N = 18), \) equally divided over the experimental Conditions 2 and 3, to attend a meeting on the PI, and the other half to attend a meeting on the PN about 2 weeks later (see Figure 2 for a graphical display of this procedure). Participants \( (N = 36) \) in these experimental conditions were randomly assigned to a triad of peers within the same condition.

**Results**

Repeated ANOVA was applied on the general outcomes, using time of measurement (before or after collaboration) as a within-subjects factor and experimental condition (either cueing/collaboration, no cueing/collaboration or no cueing/no collaboration) as the between-subjects factor. ANOVAs were applied with experimental condition as between-subject factors, and with various learning outcomes (general outcomes before and after collaboration, PI and PN scores before and after collaboration, and transfer plea scores), scores on the items of the recall questionnaire, motivation, mental effort and time-on-task scores as dependent variables. The partial eta squared statistic was used as an effect size index where values of .01, .06 and .14 correspond to small, medium and large values, respectively (Cohen, 1988). Coding scores from small group discussions during collaboration were analysed with Mann-Whitney tests with the level and types of cognitive activity as dependent variables. Finally, independent \( t \) tests were used to compare learning growth differences between experimental conditions.

**Learning outcomes before and after collaboration**

All learning outcomes before and after collaboration are summarized in Tables 3A and 3B. A repeated measures ANOVA revealed main effects for time of measurement, \( F(1, 44) = 38.36, MSE = 408.71, p < .001, \eta^2_p = .47 \) and experimental condition, \( F(2, 43) = 3.51, MSE = 408.71, p < .05, \eta^2_p = .13 \), but no interaction effect, \( F(2, 43) = 1.62, MSE = 23.66, p = .21, \eta^2_p = .07 \). The intermediate learning outcomes on PI and PN after making adjustments (either during collaboration or individually) were
significantly better than those before for all three conditions. To establish general outcomes before and after adjustment for all participants, both individual reports (PI before or PN before) and group reports (PI after or PN after) were used for participants in Conditions 2 and 3 (receiving collaboration), and average scores on both reports (PI and PN before, PI and PN after) were used for participants in Condition 1 (not receiving collaboration).

**Pleading inventory and pleading note scores.** There was a main effect of cueing on both PI and PN scores before. A one-way ANOVA showed that participants receiving cueing outperformed those that did not on the PI, $F(1, 26) = 9.80, p < .01$, and PN scores, $F(1, 26) = 26.66, p < .001$. There was also a main effect of collaboration on the PI and PN scores after. A one-way ANOVA showed that participants who collaborated delivered better PI, $F(1, 26) = 5.98, p < .05$, and PN, $F(1, 26) = 45.68, p < .001$, than participants that had to adjust the reports individually.

**General outcomes.** A main effect of cueing on all general outcomes before was found, $F(1, 44) = 5.86, \text{MSE} = 248.29, p < .05, \eta^2_p = .12$. This effect could be confirmed by a contrast test using Bonferroni correction that revealed better results for participants in the cueing condition when compared with both no cueing conditions taken together, $t(43) = 2.50, p < .01$, one-tailed. A main effect of collaboration was found on the general outcomes after, $F(1, 44) = 4.79, \text{MSE} = 184.41, p < .05, \eta^2_p = .10$. An interaction effect of cueing and collaboration was found on the general outcomes after, $F(2, 43) = 3.29, \text{MSE} = 181.44, p < .05, \eta^2_p = .13$, but not on the increase (growth) in learning outcome, $F(2, 43) = 1.30, \text{MSE} = 44.41, p = .28$.

### Table 3. (A and B). Learning outcomes (normalized to 100-point scales) for between-groups design

A. Scores on pleading inventory (PI) and pleading note (PN) subtasks before and after collaboration ($n = 28$)

<table>
<thead>
<tr>
<th>Condition</th>
<th>PI before</th>
<th>No cueing</th>
<th>No cueing/no collaboration</th>
<th>All</th>
<th>PN before</th>
<th>No cueing</th>
<th>No cueing/no collaboration</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI before</td>
<td>38.67</td>
<td>30.67</td>
<td>30.30</td>
<td>33.11</td>
<td>71.67</td>
<td>59.89</td>
<td>53.90</td>
<td>61.54</td>
</tr>
<tr>
<td>PI after</td>
<td>47.22</td>
<td>41.22</td>
<td>36.70</td>
<td>41.54</td>
<td>72.89</td>
<td>67.11</td>
<td>56.10</td>
<td>65.04</td>
</tr>
<tr>
<td>PN before</td>
<td>71.67</td>
<td>59.89</td>
<td>53.90</td>
<td>61.54</td>
<td>72.89</td>
<td>67.11</td>
<td>56.10</td>
<td>65.04</td>
</tr>
<tr>
<td>PN after</td>
<td>72.89</td>
<td>67.11</td>
<td>61.54</td>
<td>65.04</td>
<td>72.89</td>
<td>67.11</td>
<td>61.54</td>
<td>65.04</td>
</tr>
</tbody>
</table>

B. General outcomes before and after collaboration, learning growth and transfer plea scores ($N = 46$)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Outcome before</th>
<th>No cueing</th>
<th>No cueing/no collaboration</th>
<th>All</th>
<th>Growth (delta)</th>
<th>Transfer plea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome before</td>
<td>55.17</td>
<td>45.28</td>
<td>40.70</td>
<td>48.15</td>
<td>5.09</td>
<td>70.33</td>
</tr>
<tr>
<td>Outcome after</td>
<td>60.26</td>
<td>54.17</td>
<td>46.50</td>
<td>54.80</td>
<td>6.92</td>
<td>77.39</td>
</tr>
<tr>
<td>Growth (delta)</td>
<td>5.09</td>
<td>8.89</td>
<td>5.80</td>
<td>6.91</td>
<td>6.92</td>
<td>9.34</td>
</tr>
<tr>
<td>Transfer plea</td>
<td>70.33</td>
<td>77.39</td>
<td>67.20</td>
<td>72.41</td>
<td>6.92</td>
<td>9.34</td>
</tr>
</tbody>
</table>
General outcomes before and after appear to differ significantly, \( t(45) = -6.47, p < .001 \). Finally, we noted that the relative increase in learning outcome (growth) was highest for participants receiving no cueing/collaboration (Condition 2). However, independent \( t \) test comparisons of Condition 3 with Condition 2, \( t(34) = 1.43, p = 0.08 \), one-tailed, and Condition 2 with Condition 1, \( t(26) = 1.21, p = .11 \), one-tailed, only approach significance. Only a minority of five participants (of which three in Condition 3, one in Condition 2 and one in Condition 1) suffered negative learning growth on their PI or PN outcome, but decreases were small (averaging about 4 points on a 100-point scale).

**Group discussion**

Table 4 shows the aggregated results from the coding schemes on task-valid subcategories, expressed as percentages of the total number of scored items. This table also presents the portion of discussion (Subcategories 3 to 8) as measure of cognitive level. The expected main effect of cueing on the level of cognitive activity was not found. Contrary to our expectation, participants in the no cueing condition (six triads) demonstrated the highest level of cognitive activity during group discussion (\( U = 5.00, p < .05 \)). Two types of cognitive activity differed between conditions: ‘cued’ participants (six triads also) demonstrated more behaviour in Category 1: action/product (\( U = 4.50, p < .05 \)), and less in Category 3: how/product (\( U = 5.00, p < .05 \)). There were no differences on the other six categories (all \( p > .4 \)).

**Recall questionnaire**

Scores on the recall questionnaires give insight into personal *perception* of collaboration and the effect of cueing on this perception. Results show that participants felt highly motivated (\( M = 6.25, SD = .97 \), on a 9-point scale), and little mental effort was required (\( M = 3.53, SD = 1.44 \), on a 9-point scale) during collaboration. Paired \( t \) tests that compare motivation and mental effort scores during the meeting with the same scores
while individually studying the subtasks (**M** = 3.80, **SD** = 1.70 and **M** = 5.50, **SD** = 0.91, respectively) reveal strong differences, \( t(35) = 7.05, p < .001 \) and \( t(35) = -7.03, p < .001 \), respectively. Participants receiving no cueing (**M** = 6.72, **SD** = 0.83) appear most motivated during collaboration when compared with participants receiving cueing (**M** = 5.78, **SD** = 0.88; **F**(1, 34) = 11.04, **MSE** = 0.73, \( p < .01, \eta^2_p = .25 \). The perceived amount of learning increase through collaboration (**M** = 3.42, **SD** = 0.77, on a 6-point scale) could not be attributed to cueing, **F**(1, 32) = 1.12, **MSE** = 0.62, \( p = .30, \eta^2_p = .03 \). Table 2 presents results on all items of the recall questionnaire.

As expected, participants that receive cueing (**N** = 18) value leading questions (PW) more for discussing PI than for discussing PN, **F**(1, 6) = 9.78, **MSE** = 1.28, \( p < .01, \eta^2_p = .38 \); Item 7.2), and value WOE more while executing the subtask PN than while executing the subtask PI, **F**(1, 16) = 5.45, **MSE** = 2.61, \( p < .05, \eta^2_p = .25 \); Item 6.3c.

**Transfer**

An ANOVA of the transfer performance data reveals an unexpected (negative) main effect of cueing on transfer plea scores, **F**(1, 44) = 4.79, **MSE** = 93.63, \( p < .05, \eta^2_p = .10 \). The expected (positive) main effect of collaboration on the transfer plea scores, **F**(1, 44) = 7.13, **MSE** = 93.63, \( p < .05, \eta^2_p = .14 \) was also found.

**Time on task, mental effort and motivation**

An ANOVA of the motivation scores during individual study for participants receiving cueing/collaboration (**M** = 3.39, **SD** = 1.61, **N** = 18), no cueing/collaboration (**M** = 4.44, **SD** = 1.92, **N** = 18) and no cueing/no collaboration (**M** = 3.40, **SD** = 1.07, **N** = 10) reveals no differences as a function of condition, **F**(2, 43) = 2.22, **MSE** = 2.72, \( p = .12, \eta^2_p = .09 \). Average mental effort scores during individual study for these groups (**M** = 5.22, **SD** = 0.88; **M** = 5.89, **SD** = 0.90; and **M** = 5.30, **SD** = 0.82, respectively, do not differ as a function of condition, **F**(2, 43) = 2.94, **MSE** = 0.77, \( p = .06, \eta^2_p = .12 \). Finally, average time-on-task scores on the subtask (**M** = 168.06, **SD** = 63.78; **M** = 204.17, **SD** = 85.03; and **M** = 183.50, **SD** = 38.15, respectively, all in minutes) do not differ as a function of condition, **F**(2, 43) = 1.24, **MSE** = 4.771.44, \( p = .30, \eta^2_p = .05 \).

**Discussion**

Both the first hypothesis that cueing would increase performance and the second hypothesis that collaboration would increase performance could be partially confirmed. Results show that cueing improves the quality of PIs and PNs, replicating earlier findings by Hummel et al. (2004, 2006), and that collaboration further improves these reports. A comparison of general outcomes reveals main effects and an interaction effect for cueing and collaboration. Transfer measures on closing pleas revealed the expected positive effect of collaboration but not for cueing. The third hypothesis that cueing would increase the level of cognitive activity during collaboration was rejected. Results show that the level of cognitive activity and the amount of strategic discussion are higher for ‘not cued’ participants.

The interaction effect of cueing and collaboration indicates that both work together in increasing performance. The effect of collaboration increases when less cueing is provided and decreases when more cueing is provided. This explains why ‘not cued’ participants appear to benefit more from collaboration and also feel more
motivated during collaboration. ‘Not cued’ participants still receive a lot of new
information during collaboration (through peer feedback); they still have a lot of
‘choosing and planning’ (Garafolo & Lester, 1985) to catch up on. ‘Cued’ participants
have already received some of this information through PW and WOE in the program.
Phrased in schema-based learning terminology, one could state that the schemas of
the ‘not cued’ participants are still more ‘under construction’, needing a higher level
of schema elaboration and monitoring (Henri, 1992, 1994). ‘Cued’ participants, who
had received more strategic and principled cues before collaboration (from PW and
WOE), are left with ‘merely doing’ (low level of schema elaboration and monitoring)
and simply do not have that much to gain from each other any more. These results
give reason to believe that ‘students-support-each other’ is indeed a feasible option to
be combined with or (partially) substitute cueing when training complex learning
tasks.

Possible directions for future research emerge from this study. First, it would be
interesting to conduct studies to compare the benefits of face-to-face collaboration (as in
this study) with computer-supported collaboration (as in most concurrent CSCL/CSCW
research). CSCL might be less powerful (e.g. because it lacks direct and non-verbal
interaction), but can also be more feasible (less demanding to attend and more flexible
to organize).

Second, these findings should be extended to domains that share the same type of
problem-solving ontology as for law (i.e. one based on heuristic rules and strategic
approaches, rather than on strict algorithms, rules or procedures), such as the social
sciences.

Third, further experimentation on schema-based learning should and can be carried
out in the context of complex, ecologically valid, authentic training programs of longer
duration. The current study demonstrates that it is feasible to combine full experimental
control (especially on cueing and collaboration) with authentic contexts of study.
However, due to ethical considerations, differences between experimental conditions –
and consequently the effects – might have to be reduced. Even with the lack of cueing or
collaboration, some support mechanisms in the program still guaranteed that
participants, that were regular students working for credits, could still successfully
study. Inclusion of a very poor condition without support would most likely have
induced stronger effects of additional cueing and collaboration, but this was not a
realistic option here. Furthermore, although participants were urged to work
individually at home and not to discuss anything with fellow students or teachers
during the experimental period in order to maintain independence, it was impossible
for us to fully control this, which may have limited the study. Future research in an
authentic context should seek to find ways to warrant this control, for instance, by using
the ‘diary method’ (Bolger, Davis, & Rafaeli, 2003). This method provides the field of
educational psychology with ways to collect information, complementary to more
traditional designs, on study processes within everyday learning programs of longer
duration.

Fourth, the optimal balance between individual and collaborative support in
training complex problem-solving tasks should be further examined and determined.
What information can best be provided by individual cueing? Which information can
best be discussed collaboratively? It might, for instance, be more cost effective to
develop multimedia practicals if some cueing could be left to peer feedback, and at
the same time would address the teacher bandwidth problem. What would be the
optimal amount of time for both? In this study, participants spend an average time of
about 4 hours on each subtask during individual study ($M = 168.06$, $SD = 63.78$, in minutes; with some extra time outside the program; $M = 60.54$, $SD = 47.58$, in minutes), and were allowed 1.5 hours for the group discussion. Some did complain (Question 5) that time for discussion was too short, and some groups did not finish their report.

Finally, what has to be the optimal amount of structure for a CPS meeting? This study indicates that a clear purpose might be sufficient to enable efficient collaboration in small groups, and that peers do not always need more structure or ‘collaborative tools’. Although some participants did complain that no tutor was available to provide expert feedback (Question 5), it was fascinating to observe from the activities and outcomes of the group discussions that CPS can indeed simply emerge without any guiding authority. Future research should continue to examine the optimal amount of collaborative structure and its practical implications.

References


Received 29 July 2004; revised version received 5 February 2005