

Cognitive load measurements and stimulated recall interviews for studying the effects of information and communications technology

Pieter J. Beers · Henny P. A. Boshuizen ·
Paul A. Kirschner · Wim Gijsselaers ·
Jochem Westendorp

Published online: 10 November 2006

© Association for Educational Communications and Technology 2006

Abstract Many researchers use information and communications technology (ICT)-tools to augment learning in a great variety of tasks. Their effects are generally measured in terms of intended outcomes. This article argues for the use of additional, more general measures to obtain a more complete impression of the effects of ICT-tools. The first study presented in this article shows why tools should not only be studied in terms of their specific intended outcomes, but also in terms of their effects on working memory, and the cognitive mechanisms needed to achieve the intended outcomes. The second study uses cognitive load measurements and stimulated recall interviews to obtain a more comprehensive view of the effects of learning tools. Results suggest that traditional outcome measures need to be complemented with quantitative and qualitative measures of cognitive processes to substantiate conclusions about intended effects of ICT-tools.

Keywords Cognitive load · Stimulated recall interviews · ICT-tools · Complex problem solving · Negotiation · Common ground

P. J. Beers (✉)

Faculty of Technology, Policy and Management, Delft University of Technology, PO Box 5015, 2600 GA Delft, The Netherlands
e-mail: p.j.beers@tudelft.nl

H. P. A. Boshuizen · P. A. Kirschner · J. Westendorp · P. J. Beers
Educational Technology Expertise Center, Open University of the Netherlands, PO Box 2960, 6401 DL Heerlen, The Netherlands

P. A. Kirschner
Research Center for Interaction and Learning, Utrecht University, PO Box 80140, 3508 TC Utrecht, The Netherlands

W. Gijsselaers
Educational Development and Educational Research, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

Introduction

Educators and researchers increasingly use information and communications technology (ICT)-tools to augment learning in a great variety of tasks (Jonassen, 2000; Norman, 1993). For example, Buckingham Shum, MacLean, Bellotti, and Hammond (1997) studied supporting group design processes, Suthers (2001) studied facilitation of scientific reasoning, and Van Bruggen (2003) studied support of argumentation. All of these researchers have convincingly shown that they achieved the effects they intended with their ICT-tools, that is, learners' task performance was positively affected by the use of the ICT-tools. However, these results beg the question of how they were achieved, and at what cost.

Generally these tools share two characteristics. They make use of *externalization of knowledge*, and they use specific *instructions* for the externalization of that knowledge to support a specific aspect of learning. These instructions encourage learners to engage in novel ways of representing knowledge that afford learning (Jonassen, 2000; Kirschner, 2002). The ultimate goal is that learners learn to apply, or transfer, these representation methods in contexts without the ICT-tool.

It has been argued (Scaife & Rogers, 1996) that external representation of knowledge provides an off-loading effect to learners, that is, representing knowledge externally lowers working memory load because information that is externally represented does not have to be kept in that memory. This frees up working memory resources, which subsequently affords processing for the instructed activities. However, Van Bruggen, Kirschner, and Jochems (2002) found that this need not necessarily be the case. If the specific instruction for knowledge externalization is poorly designed, for example when it does not fit the task it is intended to support or when the use of the instruction requires training, external representations may increase cognitive load due to the effort needed to apply the instruction. The instruction inherent to the use of these tools may thus increase cognitive load, and, in the worst case, could even counteract learning processes. An important key to successful ICT-support, thus, seems to reside in the way the user of the ICT-support is instructed to use it for knowledge externalization.

Several measures can be taken to ensure that the instruction given to the user does not cause extraneous working memory load (not facilitative to the learning process), while still resulting in germane (intended and helpful) effects on collaboration (Van Merriënboer & Ayres, 2005). First, the instruction given to the users should be plain and simple enough to be used without much training. Furthermore, the execution of the instruction during knowledge externalization can be supported by using techniques such as coercion¹ (i.e., the extent to which users are constrained to follow the instruction) or sentence

¹ We are aware that the word coercion can be used for its negative (punitive, oppressive) connotations. We wish to stress that this article uses the word neutrally, that is, in the sense of limiting options, constraining, as it is also used by Dillenbourg (2002).

openers (i.e., beginnings for sentences which act as advance organizers for the externalization of knowledge). Such support acts as a *performance constraint* in the sense that it restricts the range of possible actions by making certain actions that are not relevant to the instructed activities unavailable (Van Merriënboer & Sweller, 2005).

By their very nature, performance constraints implicitly structure the learner's actions without giving him/her insight into the reasons why this structure is necessary. Hence, if a task is supported too well it may prevent learners from developing a deeper understanding of the actions that are allowed/required, and thus prevent transfer to other situations. This means that a good balance is needed between instruction and performance constraints (at different stages of the learning process), and that a research methodology is applied that takes cognitive load and working memory processes into account, so as to show how, and at what cost, the instruction results in the intended effect.

The goal of this research is to study whether and how cognitive load measurements and interviews can complement traditional outcome measures (i.e., intended effects) to inform the design of support and instruction in ICT-tools. First, a study with a negotiation tool (NTool; Beers, Boshuizen, Kirschner, & Gijsselaers, 2005) is revisited to illustrate the tension that can occur between performance constraints and instruction. NTool is an ICT-tool that supports complex problem solving in multidisciplinary teams by increasing a team's common ground, a cognitive frame of reference shared by all team members (Bromme, 2000). Some of the unexpected results in experiments with NTool point in the direction of unwanted, or extraneous, side-effects of the NTool support principle, which in turn may point to defective performance constraints.

Next, we report on an exploratory study with instruction but without performance constraints that builds on these results, in which cognitive load measurements (Sweller, Van Merriënboer, & Paas, 1998) and stimulated recall interviews were used to study both possible beneficial and adverse effects of the instruction. The cognitive load measurements were used to study whether there were differences in working memory load. The interview data were used to get an impression of the actual cognitive processes of the participants during collaboration, so as to assess whether they were germane or extraneous to the process, and whether they gave insight into the effect mechanisms of the instruction.

The NTool support principle for complex problem solving

The NTool support principle for complex problem solving is based on three notions. First, complex (societal) problems are generally solved in multidisciplinary teams (DeTombe, 2002; Vennix, 1996). Second, problem-solving teams need some common ground; a shared cognitive frame of reference (Bromme, 2000) to be able to construct a shared problem-representation (Ostwald, 1996). Third, expert problem solvers spend relatively more time on

problem-representation than novices, which shows that the problem-representation phase is highly important for problem solving (Lesgold et al., 1988). The negotiation of common ground can be seen as an activity that is intrinsic to solving complex problems, because common ground is needed to afford the sharing of knowledge and the subsequent construction of a shared problem-representation in multidisciplinary teams.

NTool aims at supporting multidisciplinary teams by facilitating the negotiation of common ground. It does this by requiring problem solvers to explicitly verify their understanding of each other's contributions to a conversation and to explicitly articulate their positions on those contributions. The researchers expected that this would allow easy recognition of content areas where the participants lacked common ground, which in turn would lead to more negotiation of common ground. Ultimately, the intended learning effect of NTool is that the user learns how to successfully negotiate common ground. In terms of outcome, we thus expect more explicit negotiation of common ground and more common ground with the use of NTool in multidisciplinary teams that solve complex problems.

The NTool support principle consists of five rules for communication that mimic the process of negotiation (see Table 1). Participants are required to follow these rules by having to choose from a set number of message types during on-line communication. New conversation topics are introduced using a contribution message, and verified and clarified using verification and clarification messages. Furthermore, participants can use agree- and disagree-messages to make their position known to their team mates, and they can post rejections to messages that are unintelligible or objectively incorrect in their eyes.

Study 1

The original aim of Study 1 (Beers et al., 2005) was the continued development of NTool. Three versions of NTool were tested, with different levels of coercion. Higher coercion meant fewer degrees of freedom for the user when using NTool (coercion; Dillenbourg, 2002). These levels of coercion were implemented using performance constraints. The tested hypothesis was that the higher the coercion, and thus the stronger the performance constraints, the stronger the intended effects (i.e., more explicit negotiation, more common ground) of NTool would be. All users received the same instruction about the

Table 1 The NTool support principle

-
1. Every new issue is termed a contribution
 2. Contributions require a verification by the other team members
 3. Each verification is responded to with clarification by the original contributor
 4. When all verifications are clarified, and no new verifications are performed, all team members state whether they accept or reject the statement
 5. All team members state their position about accepted statements
-

support principle and version-specific instruction about the performance constraints. Here we present again our results about coercion and negotiation, but this time we focus on the unintended, unexpected effects of the performance constraints on regulation.

Method

NTool was tested in a laboratory setting in which 17 three-person multidisciplinary teams solved the complex societal problem of high school drop-out using NTool ($N = 17$). Three different versions of NTool were used.

Idiosyncratic

This version used explicit message types for verifying and clarifying contributions and for stating one's opinion. Support consisted of on-screen descriptions of the various message types and their uses. Furthermore, prompts informed participants about contributions that had not yet been verified. Participants were unrestricted in their use of the message types (i.e., no performance constraints).

Scripted

This version used the same message types as the idiosyncratic version, and also the same on-screen descriptions and prompts. However, the scripted version prevented the posting of certain messages at certain times. More specifically, it prohibited participants from posting messages in which they stated their opinions before all contributions had been verified. This resulted in a two-phase structure of the discussion. To end the first phase, participants were required to have verified and accepted all contribution messages. Phase 2 was aimed at sharing positions on the various contributions. Using prompts, participants were informed of which phase they were in.

Stringent

This version also used the same message types as the idiosyncratic version, but applied even more constraints on collaboration than the scripted version did. The stringent version allowed discussion about only one contribution at one time. Furthermore, participants were not allowed to compose reject-, agree-, and disagree-messages before the contribution had been verified. Using prompts, participants were informed as to whether they had to verify or decide on a contribution.

Six groups used the stringent version, five used the scripted version, and the final six groups could use the NTool idiosyncratically. Groups were given 90 min to complete the task. The instruction described the support principle, that is, the various message types and their uses in communication. The exact instructions differed only with regard to the descriptions of the performance

constraints, that is, what message type was allowed at what time. There were no differences with regard to explanation of the support principle itself.

Analysis focused on the number of messages that reflected explicit negotiation of common ground, and the number of messages with regulative content. Negotiation of common ground was defined as directly or indirectly requesting information about, or giving clarification on, the intended meaning of a contribution. Regulation was defined as regulating the collaboration process, which also entailed tool use, and monitoring the problem-solving process. Negotiation and regulation were measured using a coding scheme developed by Beers et al. (2005) that, among other things, distinguishes between new contributions, verification of one's own understanding of a contribution, clarifying another's understanding of a contribution, and regulatory activities. The coding scheme was based on theory on negotiation of common ground (Bromme, 2000; Clark & Brennan, 1991) and similar coding schemes for discussion and collaboration (Avouris, Dimitracopoulou, & Komis, 2003; Fischer, Bruhn, Gräsel, & Mandl, 2002; Mulder, Swaak, & Kessels, 2002), but the regulation category was added based on the data, that is, the code was not theoretically defined before analysis, which was the case for the other codes.

Results and discussion

NTool was shown to be increasingly effective with increasing coercion; there was a significant association between coercion and the negotiation of common ground, $r_s(N = 17) = 0.51$, $p < 0.05$. This means that the fewer degrees of freedom that the participants had, and thus the more the execution of the instruction was supported and constrained, the higher the effects on the negotiation of common ground. However, the results also showed that the medium coercion groups required significantly more regulation than the other groups, $U(N = 17) = 4.00$, $p < 0.01$ (see Table 2). This unexpected result is the more interesting one for the purpose of this article.

The important question is why the scripted groups regulated so much. As all groups were at least able to solve the problem, we may assume that the increased regulation did not concern the problem-solving process itself. Furthermore, we can assume that this regulation did not benefit the negotiation of

Table 2 Negotiation of common ground and regulation

	Coercion		
	Low	Medium	High
Negotiation of common ground ^a	33.7	45.2	51.6
Regulative utterances ^b	30.7	106.0	43.7
<i>N</i>	6	5	6

^a Negotiation was associated with coercion

^b Regulation was significantly higher in the medium coercion groups than in the high and low coercion groups

common ground, because then we would have found a correlation between regulation and negotiation. This leaves the possibility that the scripted groups regulated more about the use of NTool than the idiosyncratic groups and the stringent groups. Thus, the high number of regulation messages in the scripted groups may have been *extraneous* to the task. In other words, the scripted version caused an increase in activities that did not benefit the instruction's purposes. It stands to reason that this was caused by differences in performance constraints between the three versions of NTool and/or deficiencies in the instruction itself. From a design perspective, this difference is highly important; if it turns out that the performance constraints are poorly designed, the problem can be overcome by improving their design. If, however, the instruction turns out to be deficient, the question must be raised as to whether the goal of the ICT-tool can be reached at all.

What conclusions can we draw about the performance constraints and the instruction? Let's first examine how the differences in performance constraints between the scripted and stringent versions of NTool may have affected regulation as they did. The stringent version of NTool constrained collaboration quite strongly (high coercion). This did not put additional regulative demands on the participants (compared to the idiosyncratic groups), even though the performance constraints restricted many options during collaboration. It can thus be concluded that the performance constraints were designed well enough to not require additional regulative activities. Participants using the scripted version of NTool required a great deal of regulation during the whole session to keep their discussion going.

A possible explanation lies in the design of the performance constraints. The performance constraints in the stringent version effected a sequence of communicative acts that was repeated a number of times throughout the discussion. This may have allowed participants to learn about them by experience, so they did not need to remember the instruction about the performance constraints for the entire duration of the discussion. In contrast, the scripted version effected a one-pass procedure that did not enable experiential learning about the performance constraints, and therefore the participants had to rely more heavily on recall of the instructions during collaboration. This means that participants in the scripted groups had to keep the instruction about the performance constraints in working memory for the entire duration of the experiment to be able to navigate the scripted version.

The above explanation is very interesting from the point of view of cognitive load theory (Sweller et al. 1998), which emphasizes the limited nature of working memory and its consequences for instruction. For instance, this theory holds that working memory overload can hamper learning, and that instruction should minimize extraneous cognitive load, that is, working memory activity caused by the instruction that does not benefit the learning process. Indeed, having to attend to regulatory activities would cause extraneous cognitive load within our settings. Unfortunately, we did not measure cognitive load.

Now that we have shown that the performance constraints in the scripted version may have lead to extraneous activities in terms of working memory load and regulation, and therefore were defective, it is important to also rule out the possibility of deficient instruction. Maintaining a discussion did not appear to be difficult with the idiosyncratic, as shown by the relatively low number of regulation messages. This shows that, in any event, the instructions need not result in increased regulation. However, it was also the case that the idiosyncratic version of the NTool was probably the easiest to use, as it did hardly have performance constraints. Furthermore, no intended effects of the instruction are apparent as the idiosyncratic version featured the least negotiation. In other words, the low number of regulative activities can also be the consequence of neglect of the instruction, be it good or deficient, on the part of the participants.

The stringent version resulted in the most negotiation. But what if this result was the effect of the performance constraints only? Possibly participants in the stringent groups only required the performance constraints to use NTool and solve the problem. So, did they actually learn about the support principle as per the instruction as well? In other words, they may have learned how to use the tool, but did they learn how to negotiate common ground?

For both the stringent and the idiosyncratic version, possibly either the lack or the abundance of performance constraints may have resulted in a discussion that demanded little regulative activity. So far, we cannot preclude that the same results would have been found without instruction about the support principle. In sum, notwithstanding the intended effects in Study 1, it is unclear whether the results should be attributed to the instruction itself, the performance constraints, or a combination of the two. Did participants in the idiosyncratic and stringent groups need the instruction at all? Did participants in the scripted groups require more regulation due to badly designed performance constraints, or due to deficient instruction about the support principle? Apparently, outcome measures alone do not give full insight into the effects of NTool. To distinguish between the effects of the support mechanisms and the effects of the instruction, a study with only instruction, and no performance constraints, that is no NTool, whatsoever was performed.

Study 2

In Study 2 a face-to-face setting without ICT-support was used to rule out any effects of performance constraints on grounding, and to enable studying effects of instruction only, with the use of new types of measurement. A pen-and-paper approximation of the NTool support principle was used as a face-to-face carrier of the instruction. Study 2 can thus be regarded as piloting a new measurement strategy on the basis of Study 1, which was a more extensive laboratory study. Cognitive load measurements and stimulated recall interviews were used to gain insight in the working memory effects of the NTool instruction and to distinguish the intended positive (i.e., germane)

effects of the instruction from its extraneous effects. The main aim of Study 2 was to distinguish the effects of instruction from the effects of performance constraints and to gain insight in the nature and mechanisms of the instruction effects. The main research question was: Can cognitive load measurements and stimulated recall interviews complement intended outcome measures to study the effects of ICT-tool instruction?

Cognitive load theory (Sweller et al., 1998) couples insights on working memory to the design of instruction. It holds that the human cognitive architecture consists of a limited working memory which interacts with an unlimited long-term memory. Working-memory processes information to construct and automate long-term cognitive schemas. Any learning material causes an associated *intrinsic* cognitive load that depends on the complexity of the material itself (the number of elements involved and their interactivity) and the prior knowledge of the learner. The more complex the learning matter, and the less prior knowledge of the learner, the higher the associated intrinsic cognitive load (Sweller et al., 1998).

The intrinsic load can be seen as a given for instructors, because “intrinsic cognitive load cannot be altered by instructional interventions” (Sweller et al., 1998, p. 259). There are two other types of cognitive load that can be affected by instruction. *Germane load* is caused by working memory processes that lead to schema construction and automation, whereas *extraneous load* is caused by trying to understand the instruction aimed at generating germane processes, but that is not beneficial to the learning itself. Because of working memory limitations, it is important that extraneous load is minimized, and that total load does not exceed working memory limitations (Sweller et al., 1998).

In the case of ICT-tools, we assume that the working memory processes that are important to the intended learning effects and those important for completing the task can be seen as constituting germane working memory load, whereas processes that constitute thoughts about the instruction that do not lead to the intended learning outcomes, or are not beneficial to the task, can be seen as extraneous. This means that the instruction needs to be simple so as to minimize extraneous load, and effective so as to increase germane load.

Research focused on (1) the extraneous and germane effects of the tool in terms of cognitive load, (2) the intended outcomes of the tool (i.e., negotiation, common ground), and (3) the possible mechanisms for the effect of instruction. Cognitive load was studied quantitatively via a questionnaire (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Furthermore, stimulated recall interviews were held to gain qualitative insight in the actual working memory processes, and instruction effect mechanisms.

The instruction was expected to result in more negotiation of common ground and more common ground. As a consequence, it was expected to result in more cognitive load during collaboration, either due to an increase in germane load as a result of increased grounding activities or due to an increase in extraneous load due to increased participant attention to the

instruction. The interview data were used to qualitatively identify the actual processing activities and to distinguish whether they were germane or extraneous to the task.

Method

As in Study 1, participants worked in three-person multidisciplinary teams solving the complex problem of high school drop-out. The same overall procedure as in Study 1 was used, but with additional measures for cognitive load. The main difference between the studies was the implementation of the NTool support principle (using ICT in Study 1 versus a pen-and-paper approximation in Study 2).

Participants

Participants were 12 undergraduate students from the fields of psychology, economics, and cultural sciences from a Dutch university. Four multidisciplinary teams were formed by assigning one student from each discipline to a triad.

Materials

The pen-and-paper version of the NTool support principle was used with a whiteboard and four colored whiteboard markers (black, blue, red, and green) given to all teams for creating an external representation during group collaboration. Two teams received specific instruction for using the whiteboard and the markers (instruction condition); the other teams could use their whiteboard and markers any way they wanted (idiosyncratic condition). The instruction required participants to write new contributions on the whiteboard (i.e., externalization of concepts), and to react on others' new contributions by giving their own perspective on them (i.e., negotiation of concepts). Participants were assigned personal colored whiteboard markers to allow easy recognition of contributors. The execution of the instruction was not supported during collaboration.

Cognitive load measurement was done through self-report of invested mental effort on a nine-points symmetrical scale ranging from very, very low mental effort to very, very high mental effort (Paas et al., 2003). Mental effort refers to the cognitive capacity that is actually allocated to solve the problem and can be considered to reflect the actual cognitive load (Sweller et al., 1998).

Procedure

Each team was given 30 min to collaboratively analyze the drop-out problem so as to come up with a solution. The team collaborations were videotaped. Participants were required to write down their individual perception of the problem and their solution(s) both prior to (pre-test) and after (post-test) the

team collaboration. Cognitive load was measured three times, namely after the pre-test, after the team collaboration, and after the post-test. Within 2 h after the post-test, open and stimulated recall interviews were carried out in which participants were asked to recall their thoughts during collaboration.

Quantitative analyses

The first measure of cognitive load, carried out after the pre-test, served as a baseline measure. No differences were expected. The second measure was carried out after group collaboration. Participants in the instruction groups were expected to report higher cognitive load, either due to more negotiation (i.e., germane cognitive load) or due to using the instruction (i.e., extraneous cognitive load), or both. The third measure was carried out after the post-test. Again, the instruction groups were expected to report higher cognitive load. Differences were expected due to differences in negotiation processes during collaboration, as caused by the germane effects of the instruction. In other words, higher cognitive load after the post-test was assumed to reflect germane load.

Negotiation was again measured with the coding scheme developed by Beers et al. (2005) to distinguish between new contributions, verification of one's own understanding of a contribution, and clarification of another's understanding of a contribution.

Common ground was measured by comparing content overlap in pre-tests and post-tests. Content was operationalized as contributions from the group discussion that appeared in the pre-tests and post-tests. For each contribution, it was determined in how many pre-tests and post-tests it showed up, after which the group mean for the average contribution was used as an operationalization for overlap between individual representations. The higher the change in overlap between pre-test and post-test, the more common ground.

Qualitative analysis of the interview data

The interviews were qualitatively analyzed to gain insight in the participants' actual thought processes and how they related to germane and extraneous load. Analysis focused on participants' thoughts about grounding processes, about knowledge construction, about the instruction, and about their explicit thoughts about effort involved in following instructions.

In the case of grounding, thoughts about understanding each other, about each other's positions, and about perspectives on the task were seen as indicative for negotiation activities, and therefore germane to the task. Thoughts about co-construction of knowledge were seen as intrinsic to the task, but not specific to negotiation activities. Thoughts about the use of the whiteboard were seen as germane to task if they were linked to negotiation activities and co-construction of knowledge. Other thoughts about the instruction and the use of the whiteboard were seen as extraneous to the task.

All reported thoughts from the interview transcripts about negotiation processes and instruction were gathered and categorized. The qualitative analysis was carried out by the author.

Results

Cognitive load

No differences in cognitive load after the pre-tests were found (see Table 3). Contrary to expectations, no differences were found after group collaboration. However, as expected after the post-tests, participants in the instructed groups reported significantly more cognitive load, $U(N = 12) = 7.00$, $p < 0.05$, one-tailed, suggesting that these participants required more processing in the post-test than the participants in the idiosyncratic groups.

Negotiation and common ground

No statistically significant differences were found with regard to negotiation and common ground. Contributions were most heavily negotiated in the idiosyncratic groups (see Table 4), instructed groups made more contributions.

The idiosyncratic groups achieved the most common ground after collaboration, as shown by the content overlap in post-tests, but they also started with some unexpected overlap in pre-tests, which the instructed groups did not have (see Table 5). Differences in common ground were not statistically significant.

Table 3 Cognitive load

	Cognitive load measured after the	
	Condition	
	Idiosyncratic	Instructed
Pre-test	5.67	5.50
Group collaboration	5.83	6.00
Post-test ^a	5.00	6.17
<i>N</i>	6	6

^a Cognitive load after the post-test was higher in the instructed groups than in the idiosyncratic groups

Table 4 Negotiation and other utterances

	Condition	
	Idiosyncratic	Instructed
New contributions	8.0	13.0
Negotiation of common ground	33.0	13.5
Other task-oriented utterances	124.0	91.5
Regulative utterances	25.0	13.5
Other utterances	8.5	14.0
<i>N</i>	2	2

Table 5 Common ground

Content overlap ^a in the	Condition	
	Idiosyncratic	Instructed
Pre-tests	1.35	1.00
Post-tests	2.01	1.41
<i>N</i>	2	2

^a A content overlap of 1 is the minimum score, meaning that conversation topics were mentioned by only one person. A content overlap of 3 means that all conversation topics present pre-tests and post-tests were mentioned by all team members

Interview data

Grounding processes and knowledge construction Interview data reveal that participants are aware of instances of agreement and disagreement, and also of mutual understanding and misunderstanding. In other words, they know the status of a contribution to the discussion:

S1E:² “It then crosses my mind that she makes points that I have thought of myself, so I thought that’s OK, so I also agreed with her.”

I1C: “But... yeah, I got that feeling, I don’t think you get it, and then that turns out to be right.”

S2P: “I thought that there might be possibilities, but also that it would be difficult, the way she put it.”

These excerpts show that participants are aware of the distinction between understanding and agreeing, as also shown by the following excerpt, which reflects agreement to disagree:

S2C: “I think I understand what she thinks about this, and I also understand that the positions she’s taking and I am taking, uh, differ.”

Awareness of agreement to disagree was only mentioned by participants from the instructed groups.

The interview data also give insight in the way participants react to perceived (mis)understanding and (dis)agreement. For instance, participants report thoughts about the background of others’ contributions to the discussion, and others’ educational and even philosophical backgrounds are mentioned as well. From these excerpts it seems that participants actively attribute contributions to the contributor’s background:

² Interview excerpts are coded as follows; the first letter signifies experimental condition (“S” for instructed groups and “I” for idiosyncratic groups); the digit represents the group number (starting at “1” for each experimental condition); and the last letter signifies educational background (“C” for Cultural Sciences, “E” for Economics and Business Administration, and “P” for Psychology).

I1C: “With regard to the problem we’re working on, how ... um ... why does she mention this.”

S2C: “At this moment, I’m trying to understand what the philosophical view of the other positions ... is.”

I2P: “The difference in background that appears quite quickly ... that he clearly has a background in economics.”

Several qualitatively different reactions to differences in understanding and position become apparent from the interview data, ranging from outright neglect of mutual misunderstanding to accepting something on the basis of another’s expertise:

I1E: “I got what she meant, but well, you don’t completely go like, what’s your view then and bla, bla, bla, that doesn’t matter too much.”

I1C: “... and then you hope they’ll nod, like yeah, that’s right what you’re saying or, or, we understand, and that didn’t happen.”

These excerpts show that possible misunderstandings are sometimes detected, and not actively addressed. Another reaction was waiting, leaving the contributor time to explain his/her intentions, as opposed to immediately giving one’s primary response:

S1C: “I was like, wow, that’s radical, and then I thought, well, maybe it has something.”

S2C: “... assuming that you’ve understood something, that ... that, that what the other says is something you recognize, I’m very cautious about that.”

These excerpts show that participants sometimes consciously wait for the contributor’s clarification to see whether their primary (negative) reaction toward a contribution is justified. It seems that withholding one’s reaction for a short time may allow for understanding that otherwise would not have emerged. Such active attention to mutual understanding was also mentioned on a more general note:

S2C: “At that moment I thought, OK, we’re really doing our best to understand each other and get somewhere.”

I1P: “I actually said that just to see if I had understood.”

Finally, in both conditions, the interview data show that participants actively build on each other’s knowledge, and are also capable of revising their own and each other’s ideas:

I2P: “Here I don’t wholly agree ... so that’s why I’m putting it down just a bit differently.”

S2C: “He takes my criticism seriously. And that he’s willing to consider it, then also continue thinking along that line.”

S1P: “And this was more of building on each other.”

S1E: “I thought that’s a smart move that she more or less linked her solutions and the points I was contributing at that point.”

Effects and use of whiteboard and instruction The interview data suggest that using the whiteboard and colored markers helped structure the discussion and keep track of individual contributions:

S2E: “I thought it contributed to structure, the writing down.”

I1C: “... to keep track of what we had, ... and keeping track of what needs to be discussed and things we might forget.”

S2P: “I saw quite a bit of blue and red appear on the board, but hardly any green, so ...”

Some interview excerpts suggest ways in which the whiteboard may have contributed to the negotiation of common ground. These excerpts indicate that the whiteboard takes away the necessity of immediately understanding something or agreeing on something and reacting, because the participants can refer to what is on the whiteboard later in the discussion. In other words, if something is on the whiteboard, it is easier to withhold a primary response:

S2C: “In a discussion you have to react immediately if you want to support or challenge someone. And with a whiteboard, you can, it’s on there. So you can reconsider.”

I1E: “Sometimes when you didn’t completely understand or had missed what someone had said then you could also see it on the whiteboard.”

Furthermore, some interviewees linked using the whiteboard to the negotiation of meaning. From these excerpts it seems that, on the one hand, the requirement to write something down requires participants to be clear and on the other hand, that the presence of something on the whiteboard facilitates access to the meaning of others’ contributions.

S2C: “... and maybe that the whiteboard requires you to be a bit more explicit.”

S1C: “I liked it this way [with the whiteboard] because this way you can see what someone really means.”

From further interview excerpts it seems that the co-presence of different contributions on the whiteboard may facilitate co-construction of knowledge:

I1C: “And then we looked at those things that the three of us hadn’t thought of individually, but at which we arrived as a group by reading each others notes.”

S1E: “Like I just said, you see ‘culture’ and ‘motivation’ put down [on the board] and then you combine those.”

The instructed groups were told to share individual perspectives on the whiteboard:

S2C: “In principle this [sharing and writing down opinions] goes well, if only you’ve had a small discussion before. That you first state your opinion, like this and this is what I mean.”

S1C: “I think it [following the instruction] worked out, if someone said something, then another would build on that or say whether or not they agreed.”

S1P: “I thought it went quite well. Everybody had their own opinion, the only thing is, well, we actually agreed very much on many things.”

From the interview data, it can be concluded that writing down opinions on the whiteboard did not pose a problem to the participants. It also seems that the participants valued talking about those opinions so as to clarify what they meant. These excerpts, thus, suggest that requiring groups to share individual perspectives may encourage them to negotiate the meaning of individual contributions before they get written down. However, one interviewee mentioned that the whiteboard was not used as intended:

S1E: “... because after that [the beginning] we only used the whiteboard to summarize our discussion.”

The interview data also suggest some areas of difficulty encountered when using the whiteboard:

I1C: “It can slow down a discussion, because you keep, at least in my case because I was taking the notes, busy in your head with how to write something down.”

S2E: “You talk faster than you write.”

It appears that being required to both take notes and discuss something at the same time can sometimes be taxing.

Discussion

Differences in cognitive load were found after the individual post-test, but, unexpectedly, not after group collaboration. The latter result may be explained by the interview data which showed that participants did not find using the whiteboard this way effortful, so even if discussions were different, this may have caused the same cognitive load.

In line with the expectations, cognitive load was highest when participants solved the problem individually after the group work, in the instructed groups. The instruction may have elicited more knowledge in the group work that was new for individual participants. After the group work, they may have been busy processing the others' contributions while producing an individual solution the task. This explanation points at germane effects of the instruction.

The findings on negotiation and common ground were not in the expected direction, but were statistically inconclusive. Nevertheless, the fact that negotiation was lower in the instructed groups, while they made more contributions, may be explained by the instruction itself, that is, always stating one's perspective which may have resulted in elicitation of further contributions instead of negotiation of existing contributions.

The stimulated recall data suggest a number of mechanisms for the use of the whiteboard to facilitate negotiation processes. First, the presence of contributions on a whiteboard allows one to postpone one's reaction and wait for further clarification instead of immediately stating one's opinions. Second, the co-presence of different contributions on the whiteboard appears to allow co-construction of knowledge. Finally, the instruction to publicly write down one's perspective appears to allow the (spoken) negotiation of meaning of individual contributions, prior to actually writing them down.

General discussion

This article aimed to explore how cognitive load measurements and interview data might complement traditional outcome measures to study the effects of instruction itself, in the absence of performance constraints.

Study 1 illustrated that ICT-tools may cause unwanted side-effects that are extraneous to the goals of those tools in terms of extra effort on the part of the participants. Although all intended effects (i.e., negotiation of common ground) were achieved as expected (i.e., more coercion resulted in more negotiation), the results did not preclude that the effects were produced by performance constraints alone.

It was hypothesized that cognitive load measurements and qualitative interview data might give insights in the effects of ICT-tool instruction. Study 2 showed that cognitive load measurements indeed can yield information about germane and extraneous cognitive load effects of ICT-tools. No differences were found after group collaboration, but the higher cognitive load in the instructed groups after the individual post-test suggests some germane effects of the instruction. With regard to adverse effects of NTool, the interview data indicated that the instruction for using the whiteboard and the markers in Study 2 does not significantly increase extraneous load compared to the groups without the instruction, and that it may have had some germane effects on common ground. In sum, the cognitive load measurements in Study 2 suggest an increase in germane load, which may have been caused by increased grounding activities. However, these conclusions cannot be substantiated by the negotiation and common ground measurements.

Study 2 also exemplified how qualitative interview data may give insight in the effect of instruction. Notwithstanding the absence of statistically significant differences in outcome measures, the interview data suggest how the instruction might result in germane processing by allowing co-construction of knowledge and negotiation activities prior to writing down one's own perspective. This facilitates postponing one's primary reaction, which could in turn allow for consideration of another's contributions that otherwise does not take place. Such consideration is in line with the theorized effect mechanism of the NTool support principle, namely supporting the clarification and verification of another's contribution. The interview data also indicated where the instruction might have some extraneous effect, that is in dividing participants'

attention between writing things down on the whiteboard and keeping up with the discussion.

As differences between the conditions were small, and as the number of groups in this study was low, more research is needed to strengthen these hypotheses. The cognitive load measurements and the stimulated recall interview lend credence to a possible germane effect of the NTool support principle, which suggests that the intended NTool effects in Study 1 can be attributed, at least in part, to instruction alone. The necessity for regulation in the scripted (i.e., medium coercion) groups may thus be attributed mainly to defective performance constraints, and not to deficient instruction. More research on a larger scale is needed, using both intended outcome measures and other measures, to substantiate this conclusion.

For the further design of NTool, the cognitive load measurements combined with the insights from the stimulated recall interviews have helped to rule out the possibility that our instructions did have no intended effect at all. Furthermore, as mentioned before, future research using these measures can yield insights in the nature of the unintended effects of the scripted version of NTool. In general, it appears that cognitive load measures, coupled with interview data, can yield substantial information about the effects of ICT-tools, and their nature. They can be used to double-check expected outcomes in terms of learning results, and they can help explaining unexpected results. Cognitive load measures can give quantitative information about differences between versions of ICT-tools, and interview data about the actual thoughts of participants can indicate qualitative effects of ICT-tool design characteristics, and their germane or extraneous nature with regard to the task and learning goals at hand.

In sum, it appears that both cognitive load measurements and qualitative interview data provide insights in the NTool instructions that could not have been achieved through outcome measures alone. For designers of ICT-tools, the use of outcome measures and complementary measurements of cognitive processes can help to gain systematic insights regarding the effects of those tools.

Acknowledgments We wish to express our gratitude to the pub.group at OTEC and Tamara van Gog in particular, and Marjolein van Asselt for their comments on earlier drafts of this article. This research was supported by the Netherlands Scientific Organization (NWO) within the Society and Electronic Highway (MES) program under grant number 014-43-704.

References

- Avouris, N. M., Dimitracopoulou, A., & Komis, V. (2003). On analysis of collaborative problem solving: An object-oriented approach. *Computers in Human Behavior, 19*, 147–167.
- Beers, P. J., Boshuizen, H. P. A., Kirschner, P. A., & Gijssels, W. H. (2005). Computer support for knowledge construction in collaborative learning environments. *Computers in Human Behavior, 21*, 623–643.
- Bromme, R. (2000). Beyond one's own perspective: The psychology of cognitive interdisciplinarity. In P. Weingart, & N. Stehr (Eds.), *Practicing interdisciplinarity* (pp. 115–133). Toronto, Canada: University of Toronto Press.

- Buckingham Shum, S. J., MacLean, A., Bellotti, V. M. E., & Hammond, N. V. (1997). Graphical argumentation and design cognition. *Human-Computer Interaction, 12*, 267–300.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington: American Psychological Association.
- DeTombe, D. J. (2002). Complex societal problems in operational research. *European Journal of Operational Research, 140*, 232–240.
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61–91). Heerlen, The Netherlands: Open Universiteit Nederland.
- Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*, 213–232.
- Jonassen, D. H. (2000). *Computers as mindtools for schools: Engaging critical thinking*. Columbus: Prentice-Hall.
- Kirschner, P. (2002). Can we support CSCL? Educational, social and technological affordances for learning. In P. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL*. Inaugural address, Open University of the Netherlands.
- Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*. Hillsdale: Lawrence Erlbaum Associates.
- Mulder, I., Swaak, J., & Kessels, J. (2002). Assessing group learning and shared understanding in technology-mediated interaction. *Educational Technology & Society, 5*, 35–47.
- Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. Cambridge: Perseus Publishing.
- Ostwald, J. (1996). *Knowledge construction in software development: The evolving artifact approach*. Unpublished PhD thesis, University of Colorado, Boulder.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–71.
- Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies, 45*, 185–213.
- Suthers, D. D. (2001). Towards a systematic study of representational guidance for collaborative learning discourse. *Journal of Universal Computer Science, 7*(3). Retrieved January 23, 2004, from http://www.jucs.org/jucs_7_3/towards_a_systematic_study.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–295.
- Van Bruggen, J. M. (2003). *Explorations in graphical argumentation. The use of external representations in collaborative problem solving*. Unpublished PhD thesis, Open University of the Netherlands, Heerlen, The Netherlands.
- Van Bruggen, J. M., Kirschner, P. A., & Jochems, W. (2002). External representation of argumentation in CSCL and the management of cognitive load. *Learning and Instruction, 12*, 121–138.
- Van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development, 53*(3), 5–13.
- Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*, 147–177.
- Vennix, J. A. M. (1996). *Group model building: Facilitating team learning using system dynamics*. Chichester, UK: Wiley.

Pieter J. Beers has a background in environmental health sciences. He did his PhD on knowledge sharing in multidisciplinary teams. He currently works as a researcher at the Policy Analysis section at Delft University of Technology on conceptual analyses of actor models.

Henny P. A. (Els) Boshuizen is an international specialist in expertise development. She is professor of education and educational technology and dean of the Master of Active Learning programme at the Open University of the Netherlands.

Paul A. Kirschner is professor of educational psychology and ICT at Utrecht University and professor of educational technology at the Open University of the Netherlands.

Wim H. Gijsselaers is professor of education at Maastricht University. His research focuses on determinants of learning in Problem-Based Learning, Team Learning, and Professional Development.

Jochem J. Westendorp studied Cognitive Psychology at Maastricht University. His Master's Thesis treats of cognitive load and the sharing of knowledge in multidisciplinary teams.