

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

## Citation for published version (APA):

Hummel, H., Paas, F., & Koper, R. (2006). Effects of cueing and collaboration on the acquisition of complex legal skills. *British Journal of Educational Psychology*, 76(3), 613-631.  
<https://doi.org/10.1348/000709905X85629>

## DOI:

[10.1348/000709905X85629](https://doi.org/10.1348/000709905X85629)

## Document status and date:

Published: 01/09/2006

## Document Version:

Peer reviewed version

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

## Take down policy

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 12 Dec. 2021

Open Universiteit  
[www.ou.nl](http://www.ou.nl)



Running head: COLLABORATION AND CUEING

Effects of Cueing and Collaboration on the Acquisition of Complex Legal Skills

Hans G. K. Hummel

Fred Paas

E. J. R. Koper

Open University of the Netherlands

Submitted: July 29, 2004

Resubmitted: November 30, 2004

Revised: February 9, 2005

Author note

The authors would like to thank the law students and staff of the court practical at the Free University of Amsterdam in The Netherlands for participating in this study.

Correspondence concerning this article should be addressed to Hans G. K. Hummel, Educational Technology Expertise Centre, Open University of the Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands, E-mail: [hans.hummel@ou.nl](mailto:hans.hummel@ou.nl), Fax: +31 45 5762 907.

Keywords: cueing, peer feedback, collaboration, schemas, complex problem-solving tasks

### Abstract

This study examines the effects of cueing and collaboration on training outcomes and transfer pleas, and on cognitive activity during collaboration, by combining a multimedia practical with cueing and small-group collaboration with peer feedback to support the complex task of preparing a plea in court. Results reveal that 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. The theoretical and practical implications for combining cueing and small-group collaboration to support the acquisition of complex skills are discussed.

## Effects of Cueing and Collaboration on the Acquisition of Complex Legal Skills

Distance education and lifelong learning call for individualised 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 called ‘teacher bandwidth problem’ (Wiley & Edwards, 2003, p. 2). *Cueing* is one of these instructional techniques, that is defined by Hummel and Nadolski (2002) as a technique that facilitates cognitive processes that enable problem-solving transfer, i.e. 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 (see e.g., Renkl, 2002) and process worksheets with leading questions (see e.g., 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, in press).

Although multimedia practicals with cueing offer powerful individualised learning environments and decrease the working expenses of education, they are at the same time 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 e.g., 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 criticise (Jeong & Chi, 2000). Wiley and Edwards (2003) investigated the potential of Online 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, by which means and to which extent 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 operationalised 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 worked-out examples 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 worked-out examples 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, e.g., 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 (1992, 1994) 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*

Fifty junior law students at a Dutch university volunteered to participate in the experiment, which was organised 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 randomised 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 a 18-point scale) and did not differ as a function of experimental condition ( $F(2, 43) = 0.39$ ,  $MSE = 10.49$ ,  $p = .68$ ,  $\eta_p^2 = .02$ ).

### *Learning materials*

Two versions of the multimedia practical *Preparing a plea* (Wöretshofer 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 individualised learning, consisted of a combination of evaluation criteria and leading questions contained in process worksheets (PW), and accompanying worked-out examples (WOE) provided by the (virtual) coach in the programme. 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.

-----  
INSERT FIGURE 1 ABOUT HERE  
-----



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; the filled-in PW (reports) could however only be sent in for feedback within the appropriate step. Besides cueing, both versions presented identical support tools, like a 'plea checker' to analyse pleas, discussions of ethical issues in pleading, numerous files and documents, and two non-compulsory practice dossiers. The programme 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 familiarising 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. Following constituent skills for holding a plea are trained 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 and plea making; (6) writing a pleading note; (7) transforming the pleading note into a plea; (8) practicing 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 programme. 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. Both within the 'Bosmans' task, the two subtasks under study were: a. the construction of a pleading inventory (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 pleading note (outcome of step 6), which is a (more) product-oriented subtask aimed at finalising the written pleading note.

*Experimental procedure*

At the start of the experiment, participants were informed by a recruitment text, a written instruction and programme manual about the study load of the programme, 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 three 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 programme.

Participants were allowed five weeks to study the practical before they had to send in their individual pleading inventory (subtask 3 for 'Bosmans' case), and another two weeks to send in their individual pleading note (subtask 6 for 'Bosmans' case), averaging a total of about 25 study hours. Participants were urged, by the instruction, and controlled, by comparing reports, to work individually on the programme 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 pleading inventory and pleading note and attended the meeting; consequently, there were six trios to discuss the pleading inventory and six trios to discuss the pleading note (see Figure 2).

-----  
INSERT FIGURE 2 ABOUT HERE  
-----

Besides the practical, students were assigned to study one 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 organised within this law faculty. It was

however controlled 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 two months after the experimental period (meetings about the pleading note), the court practical ended with students holding their transfer pleas in a real courthouse.

Performance on the pleading inventory and pleading note reports were measured as intermediate learning outcomes; 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, who were almost graduated law students that received a short training on the pleading measurement instruments and coding scheme.

#### *Procedure for collaboration*

At the start of each meeting, each triad of peers was read the standardised 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 programme 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 furthermore no extra directives were given and no structure was offered. From instruction to deadline, group members were allowed one-and-a-half hour for reading, discussion and writing. This period of time was videotaped for each group. Fifteen minutes after starting and fifteen 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 worked-out example 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 pleading inventory and pleading note.

#### *Questionnaires and Pleading measurement instruments*

The prior knowledge questionnaire (Nadolski, Kirschner, & Van Merriënboer, in press) 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.

-----  
INSERT TABLE 1 ABOUT HERE  
-----

Specific pleading measurement instruments (see also Nadolski, Kirschner, & Van Merriënboer, in press; Hummel, Paas, & Koper, 2004, in press) were used to determine the quality of the pleading inventory (PI), pleading note (PN) and transfer plea. One teacher scored the transfer pleas using the ‘plea-checker’ tool from the programme, which consists of nine criteria (like 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 sixty items that pertain to both correctness of legal content and

adequateness of presentation. Scores were normalised 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 AMR (average measure reliability) 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, Paas, and Koper (in press).

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 min), 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 solvers 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 (2004), 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 characterised 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 characterised as more product-oriented when aimed at finalising 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.

-----  
 INSERT TABLE 2 ABOUT HERE  
 -----

The portion of discussion was taken as measure for the *level* of cognitive activity, in line with the approach taken by Garafolo 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, behaviour is considered to indicate meta-cognitive activity.

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. Video tapes 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 twelve subcategories ( $K = .85$ ,  $N = 1,080$ ). Leaving out the proportion (27.5%) of task-irrelevant behaviour (subcategories 9-12), 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 programme 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 pleading inventory, and the other half to attend a meeting on the pleading note about two 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 analysis of variance (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. ANOVA were applied with experimental condition as between-subject factors, and with various learning outcomes (general outcomes before and after collaboration, pleading inventory and pleading note 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 summarised in Tables 3A and 3B. A repeated measures ANOVA revealed main effects for time of measurement  $F(1, 44) = 38.36$ ,  $MSE = 408.71$ ,  $p < .001$ ,  $\eta_p^2 = .47$ ) and experimental condition ( $F(2, 43) = 3.31$ ,  $MSE = 408.71$ ,  $p < .05$ ,  $\eta_p^2 = .13$ ), but no interaction effect ( $F(2, 43) = 1.62$ ,  $MSE = 23.66$ ,  $p = .21$ ,  $\eta_p^2 = .07$ ). The intermediate learning outcomes on pleading inventory (PI) and pleading note (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. One-way ANOVA show that participants receiving cueing outperform 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. One-way ANOVA show that participants who collaborated finally 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.

-----  
 INSERT TABLE 3A ABOUT HERE  
 -----

*General outcomes.* A main effect of cueing on all general outcomes before was found ( $F(1, 44) = 5.86, MSE = 248.29, p < .05, \eta_p^2 = .12$ ). This effect could be confirmed by a contrast test using Bonferroni correction, that revealed better results for participants in the ‘cueing’ condition when compared to 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 ( $F(1, 44) = 4.79, MSE = 184.41, p < .05, \eta_p^2 = .10$ ). An interaction effect of cueing and collaboration was found on the general outcomes after ( $F(2, 43) = 3.29, MSE = 181.44, p < .05, \eta_p^2 = .13$ ), but not on the increase (growth) in learning outcome ( $F(2, 43) = 1.30, MSE = 44.41, p = .28, \eta_p^2 = .06$ ). 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 conditions 3 with 2 ( $t(34) = 1.43, p = 0.08$ , one-tailed) and 2 with 1 ( $t(26) = 1.21, p = .11$ , one-tailed) only approach significance. Only a

minority of 5 participants (of which three in condition 3, one in condition 2, and one in condition 1) suffered negative learning growth on their pleading inventory or pleading note outcome, but decreases were small (averaging about four points on a 100-point scale).

-----  
 INSERT TABLE 3B ABOUT HERE  
 -----

### *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 3-8) 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  $p > .4$ ).

-----  
 INSERT TABLE 4 ABOUT HERE  
 -----

### *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 nine-point-scale), and little mental effort was required ( $M = 3.53, SD = 1.44$ , on a nine-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  $M = 5.50, SD = .91$  respectively) reveal strong differences ( $t(35) = 7.05, p < .001$  and  $t(35) = -7.03, p < .001$  respectively). Participants receiving ‘no cueing’ ( $M = 6.72, SD = .83$ ) appear most motivated during collaboration when compared to participants receiving ‘cueing’ ( $M = 5.78, SD = .88; F(1, 34) = 11.04, MSE = .73, p < .01, \eta_p^2 = .25$ ). The *perceived* amount of learning increase through collaboration ( $M = 3.42, SD = .77$ , on a six-point scale) could not be attributed to cueing ( $F(1, 32) = 1.12, MSE = .62, p = .30, \eta_p^2 = .03$ ). 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 ( $F(1, 16) = 9.78, MSE = 1.28, p < .01, \eta_p^2 = .38$ ; item 7.2), and value Worked-Out Examples (WOE) more while executing the subtask PN than while executing the subtask PI ( $F(1, 16) = 5.45, MSE = 2.61, p < .05, \eta_p^2 = .25$ ; item 6.3c).

### *Transfer*

ANOVA of the transfer performance data reveals an unexpected (negative) main effect of cueing on transfer plea scores ( $F(1, 44) = 4.79, MSE = 93.63, p < .05, \eta_p^2 = .10$ ). The expected (positive) main effect of collaboration on the transfer plea scores ( $F(1, 44) = 7.13, MSE = 93.63, p < .05, \eta_p^2 = .14$ ) 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 ( $F(2, 43) = 2.22, MSE = 2.72, p = .12, \eta_p^2 = .09$ ). Average mental effort scores during individual study for these groups ( $M = 5.22, SD = .88; M = 5.89, SD = .90$ ; and  $M = 5.30, SD = .82$  respectively) do not differ as a function of

condition ( $F(2, 43) = 2.94, MSE = .77, p = .06, \eta_p^2 = .12$ ). Finally, average time-on-task scores on the subtask ( $M = 168.06, SD = 63.78; M = 204.17, SD = 85.03; \text{ and } M = 183.50, SD = 38.15$  respectively, all in minutes) do not differ as a function of condition ( $F(2, 43) = 1.24, MSE = 4,771.44, p = .30, \eta_p^2 = .05$ ).

## 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 pleading inventories and pleading notes, replicating earlier findings by Hummel et al. (2004, in press), 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 are still to 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 programme. 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 anymore. 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 organise).

*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), like 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 programmes 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 programme 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 and controlled 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 is challenged 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 programmes 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 four hours on each subtask during individual study ( $M = 168.06$ ,  $SD = 63.78$ , in minutes; with some extra time outside the programme ( $M = 60.54$ ,  $SD = 47.58$ , in minutes)), and were allowed one-and-a-half hour 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 collaborative problem-solving (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

- Barlow, C, Phelan, A., Harasym, P, & Myrick, F. (2004). Peer Collaboration as a Model for Workplace Learning in Health care: Possibilities and Challenges. [Retrieved online on November 23, 2004 from: <http://www.wln.ualberta.ca/papers/pdf/03.pdf>]
- Beers, P. J., Boshuizen, H. P. A., & Kirschner, P. A. (2003). Agreeing to Disagree: Perspective, Negotiation and Common ground. In P. A. Kirschner (Chair), *The Social psychological Dimension of Social interaction and the Effects of Cultural Backgrounds in CSCL*. Symposium conducted at the 10th EARLI Conference, Padova, Italy.
- Bolger, N., Davis, A, & Rafaeli, E. (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology*, 54, 579-616.
- Carletta, J., Isard, A., Kowto, J.C., Doherty-Sneddon. G., & Anderson, A.H. (1997). The Reliability of Dialogue Structure Coding Scheme. *Computational Linguistics*, 23, 13-31.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Garafalo, J., & Lester, F.K. (1985). Metacognition, Cognitive Monitoring, and Mathematical Performance. *Journal for Research in Mathematics Education*, 16, 163-176.
- Gunawardena, C.N., Carabajal, K., & Lowe, C.A. (2001). *Critical Analysis of Models and Methods used to evaluate Online Learning Networks*. Seattle: AERA.
- Gunawardena, C.N., Lowe, C.A. & Anderson, T (1997). Analysis of a Global Online Debate and the Development of an Interaction Analysis Model for Examining Social Construction of Knowledge in Computer Conferencing. *Journal of Educational Computing Research*, 17(4), 395-429.

- Henri, F. (1992). Computer-conferencing and Content Analysis. In A.R. Kaye (Ed.), *Collaborative Learning through Computer Conferencing: The Najaden papers* (pp. 117-136). London: Springer-Verlag.
- Henri, F. (1994). Distance Learning and Computer-Mediated Communication: Interactive, Quasi-interactive or Monologue? In C. O'Malley (Ed.), *Computer Supported Collaborative learning* (pp. 145-161). London: Springer-Verlag.
- Heuvelmans, A.P.J.M., & Sanders, P.F. (1993). Beoordelaarsovereenstemming [Inter-rater reliability]. In T. Eggen & P. Sanders (Eds.), *Psychometrie in de praktijk* [Psychometry in practice] (pp. 443-471). Arnhem, The Netherlands: CITO.
- Hummel, H. G. K., & Nadolski, R.J. (2002). Cueing for Schema Construction: Designing Problem-solving Multimedia Practicals. *Contemporary Educational Psychology*, 27(2), 229-249.
- Hummel, H. G. K., Paas, F., & Koper, E. J. R. (2004). Cueing for Transfer in Multimedia Programmes: Process-worksheets vs. Worked-out Examples. *Journal of Computer Assisted Learning*, 20(5), 387-397.
- Hummel, H. G. K., Paas, F., & Koper, E. J. R. (in press). Timing of Cueing in Complex Problem-Solving Tasks: Learner versus System Control. *Computers in Human Behavior*.
- Jeong, H., & Chi, M.T.H. (2000). Does Collaborative Learning Lead to the Construction of Common Knowledge? [Retrieved online on July 9, 2004 from:  
[http://www.ircs.upenn.edu.edu/cogsci2000/PRCDNGS/SPRCDNGS/posters/jeo\\_chi.pdf](http://www.ircs.upenn.edu.edu/cogsci2000/PRCDNGS/SPRCDNGS/posters/jeo_chi.pdf)]
- Land, S. M. (2000). Cognitive Requirements for Learning with Open-ended Learning Environments. *Educational Technology Research and Development*, 48, 61-78.
- Mevarech, Z.R., & Kramarski, B. (2003). The Effects of Metacognitive Training versus Worked-out Examples on Students' Mathematical Reasoning. *British Journal of Educational Psychology*, 73, 449-471.



- Nadolski, R. J., Kirschner, P. A., & Van Merriënboer, J. J. G. (in press). Optimizing the Number of Steps in Learning Tasks for Complex Skills. *British Journal of Educational Psychology*.
- Nelson, L.M. (1999). Collaborative Problem Solving. In C.M. Reigeluth (Ed.). *Instructional Design Theories and Models: A new Paradigm of Instructional Theory* (pp. 241-267). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Owen, M. (2000). Structure and Discourse in a Telematic Learning Environment. *Educational Technology & Society*, 3.3. [Retrieved online on December 2, 2003 from: [http://ifets.ieee.org/periodical/vol\\_3\\_2000/b04.html](http://ifets.ieee.org/periodical/vol_3_2000/b04.html)]
- Paas, F. (1992). Training Strategies for Attaining Transfer of Problem-solving Skill in Statistics: a Cognitive Load Approach. *Journal of Educational Psychology*, 84, 429-434.
- Pawar, K.S., & Sharifi, S. (1997) Physical or Virtual Team Collocation: Does it matter? *International Journal of Production Economics*, 52(3), 283-290.
- Pearce, J. A. I., & Ravlin, E. C. (1987). The Design and Activation of Self Regulating Work groups. *Human Relations*, 40(11), 751-760.
- Renkl, A. (2002). Worked-out Examples: Instructional Explanations Support Learning by Self-explanations. *Learning and Instruction*, 12, 529-556.
- Shrout, P. E., & Fleiss, J. L. (1979). Intra Class Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin*, 86, 2, 420-428.
- Strijbos, J.W., & Martens, R.L. (2001). Coördinatieprocessen tijdens computerondersteund samenwerkend leren [Coordination processes during CSCL]. In P.A. Kirschner (Ed.), *Factoren die collaboratief leren beïnvloeden [Factors to influence collaborative learning]*. Heerlen, The Netherlands: Open University of the Netherlands.

- Taylor, K.L., & Dionne, J-P. (2000). Accessing Problem-Solving Strategy Knowledge : The Complementary Use of Concurrent Verbal Protocols and Retrospective Debriefing. *Journal of Educational Psychology*, 92, 3, 413-425.
- Van Bruggen, J., Kirschner, P. A. & Jochems, W. (2002). Representation of Argumentation in CSCL and Management of Cognitive Load. *Learning & Instruction*, 12(1), 121-138.
- Van Gog, T., Paas, F., Van Merriënboer, J.J.G., & Witte, P. (2004). A Comparison of Problem-Solving Information Elicited by the Think Aloud Method, Retrospective Reporting and Cued Retrospective Reporting. Manuscript submitted for publication.
- Veldhuis-Diermanse, A.E. (2002). *CSC Learning?: Participation, Learning activities and Knowledge construction in Computer-Supported Collaborative Learning in Higher Education*. Unpublished Thesis. Wageningen University, The Netherlands.
- Wever, B. de, Valcke, M., & Winckel, M. van (2003). The Impact of ‘Structure’ in CSCL environments: A Study with Medical students. *Paper presented at CSCL 2003 Doctoral Consortium*. [Retrieved online on July 14, 2004 from:  
<http://www.intermedia.uib.no/cscl/dic/files/deweever.pdf>]
- Wiley, D.A., & Edwards, E.K. (2003). Online Self-Organizing Social Systems: The Decentralized Future of Online Learning. [Retrieved online on March 12, 2004 from:  
<http://wiley.ed.usu.edu/docs/ososs.pdf>]
- Wöretshofer, J., Nadolski, R. J., Starren-Weijnenberg, A. M. A. G., Quanjel-Schreurs, R. A. M., Aretz, C. C. W. M., Van der Meer, N.H.W., Martyn, G., Van den Brink, H.J., Slootmaker, A. & Berkhout, J. (2000). Pleit voorbereid [Preparing a plea] (version 1.0) [computer software]. Heerlen, The Netherlands: CIHO.
- Yaffee, R. A. (1998). *Enhancement of Reliability Analysis: Application of Intra Class Correlations with SPSS*. [Retrieved online on March 11, 2003 from:  
<http://www.nyu.edu/acf/socsci/Docs/intracls.html>]

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)

| Nr   | Question  | Scale / options   | <i>M</i> | <i>SD</i> |
|------|---|---|----------|-----------|
| 1    | How much mental effort did you feel during the group discussion?                                | Very, very little (1) – Very, very much (9)                   | 3.53     | 1.44      |
| 2    | How motivated were you during the group discussion?   | Very, very little (1) – Very, very much (9)                   | 6.25     | .97       |
| 3a   | Indicate which statements are true, by dividing 10 points over .... [a-e]                       | Discussion took place in a positive atmosphere                | 2.39     | .78       |
| 3b   | .....   | Discussion led to new knowledge and improvement of the report | 1.89     | .83       |
| 3c   | .....   | I made a substantial contribution to the group report         | 2.22     | .44       |
| 3d   | .....   | I was able to clarify my opinions                             | 1.89     | .60       |
| 3e   | .....   | There was considerable mutual misunderstanding and conflict   | 1.44     | 1.24      |
| 4    | Indicate to which extent the discussion led to new knowledge and improvement of the report      | Very little (1) – Very much (5)                               | 3.42     | .77       |
| 5    | Which improvements will make the meeting more efficient?  | Open question   | -        | -         |
| 6.1  | Did you make use of the worked-out examples (WOE) when writing your individual report?          | Very little (1) – Very much (5)                               | 3.22     | 1.66      |
| 6.2  | Did you make use of the worked-out examples (WOE) when writing the group report?                | Very little (1) – Very much (5)                               | 2.83     | 1.15      |
| 6.3a | Indicate the contribution of WOE on the group discussion, by dividing 10 points over .... [a-e] | Used while orienting on the task                              | 1.61     | 1.92      |
| 6.3b | .....   | Used while planning the task                                  | 1.17     | 1.30      |
| 6.3c | .....   | Used while executing (process) the task                       | 2.00     | 1.82      |
| 6.3d | .....   | Used while finalising (product) the task                      | 2.94     | 2.51      |
| 6.3e | .....   | Did not use them  | 2.28     | 2.89      |
| 7.1  | Did you make use of the leading questions (PW) when writing your individual report?             | Very little (1) – Very much (5)                               | 3.00     | 1.57      |
| 7.2  | Did you make use of the leading questions (PW) when writing the group report?                   | Very little (1) – Very much (5)                               | 1.94     | 1.40      |
| 7.3a | Indicate the contribution of PW on the group discussion, by dividing 10 points over ....[a-e]   | Used while orienting on the task                              | .94      | 1.35      |
| 7.3b | .....   | Used while planning the task                                  | 1.11     | 1.45      |
| 7.3c | .....   | Used while executing (process) the task                       | 3.67     | 2.50      |
| 7.3d | .....   | Used while finalising (product) the task                      | .72      | 1.02      |
| 7.3e | .....   | Did not use them  | 3.56     | 3.18      |

Table 2

*Coding scheme for cognitive activity during collaboration*

| Nr   | Main Category   | Sub Category | Activity related to ...   | Prototype examples Bosmans case  |
|------|-----------------|--------------|---|--|
| 1    | PRODUCT         | ACTION       | <i>Executing actions:</i><br>applying information, writing, dictating, editing, ....  | <ul style="list-style-type: none"> <li>- How are we going to phrase this argument?</li> <li>- We should place to most important argument first in the list.</li> <li>- Let's delete that sentence anyway.</li> </ul>   |
| 2    | PROCESS         | ACTION       | <i>Preparatory actions:</i><br>searching information, reading aloud, selecting usable information, ....   | <ul style="list-style-type: none"> <li>- Reading aloud the exact text of the demanding party (what exactly is claimed here).</li> <li>- What is mentioned about this type of Honda?</li> <li>- What was the story behind the insurance?</li> </ul>   |
| 3    | PRODUCT         | HOW          | Discussing the <i>chosen</i> strategies or tactics, e.g. how to apply the solution or worked-out examples in the report.                                    | <ul style="list-style-type: none"> <li>- Are we claiming or disbanding the contract?</li> <li>- What is primary, subsidiary, ..?</li> <li>- Are we going to use liability?</li> <li>- Are we going to charge the process costs?</li> <li>- We better combine a neutral plea with emotions, but only when relevant.</li> </ul>                            |
| 4    | PROCESS         | HOW          | Discussing <i>possible</i> approaches or heuristics for report, e.g. by examining case law, consulting experts, or applying leading questions and criteria. | <ul style="list-style-type: none"> <li>- Can we use article 717 sub 4 as an exemption to non-conformity?</li> <li>- Does plaintiff claim miscarriage?</li> <li>- Should we include the meaning of opposing party in this argumentation?</li> <li>- Could we urge for minority as excuse?</li> <li>- Should we speak about mutual miscarriage?</li> </ul> |
| 5    | PRODUCT         | WHY          | Discussing juridical principles, rules and facts behind the <i>chosen</i> solution.   | <ul style="list-style-type: none"> <li>- Mentioning default is redundant here.</li> <li>- If article 218 sub c, then we refer to 6:230</li> <li>- Are we addressing this issue in a relational or more objective tone?</li> <li>- We should restrict to sub c, because ....</li> <li>- Which facts are still missing ?</li> </ul>                        |
| 6    | PROCESS         | WHY          | Discussing juridical principles, rules and facts behind <i>possible</i> solutions.  | <ul style="list-style-type: none"> <li>- Does default apply here?</li> <li>- Now, what exactly is the juridical question?</li> <li>- Is there a principal difference between making one or two test drives?</li> <li>- What is the technical state of the Honda ?</li> <li>- Does a duty of giving notice apply here?</li> </ul>                         |
| 7    | PRODUCT         | META         | Orientation, monitoring and evaluating chosen solution  | <ul style="list-style-type: none"> <li>- Let's leave those headings bold-faced ...</li> <li>- What should happen with this report?</li> <li>- Do you still think this sums it up well?</li> </ul>  |
| 8    | PROCESS         | META         | Orientation, monitoring and evaluating the collaboration  | <ul style="list-style-type: none"> <li>- Is everybody satisfied?</li> <li>- We should start dividing tasks.</li> <li>- Lets first have a look at what everybody has as extras.</li> </ul>  |
| 9-12 | TASK-IRRELEVANT | [various]    | Praise / complaints about programme or meeting. Reading or writing individually. Fragments that cannot be scored.   | <ul style="list-style-type: none"> <li>- How irritating that you cannot scroll through or print those documents</li> <li>- Replacing the computer or flap-over.</li> <li>- Audio fragment is not audible (bad quality of recording).</li> </ul>  |

Table 3 (A and B)

*Learning outcomes (normalised 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)*

| Condition | Cueing /       |           | No cueing /    |           | No cueing / no  |           | All           |           |
|-----------|----------------|-----------|----------------|-----------|-----------------|-----------|---------------|-----------|
|           | collaboration  |           | collaboration  |           | collaboration   |           | <i>n</i> = 28 |           |
|           | <i>(n</i> = 9) |           | <i>(n</i> = 9) |           | <i>(n</i> = 10) |           |               |           |
|           | <i>M</i>       | <i>SD</i> | <i>M</i>       | <i>SD</i> | <i>M</i>        | <i>SD</i> | <i>M</i>      | <i>SD</i> |
| PI before | 38.67          | 8.87      | 30.67          | 5.20      | 30.30           | 5.17      | 33.11         | 7.45      |
| PI after  | 47.22          | 5.83      | 41.22          | 8.27      | 36.70           | 8.16      | 41.54         | 8.49      |
| PN before | 71.67          | 5.59      | 59.89          | 3.33      | 53.90           | 9.54      | 61.54         | 9.98      |
| PN after  | 72.89          | 3.33      | 67.11          | 3.95      | 56.10           | 6.17      | 65.04         | 8.50      |

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

| Condition      | Cueing /       |           | No cueing /    |           | No cueing / No |           | All           |           |
|----------------|----------------|-----------|----------------|-----------|----------------|-----------|---------------|-----------|
|                | collaboration  |           | collaboration  |           | collaboration  |           | <i>N</i> = 46 |           |
|                | <i>(n</i> =18) |           | <i>(n</i> =18) |           | <i>(n</i> =10) |           |               |           |
|                | <i>M</i>       | <i>SD</i> | <i>M</i>       | <i>SD</i> | <i>M</i>       | <i>SD</i> | <i>M</i>      | <i>SD</i> |
| Outcome before | 55.17          | 18.44     | 45.28          | 15.62     | 40.70          | 9.79      | 48.15         | 16.59     |
| Outcome after  | 60.26          | 13.98     | 54.17          | 14.73     | 46.50          | 9.36      | 54.80         | 14.14     |
| Growth (delta) | 5.09           | 6.92      | 8.89           | 7.12      | 5.80           | 5.01      | 6.91          | 6.71      |
| Transfer plea  | 70.33          | 9.34      | 77.39          | 6.96      | 67.20          | 13.82     | 72.41         | 10.35     |

Table 4

*Coding results from group discussions: categories and level of cognitive activity*

(expressed as percentages of the total number of scored and task-valid items)

| Condition                                   | Cueing   |           | No cueing |           | All      |           |
|---|----------|-----------|-----------|-----------|----------|-----------|
|   | (n = 6)  |           | (n = 6)   |           | N = 12   |           |
|   | <i>M</i> | <i>SD</i> | <i>M</i>  | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Category 1. Action / product <sup>#</sup>   | 37.75    | 4.58      | 29.42     | 5.79      | 33.58    | 6.61      |
| Category 2. Action / process                | 18.67    | 8.80      | 18.92     | 8.66      | 18.79    | 8.32      |
| Category 3. How / product <sup>#</sup>      | 7.42     | 3.01      | 13.67     | 5.69      | 10.54    | 5.43      |
| Category 4. How / process                   | 11.08    | 9.01      | 11.42     | 4.09      | 11.25    | 6.67      |
| Category 5. Why / product                   | 7.08     | 2.91      | 6.67      | 2.94      | 6.88     | 2.80      |
| Category 6. Why / process                   | 8.67     | 1.75      | 10.42     | 4.12      | 9.54     | 3.15      |
| Category 7. Meta / product                  | 4.42     | 3.50      | 5.58      | 1.85      | 5.00     | 2.74      |
| Category 8. Meta / process                  | 4.83     | 1.97      | 3.92      | 2.50      | 4.38     | 2.20      |
| Level (of cognitive activity) <sup>#*</sup> | 43.58    | 4.50      | 51.68     | 6.30      | 47.63    | 6.71      |

<sup>#</sup> Significant difference ( $p < .05$ ) between conditions

\* Level of cognitive activity is portion of discussion (categories 3-8)

## Figure Captions

*Figure 1. Excerpts taken from concrete cueing examples*

When studying the file of case X (step 3 of the task procedure) students draw up a pleading inventory for case X. Some of the leading questions that have to be considered can be found on the left side (excerpts from the PW); part of the expert solution (i.e., possible answer to leading question 6) can be found on the right side (excerpts from the WOE), with article numbers referring to Dutch Law.

*Figure 2. Outline of experimental procedure*

PI = pleading inventory report

PN = pleading note report

coll = collaboration during small-group discussions





