Computer-Supported Collaborative Learning DOI 10.1007/s11412-008-9059-z

 $\frac{1}{2}$

Productive failure in CSCL groups

Manu Kapur • Charles K. Kinzer

3

5

6

7

8

4

Received: 7 February 2008 / Accepted: 21 November 2008 © International Society of the Learning Sciences, Inc.; Springer Science + Business Media, LLC 2008

Abstract This study was designed as a confirmatory study of Kapur's (Cognition and 9 Instruction, 26(3), 379–424, 2008) work on productive failure. N=177, 11th-grade science 10students were randomly assigned to solve either well- or ill-structured problems in a 11 computer-supported collaborative learning (CSCL) environment without the provision of 12any external support structures or scaffolds. After group problem solving, all students 13individually solved well-structured problems followed by ill-structured problems. Com-14 pared to groups who solved well-structured problems, groups who solved ill-structured 15problems expectedly struggled with defining, analyzing, and solving the problems. 16However, despite failing in their collaborative problem-solving efforts, these students 17 outperformed their counterparts from the well-structured condition on the individual near 18 and far transfer measures subsequently, thereby confirming the productive failure 19 hypothesis. Building on the previous study, additional analyses revealed that neither 20preexisting differences in prior knowledge nor the variation in group outcomes (quality of 21solutions produced) seemed to have had any significant effect on individual near and far 22transfer measures, lending support to the idea that it was the nature of the collaborative 23process that explained productive failure. 24

Keywords Ill-s	structured problem solving \cdot Well-structured problem solving \cdot	25
Synchronous co	llaboration · Problem-solving failure	26

27

Generally speaking, situative socio-constructivist perspectives underpin much of CSCL 28 research (Brown et al. 1989; Scardamalia and Bereiter 2003; Stahl 2005). An integral 29 proposition of this perspective is that learners need to be engaged in solving authentic, ill-30

M. Kapur (🖂)

C. K. Kinzer Teachers College, Columbia University, 525 W 120th Street, New York, NY 10027, USA e-mail: kinzer@tc.edu

National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616, Singapore e-mail: manu.kapur@nie.edu.sg

structured problems for deep learning to take place (Hmelo-Silver 2004; Spiro et al. 1992). 31Because ill-structured problems tend to be complex and often beyond the existing 32 knowledge and skill sets of learners, a substantial amount of CSCL research seeks ways of 33 supporting the collaborative, problem-solving interactions by providing some structure in 34one way or another. Furthermore, a substantial amount of research shows that productive 35collaboration does not often happen when learners are left to their own devices, without the 36 provision of support structures or scaffolds (e.g., Bromme et al. 2005; Hewitt 2005; 37 Sandoval and Millwood 2005). It is not surprising, therefore, that CSCL research has 38 tended to focus more on what is gained from such structuring of collaborative interactions 39but not as much on what could be lost (for a more general argument, see Reiser 2004). 40

Structure in a problem-solving activity, broadly conceived, can be operationalized in a 41 variety of forms such as structuring the problem or task itself, scaffolding, provision of 42tools and resources, micro and macro scripting, expert help, and so on (more examples in 43the following section). Thus conceptualized, structure corresponds to the degrees of 44 freedom in an activity; the greater the structure, the lower the degree of freedom (Cronbach 45and Snow 1977; Kauffman 1995; Reiser 2004; Voss 2005; Woods et al. 1976). Our 46argument is not that one should not structure collaborative interactions at all. Believing in 47the efficacy of structuring what might otherwise be a complex, divergent, and unproductive 48process is indeed well supported by CSCL research (Fischer et al. 2007). Instead, our 49argument is to allow for the concomitant possibility that under certain conditions, even ill-50structured, complex, divergent, and seemingly unproductive processes have a hidden 51efficacy about them. It is, perhaps, not so entirely inconceivable that by not structuring the 52collaborative problem-solving process—leaving groups of learners to struggle and even fail 53at tasks that are ill-structured and beyond their skills and abilities—may be a productive 54exercise in failure. It is this possibility that we have explored in our ongoing research 55program on productive failure in CSCL groups. 56

This manuscript comes in four sections. In the first section, we briefly review CSCL 57research with regard to how structure has been designed to support CSCL group 58interactions. We argue that just as it is important to investigate conditions under which 59ill-structured problem-solving activities can be structured so that they lead to productive 60 outcomes, it is equally important to investigate conditions under which ill-structured 61 problem-solving activities, when left without any external support structures, may lead to 62 productive failure. In the second section, we briefly describe one such investigation that 63 formed our initial study on productive failure in CSCL groups (Kapur 2008). Noting the 64limitations of one study and the paucity of confirmatory work in education research, we 65 argue for and describe a confirmatory study of productive failure. Furthermore, we extend 66 the findings from the initial study and report additional analyses and findings not reported 67 in the initial study. The confirmatory study with its research design and procedures, 68 instrumentation, data analyses, and results (including the additional analyses) form the main 69 focus of this manuscript and are described in the third section. The fourth and final section 70concludes the manuscript with a discussion of our findings, the limitations, and implications 71for future research. 72

Structure in CSCL

73

A focus on structuring collaborative problem-solving interactions has resulted in a healthy diversity of conceptions of what constitutes structure. Examples include: a) problem structuring (e.g., Jonassen and Kwon 2001; Kapur and Kinzer 2007), b) metacognitive 76 Computer-Supported Collaborative Learning

support through reflection prompts (e.g., Lin et al. 1999), c) content support (e.g., Fischer 77 and Mandl 2005), d) interactional support through question prompts (e.g., Ge & Land 782003), e) supporting group discourse through argumentation tools and representational 79guidance (e.g., Cho and Jonassen 2002; Lund et al. 2007; Mirza et al. 2007; Suthers and 80 Hundhausen 2003), f) designing interdependencies through a division of labor (e.g., 81 Dillenbourg and Jermann 2007; Schellens et al. 2007), and g) supporting the problem-82 solving process through several ways of scripting CSCL interactions (e.g., Ertl et al. 2007; 83 King 2007; Kobbe et al. 2007; Rummel and Spada 2007; Weinberger et al. 2007). 84

Regardless of the type of structure, it reduces the degrees of freedom in the problem-85 solving activity and is temporally administered to support learners while they are engaged 86 in solving a particular problem. Invariant across the multiple conceptions of structure is the 87 argument that structure increases the likelihood of group interactions being more 88 productive, thereby helping learners and groups accomplish what they might not otherwise 89 be able to in its absence. As noted above, much empirical work and analysis supports this 90 with research consistently showing that minimally-structured problem solving rarely leads 91to productive learning outcomes (Fischer et al. 2007; for a more general argument, see 92Kirschner et al. 2006). However, this has also led to the emergence of a deeply-ingrained 93 belief that collaborative problem solving needs to be structured a priori to support learners 94as they engage in solving complex, ill-structured tasks, for, without such structure, they 95may fail. 96

But, the belief does not necessarily imply that there is little or no efficacy embedded in 97 the *failure* that often results when groups solve complex, ill-structured problems in the 98absence of any support structures. Why? It is one thing to infer learning from observed 99 success on measures of performance. However, the conclusion that a lack of success on 100those measures implies a lack of learning does not logically follow. In other words, even if 101 A (success on performance measures) were to imply B (learning), not-A does not necessarily 102imply not-B. One is also limited by the validity and scope of the assessment measures one 103adopts (Chatterji 2003; Schwartz and Martin 2004). Furthermore, past research provides 104empirical evidence that there may very well be a hidden efficacy in ill-structured problem-105solving processes; processes that rely on generativity on the part of the learners (Mestre 1062005); processes that are often complex and divergent (Goel and Pirolli 1992); processes 107that seemingly lead to failure in the shorter term but can potentially be productive in the 108longer term (McNamara et al. 1996; McNamara 2001; Schwartz and Bransford 1998; 109Schwartz and Martin 2004; VanLehn 1999; VanLehn et al. 2003). For example, Schwartz 110and Bransford's (1998) work on preparation for future learning demonstrated that when 111 students examined similarities and differences among contrasting cases representing a target 112concept, it prepared them to derive greater benefit from a subsequent lecture or reading on 113that concept. A similar series of design experiments by Schwartz and Martin (2004) 114demonstrated a hidden potential of invention activities in the learning of descriptive 115statistics even though these activities *failed* to produce canonical conceptions and solutions 116during the invention phase. Therefore, the challenge for researchers and instructional 117designers really is to realize this potentiality. 118

Exploring productive failure

We undertook the challenge to realize the above-mentioned potentiality in a study of 120 productive failure (Kapur 2008; hereinafter referred to as the initial study). In contrast to a 121 substantive amount of CSCL research that examines students solving ill-structured 122

problems *with* the provision of various support structures, the initial study examined 123 students solving complex, ill-structured problems *without* the provision on any external 124 support structures. Because this study forms the backdrop for the confirmatory study 125 reported in this paper, we describe the initial study briefly. 126

In the initial study, 11th-grade student triads solved either ill- or well-structured physics 127problem scenarios in an online, chat environment (the design and validation of well- and ill-128structured problems is described in the following section). After participating in group 129problem solving, all students individually solved well-structured problems followed by ill-130structured problems. The analyses showed that ill-structured group discussions were 131significantly more complex and divergent than those of their well-structured counterparts, 132leading to poor group performance as evidenced by the quality of solutions produced by the 133groups. However, findings also suggested a hidden efficacy in the complex, divergent 134interactional process even though it seemingly led to failure; students from groups that 135solved ill-structured problems outperformed their counterparts from the well-structured 136condition in solving the subsequent well- and ill-structured problems individually, 137suggesting a latent productivity in the failure. 138

The argument that explains and forms the core of the productive failure hypothesis is 139simple. The structure received by students who solved ill-structured problems collabora-140tively before solving well-structured problems individually helped them make better sense 141 of the preceding ill-structured problems-a retrospective transferring effect (Marton 2007). 142At the same time, it also helped these students solve the very well-structured problems that 143provided the contrast better-an explanation consistent with what Bransford and Schwartz 144(1999) refer to as a transferring-in effect. In other words, solving the ill-structured 145problems not only influenced how students solved the well-structured problems but also 146that solving the well-structured problems retrospectively helped them make sense of the 147preceding ill-structured problems. In turn, this contrast of ill- followed by well-structured 148problems helped students discern (Marton 2007; Schwartz and Bransford 1998) how to 149structure an ill-structured problem, thereby becoming better solvers of ill-structured 150problems subsequently—a *transferring-out* effect (Bransford and Schwartz 1999). 151

Going forward

Notwithstanding a preliminary support for productive failure, the population from which 153the sample was drawn was constrained and, therefore, the findings from the initial study 154could not be extended to other contexts and settings. Hence, to assuage the concern that the 155findings from the initial study could very well be chance findings, a confirmatory study was 156necessary so as to bolster the pedagogical validity of productive failure in terms of its 157practicality and feasibility in real classroom contexts. Furthermore, while we had good 158reasons (described shortly) to choose problem structuring as the experimental manipulation 159of structure, it also limited the study's external validity to other manipulations of structure. 160Future research would do well to examine the productive failure hypothesis using other 161 operationalizations of structure. However, instead of extending productive failure across 162other operationalizations of structure, we chose to design a confirmatory study first. The 163reason was simple: If we are unable to find confirmatory evidence for productive failure in 164CSCL groups, then the question of extending the research to other operationalizations of 165structure becomes somewhat moot. 166

Our decision was further bolstered by the apparent paucity of confirmatory work in 167 educational research. It may at first seem reasonable to view confirmatory work 168

Deringer

Computer-Supported Collaborative Learning

unfavorably as not reporting "anything new." However, one must note that confirmatory 169work, albeit with its own set of issues, forms a cornerstone of scientific research (Collins 1701985); new findings are usually treated as tentative until they have been replicated or 171confirmed a number of times (Giles 2006). Of course, the situated and context-dependent 172nature of educational settings may well make confirmatory work problematic if not 173impossible (Barab and Squire 2004). However, one could also argue, as indeed we do, that 174the very context-dependence of educational settings underscores the importance of 175confirmatory work even more because such work can serve to reveal patterns and findings 176that remain *invariant* across contexts and settings. For phenomena that are especially 177complex and context-dependent, invariant patterns across contexts in and of themselves 178constitute important and new findings. 179

Finally, it is important to note that the reality of doing classroom-based research across 180multiple schools required us to adhere to the curriculum time allotted for the targeted unit. 181 This meant that the productive failure hypothesis had to be tested in a relatively short period 182of curriculum time, thereby restricting us to small manipulations. Problem structuring 183formed one such manipulation that was not only contextually-meaningful (given the heavy 184emphasis on problem solving in the curriculum) but also theoretically and empirically 185sound (Hmelo-Silver 2004; Jonassen 2000; Kapur and Kinzer 2007; Spiro et al. 1992). 186Indeed, we could have sought more substantive manipulations in terms of micro and macro 187 scripting using more sophisticated CSCL tools with built-in argumentation tools, prompts, 188 representational guidance, process scaffolds, and so on, than what the school had, but this 189would have also meant substantive undertakings in terms of teacher training and technology 190development, testing, and deployment. From a research design standpoint, it would have 191also made controlling for teacher effects across the classrooms highly problematic if not 192impossible. Instead, we chose to work within the schools' existing social and technological 193infrastructure (Bielaczyc 2006) with the hope that if the productive failure hypothesis could 194be demonstrated and replicated with minimal changes to the school curriculum, teacher 195training, and technological infrastructure, and that, too, within a relatively short timeframe, 196then it would only speak well of the productive failure design's practical significance. A 197 minimalist approach also made it easier to carry out the confirmation study in the contexts 198of two different schools. The following sections describe the confirmatory study of 199productive failure in CSCL groups. 200

Purpose

The primary purpose of the study described herein was to provide confirmatory evidence 202for productive failure. A secondary purpose was to build on the initial study and carry out 203additional analyses to further unpack the productive failure effect (described in greater 204detail in the sections that follow). For the confirmatory part, we replicated the research 205design and procedures from the initial study (Kapur 2008), including most of the group-206and individual-level analyses procedures. For the sake of completeness, the research design, 207instrumentation, and data analyses procedures are, for the most part, reproduced here in 208their entirety. 209

Participants

Participants were N=177, 11th-grade science students (120 male, 57 female) from two coeducational, English-speaking high schools in the National Capital Region of India. These 212

201

229

t1.1

schools were of similar academic standing compared to the schools in the initial study. 213Students in the science stream typically study Mathematics, Physics, Chemistry, and 214 English as their main academic subjects. The proportion of males to females in this sample 215is considered typical for the science stream in the senior secondary years (11th & 12th 216grades) in India. The school's curriculum was prescribed by the Central Board of Secondary 217Education (CBSE) of India. Using data from the 10th-grade CBSE national standardized 218test scores in science, an ANOVA did not find any significant difference between the two 219schools in terms of student ability in science, p=.227. As is typically the case, students 220came from upper-middle class families and were considered technologically savvy. The 221study was designed to reflect the schools' mathematics and science curricula. Prior to the 222study, all students had completed the curricular unit on Newtonian kinematics-the targeted 223conceptual domain of the study. It must be noted that problem solving is an integral component 224of the curricula especially in light of the high-stakes competitive entrance examinations for top 225universities in India. However, much problem solving centers on textbook-type problems, 226ranging from the simple to the very difficult. Hence, dealing with well- and ill-structured 227problem scenarios was a novel experience for both teachers and students. 228

Research design & procedures

A randomized experimental design was used. The N=177 students were first randomly 230 grouped into triads, resulting in n=59 groups. These groups were then randomly assigned 231 to one of two conditions: an ill-structured problem-solving condition (28 groups) or a well-232 structured problem-solving condition (31 groups). Table 1 shows the three phases in which 233 the study was carried out. 234

- *Phase 1*: Roughly 3 days before group work, all students individually took a 25-item multiple-choice pretest on concepts in Newtonian kinematics (*Cronbach's alpha=.78*).
 Appendix A presents three sample items.
- ii. Phase 2: Following the pretest was the collaboration phase. Groups in the ill-structured 238problem condition (IS groups) were asked to solve two ill-structured problems 239without the provision of any external support structure or scaffolds. They were 240given the ill-structured problem scenarios and then left to their own devices to 241discuss and solve the problems. Groups in the well-structured problem condition 242(WS groups) were given the same problems but in a more structured format 243(Jonassen 2000; Voss 1988, 2005; Woods et al. 1976). All problems dealt with car-244accident scenarios requiring students to apply concepts in Newtonian kinematics, 245were content validated by physics teachers, and pilot tested (for problem design and 246validation, see the following section). Each group solved two ill-structured or two 247

Phase 1 (<i>Individual</i>)	Phase 2 (Group)		Phase 3 (Individual)		t1.
Pretest	Well-structured Condition	Solve 2 well-structured problems in a counter- balanced order	Well-structured problems posttest	Ill-structured problems posttest	t1.
	Ill-structured Condition	Solve 2 ill-structured problems in a counter- balanced order			t1.

 Table 1
 The three-phase research design

Computer-Supported Collaborative Learning

well-structured problems (their order counterbalanced) as appropriate to their 248assigned condition. No other help, support structures, or scaffolds were provided to 249any group during problem solving. Phase 2 was carried out in the schools' computer 250laboratories, where group members communicated with each other only through 251synchronous, *text-only* chat much like an instant-messaging application. As such, 252the chat application did not have any additional features such as a shared 253whiteboard, visualization and simulation tools, representational tools, and so on. 254The chat application automatically archived the transcript of their discussion and 255group solutions. Groups were given 1.5 h per problem and solved both the problems 256in one seating. It is important to note that the focus of this study was not so much on 257the design of the chat environment; we wanted to leverage existing technological 258resources—a simple, text-only chat environment that students were using pervasively 259in their daily lives-to design collaboration problem solving as an instructional and 260educational activity. 261

Phase 3: The day following group work, all students individually solved well-structured iii. 262problems (WS posttest), creating a contrast for students from the IS groups. This is 263because IS group students solved ill-structured problems first, and then contrasted that 264with solving well-structured problems individually. Needless to say, the contrast 265received by IS group students can itself be seen as an external structuring mechanism. 266However, this structuring mechanism operated across the two problem-solving 267activities as opposed to operating within them, thereby setting up conditions for 268testing the hypothesis of productive failure. Finally, all students individually solved 269ill-structured problems (IS posttest). Both posttests dealt with two car accident 270scenarios each, and were content validated and pilot tested. The WS posttest was 271similar to the group problems, for which a maximum of 1.5 h were given. The IS 272posttest required students to apply more advanced concepts in Newtonian mechanics. 273A maximum of 2 h were given for the IS posttest. 274

Design and validation of problem scenarios

Design of problem scenariosThe design of ill- and well-structured problem scenarios was277closely aligned to a design typology for problems put forth by several researchers (e.g.,278Goel and Pirolli 1992; Jonassen 2000; Spiro et al. 1992; Voss 1988, 2005). Accordingly, ill-279structured problems were designed such that they possessed many problem parameters with280varying degrees of *relevance* and *specificity*. Furthermore, some of these parameters281interacted with one another in ways that allowed for multiple solutions and solution paths,282thereby making the problem intrinsically more complex.283

After designing the ill-structured problem scenarios, their well-structured counterparts were 284designed by reducing the degrees of freedom in the ill-structured problem scenarios (Jonassen 2852000; Voss 2005; Woods et al. 1976). However, the targeted content in Newtonian kinematics 286was kept the same across the well- and ill-structured problem scenarios. As a result, the well-287structured problem scenarios possessed relatively fewer problem parameters (limited 288primarily to parameters that were directly relevant to the problem) that were given with full 289specificity. Unlike in the ill-structured problems, these parameters did not interact with each 290other, thereby making the well-structured problem intrinsically simpler. As such, the well-291structured versions of the problems did not admit multiple solutions and solution paths. 292

Four problem scenarios, two well- and two ill-structured, were developed for the 293 collaborative phase (phase 2) of this study (see Appendix B for one such pair of an ill-294

275

structured and its corresponding well-structured problem). The problem scenarios were295aligned to the curricula objectives of the CBSE physics curriculum, which the participating296schools followed. Because the difference in the structure of well- and ill-structured297problems is central to the study's experimental manipulation, we exemplify this difference298by considering the ill- and well-structured problem pair in Appendix B).299

Clearly, the ill-structured problem contains many more parameters than the well-300 structured problem. For example, parameters such as weight, age, traffic conditions, prior 301 violations, and alcohol screening test, among others, are not in the problem space of the 302 well-structured problem. However, the relevance of the additional parameters varies; some 303 being more relevant to the problem than others as one would expect in an authentic 304scenario. For example, driver's weight, age, and prior violations are perhaps not as relevant 305 as traffic conditions or the alcohol screening test. Complicating the matter further, not all 306 the parameters in the ill-structured problem are known with or specified to a high degree of 307 certainty, requiring students to either deduce, simulate, or rely on assumptions, opinions, or 308 beliefs. For example, the coefficient of friction is given as a range in the ill-structured 309 problem together with a statement about the bad road conditions in the city. This is in 310contrast to the well-structured problem where the coefficient of friction is set at 0.6 without 311any further qualification. Finally, there are a greater number of interactions between the 312 parameters in the ill-structured problem than in the well-structured problem. Interactions do 313 not allow the parameters to be considered in an isolated, additive manner, making the ill-314structured problem significantly more complex than the well-structured problem. For 315example, it is easier to take the stopping distance to be 15 m, that is, the length of skid 316 marks as stated in the well-structured problem. However, the information (from the 317 mechanic's account) of the wear and tear and the status of the braking fluid interacts with 318 the length of the skid marks in the sense that the skid marks may not directly correspond to 319the stopping distance. With more wear and tear and the braking fluid running out, the 320 stopping distance is likely to be greater than the length of the skid marks. How much longer 321 is again unknown, which is precisely an example of the complexity and lack of structure 322that students engaged in while solving ill-structured problems. 323

Thus, the varying levels of parametric *relevancy* and *specification* coupled with greater 324number of parameters as well as interactions between them make the ill-structured problem 325more complex and ill-structured in comparison to the well-structured problem (Spiro et al. 326 1992). Consequently, relative to the well-structured problem, the ill-structured problem 327 admitted many more problem definitions, solutions, and solution paths as well as criteria for 328 evaluating those solutions. Note, however, that the problems were similar in their respective 329goals: both types of problems required learners to take on the same role-a lawyer-and 330 come to an evidence-based decision. Recall that the targeted content in Newtonian 331 kinematics was kept the same across the well- and ill-structured problem scenarios. 332

The ill- and well-structured problems were intentionally designed to be just beyond the 333 skills and abilities of the students. Four additional problem scenarios, two well- and two ill-334structured, were developed for the individual post-tests, that is, for the phase 3 of the study. 335 The well-structured problems on the posttest were similar to the well-structured problems in 336 the collaborative phase. Being similar to the WS group problems, the WS posttest did not 337 contain the parametric complexity (number, specification, relevancy, interactions) that 338 would somehow privilege students from the IS groups (who experienced such complexity). 339Nor did the WS posttest problem contain any additional content in Newtonian kinematics 340 that would privilege students from WS or IS groups in terms of additional content exposure. 341 Thus, the WS posttest problems can be seen as a measure of near transfer in terms of the 342 concepts and skills required to solve them. However, the two problems on the IS posttest 343

Computer-Supported Collaborative Learning

required students to use more advanced concepts than those required to solve the problems 344in the collaborative phase. These included laws of conservation of momentum, energy, and 345impact. Neither the WS nor the IS group students encountered these concepts in the 346 collaboration phase. Thus, it is reasonable to posit that exposure to content during the 347 collaboration phase could not have privileged one group over the other because this 348 additional content was not even targeted during the collaboration phase and the WS 349posttest. Hence, performance on the IS posttest was conceivably a measure of far transfer 350albeit still within the domain of Newtonian kinematics. Needless to say, we did not expect 351students to be able to solve these problems completely. What was important was that the IS 352posttest allowed for the possibility to examine if IS group students were better than their 353 WS counterparts in flexibly adapting and building on their collaborative problem-solving 354experiences to deal with the novel IS posttest problems. 355

Validation of problem scenarios Validation of all the problem scenarios was achieved in 356 multiple ways. First, two physics teachers from the schools with experience in teaching the 357 subject at the senior secondary levels (11th & 12th grades) helped content validate the 358problems. Second, a senior-secondary English language teacher from one of the schools 359 assessed the problem statements for language and readability. This was done to ensure that 360 language and readability were not confounding factors. Third, problem classification 361 validation was also undertaken by having the two physics teachers classify the problems 362 into categories. Their classification was consistent with the researchers'. Fourth, all problem 363 scenarios were iteratively validated through a small pilot study first (Kapur and Kinzer 3642007), followed by the initial study with the previous cohorts of 11th-grade science students 365from one of the participating schools. At each stage, feedback from the teachers and 366 students for content, language, and classification was incorporated to make the necessary 367 changes to the problem scenarios. These studies also informed the time allocation for group 368 and individual tasks so as to ensure that insufficient time was not a confounding factor for 369 differences between the WS and IS groups. Indeed, as expected, students in the 370 confirmatory study were able to complete the tasks, and the time stamp in the chat 371environment indicated that groups tended to make full use of the allotted time. Also, 372requests for extra time from groups were too few and far between to be of any significance. 373 Thus, time taken for all group and individual tasks was effectively treated as having been 374held constant. 375

Finally, we were also mindful of the possibility that the narrow bandwidth afforded by 376 chat communication may have differential effects for the IS and WS groups, thereby 377 confounding the study's results and findings. While this effect can never be fully mitigated, 378 careful design and validation of study's instruments did help alleviate this concern. Both the 379pilot and exploratory studies with the previous cohort of 11th-grade students and teachers 380 helped refine the problem scenarios and the study's design to mitigate this limitation as far 381as possible. Nor did these studies reveal any such differential effects. Plus, the fact that the 382chat environment used for this study was one that students used on a daily basis for 383 collaborating and chatting also worked to the study's advantage. 384

Data coding

Quantitative Content Analysis (QCA; Chi 1997) was used to segment and code utterances.387The unit of analysis was semantically defined as the function(s) that an intentional utterance388served in the problem-solving process (Suthers 2006). Thus, every utterance was segmented389into one or more interaction unit(s), and coded into categories adapted from the Functional390

419

422

428

Category System (FCS)—an interaction coding scheme developed by Poole and Holmes 391 (1995). Accordingly, each interaction unit was coded into one of *seven* categories: 392

1.	Problem Analysis (PA): Statements that define or state the causes behind a problem (e.g.,	393
	"I think the man was driving too fast"),	394

- Problem Critique (PC): Statements that evaluate problem analysis statements (e.g., 395 "how can you be sure that the man was driving fast"), 396
- Orientation (OO): Statements that attempt to orient or guide the group's process, including 397 simple repetitions of others' statements or clarifications; statements that reflect on or 398 evaluate the group's process or progress (e.g., "*let's take turns giving our opinions*"), 399
- 4. *Criteria Development* (CD): Statements that concern criteria for decision making or 400 general parameters for solutions (e.g., "we need to find the initial speed of the car"), 401
- Solution Development (SD): Suggestions of alternatives, ideas, proposals for solving 402 the problem; statements that provide detail or elaborate on a previously stated 403 alternative. They are neutral in character and provide ideas or further information about 404 alternatives (e.g., *"use the second equation of motion"*), 405
- 6. Solution Evaluation (SE): Statements that evaluate alternatives and give reasons, 406 explicit or implicit, for the evaluations; this also included statements that simply agreed 407 or disagreed with criteria development or solution suggestion statements; statements 408 that state the decision in its final form or ask for final group confirmation of the 409 decision. (e.g., "yes, but how do we get acceleration"), or 410
- Non-Task (NT): Statements that do not have anything to do with the decision task. 411 They include off-topic jokes and tangents (e.g., "let's take a break!"). 412

The pilot and exploratory studies provided the necessary training for two doctoral 413 students to independently code the interactions with an inter-rater reliability (*Krippendorff's* 414 *alpha*) of .81. The lead author and a physics teacher independently rated the quality of all 415 group solutions as well as the individual posttest performances of all students. The physics 416 teacher was blind to the treatment conditions. *Krippendorff's alphas* for rating group 417 solutions, WS posttest, and IS posttest were .85, .93, and .86, respectively. 418

Summary of data sources & measures

Before describing the data analysis procedures and methods, Table 2 summarizes the data420sources as well as the various measures that were derived from them.421

Data analysis

Data analysis was carried out at the group and the individual levels. Because the productive423failure hypothesis rested heavily on the nature of group dynamics, a multipronged group-level424analysis was undertaken to understand differences between the WS and IS groups in terms of:425

- 1. the functional content of their discussions,4262. the sequential patterns in their discussions, and427
- 3. the quality of solutions they produced as a group.

The first two measures can be seen as process measures and the third as a measure of 429 group outcome. The analysis of functional content provided a quick and dirty sense of 430 "what" the groups discussed vis-à-vis the FCS categories. For example, did the discussion 431 focus more on problem analysis and critique or solution development, and so on? Although 432

+9.1

Computer-Supported Collaborative Learning

Data Source	Derived Measure
i. 10th-grade national CBSE standardized test	Individual Science Ability (individual Science score)
scores in Science and English for each participant	 Group English proficiency (Mean English score of the group)
ii. Pretest performance score for each participant	Individual prior knowledge (individual pretest score)Group prior knowledge (Mean pretest score of
	the group)
iii. Automatically archived transcripts containing	■ Functional content of interactional activity (proportion
the problem-solving interactions of groups,	of interactional activity in the FCS categories)
including the solutions to the problems	 Sequential patterns (Lag-sequential analysis of
	significant transitions)
	Group performance (rated quality of group solutions)
iv. Well-structured posttest solution transcripts	 Well-structured problem-solving ability (rated score
of each participant	on WS posttest)
v. Ill-structured posttest solution transcripts of	■ Ill-structured problem-solving ability (rated score on
each participant	IS posttest)

 Table 2
 Summary of data sources and measures

such "coding and counting" analysis is common in CSCL research and is informative 433 (Suthers 2006), it also has its limitations. Therefore, we also examined the nature of 434 interactional sequences and patterns with the view that perhaps certain interactional 435 sequences were more likely in IS groups than in WS groups and vice versa (Barron 2003; 436 Erkens et al. 2003). For example, were attempts at problem analysis followed by more 437 problem analysis or perhaps by problem critique, and so on? Finally, analysis of group 438 performance revealed differences in the quality of solutions produced by IS and WS groups.

At the individual level, we compared the performance of students from IS and WS groups on 440the WS and IS posttests. Given the nested nature of the data set (students nested within groups), 441 we carried out the same analysis at the group level using group mean scores just to make sure 442 our results were not dependent upon the nested nature of the data set. This was a reasonable 443 strategy because we were not testing any cross-level interactions between group- and 444 individual-level constructs in our study (Snijders and Bosker's 1999). Therefore, by examining 445for consistency between group- and individual-level results, we could steer clear from 446 committing ecological or atomistic fallacies, that is, applying inferences derived from group-447 level analysis at the individual-level, and vice versa (Snijders and Bosker's 1999; Raudenbush 448 and Bryk 2002). Given the extensive nature of the analyses, variables used in the data 449analyses as well as the procedures are described together with the results in the following 450section. It is important to note that in all the results reported herein—at the group and the 451individual levels-the effects of confounding factors (e.g., counterbalanced problem order, 452school, etc.) and covariates (e.g., individual pretest score, group prior knowledge as measured 453by mean pretest score, etc.) were statistically controlled for. Recall that time-on-task, both at 454the group and individual levels, was effectively held constant by design. 455

Results and discussion

Functional content of group discussions

Controlling for the effects of school, counterbalanced problem order, group prior 458 knowledge, and group English proficiency, a MANCOVA (recall that the proportion of 459

456

0.1

479

Functional Category	Experiment 2			
	WS Groups		IS Groups	
	М	SD	M	SD
PA: Problem Analysis	.046	.022	.081*	.031
PC: Problem Critique	.032	.016	.053*	.020
OO: Orientation	.355	.128	.382	.079
CD: Criteria Development	.045	.019	.053*	.018
SD: Solution Development	.354*	.126	.272	.087
SE: Solution Evaluation	.151	.052	.143	.046

1 1 1

* denotes significance at p < .05 or better

interactional activity in the six functional categories PA, PC, OO, CD, SD, and SE formed 460the six dependent variables) revealed a significant multivariate effect of WS versus IS 461 groups on the functional content of their discussions, F(6, 50)=3.46, p=.006, partial 462 $\eta^2 = .29^1$. Table 3 presents the descriptive statistics. 463

The six univariate Levene's tests for equality of error variances were statistically not 464significant. Univariate analyses showed that IS groups had significantly greater proportion 465of activity centered on: 466

•	PA: problem analysis, $F(1, 55)=16.81$, $p<.001$, partial $\eta^2=.23$,	467
•	PC: problem critique, $F(1, 55) = 12.27$, $p = .001$, partial $n^2 = .18$, and	468

- PC: problem critique, F(1, 55)=12.27, p=.001, partial $\eta^2=.18$, and
- CD: criteria development, F(1, 55)=3.79, p=.047, partial $\eta^2=.06$. 469

In contrast, WS groups had significantly greater proportion of activity centered on: 470

SD: solution development, F(1, 55)=4.37, p=.041, partial $\eta^2=.07$. 471

There was no significant difference in the OO and SE activity between WS and IS 472groups. IS groups had greater proportion of interactional activity centered on PA, PC, and 473CD whereas WS groups had greater proportion of interactional activity centered on OO, 474SD, and SE. The descriptive trends were consistent with those found in the initial study. 475The only exception in terms of statistical significance was the interactional activity centered 476on SE, which was significantly greater for WS groups in the initial study, whereas our 477 findings above revealed that it was not the case for the present study. 478

Sequential patterns in group discussions

The above analysis only provides an indication of "what" the groups focused on, and not 480the sequential patterns in their interactions. Lag-sequential analysis² (LSA)—a technique 481increasingly being used to detect such patterns-treats each interactional unit (defined 482earlier) as an observation; a coded sequence of these observations forming the problem-483 solving sequence of a group discussion (Erkens et al. 2003). It detects the various non-484

¹ As a rule of thumb, partial η^2 =.01 is considered a small, .06 medium, and .14 a large effect size (Cohen, 1977).

² The software program Multiple Episode Protocol Analysis (MEPA) developed by Dr. Gijsbert Erkens was used for carrying out the LSA. See http://edugate.fss.uu.nl/mepa/index.htm.

Computer-Supported Collaborative Learning



Circled category: Sustained activity Arrow: Transition from one type of activity to another

Fig. 1 Likely sequential patterns in the discussions of well- versus ill-structured groups

random aspects of interactional sequences to reveal how certain types of interactions follow 485 others more often than what one would expect by chance (Wampold 1992). It accomplishes 486 this by identifying statistically significant transitions from one type of interactional activity 487 to another (Bakeman and Gottman 1997; Wampold 1992). 488

LSA revealed significant differences between the discussions of WS versus IS groups 489(see Fig. 1). In fig. 1, a circled category means that groups in that condition were at least 490*twice* as likely to sustain that type of activity, that is, the activity was at least twice as likely 491to appear in coherent clusters rather than be spread throughout the discussion. For example, 492PA was at least twice as likely to be followed by more PA in WS groups than in IS groups; 493attempts at problem analysis were followed by more problem analysis. An arrow represents 494a directed transition. For example, PA activity was at least twice as likely to be followed by 495PC activity in IS groups; attempts at problem analysis were followed by problem critique, 496which, in turn, were followed by even more critique. 497

Figure 1 suggests that with regard to how groups sustained different types of activities, 498IS groups were at least twice as likely to sustain PC and SE activities. For example, 499sequences where PC was followed by PC, and inductively, more PC, were twice as likely to 500be found in IS group discussions than in WS group discussion. In contrast, WS groups were 501at least twice as likely to sustain PA, CD, and SD activities. Note how LSA reveals 502differences in interactional patterns where the earlier analysis of functional content did not. 503For example, whereas IS groups spent a greater proportion of their interactional activity on 504problem analysis, WS groups attempts at problem analysis were more clustered together 505than spread throughout the discussion. Note how structuring the problem reproduced the 506patterns of interaction that process scaffolds typically engender, that is, helping groups 507 carry out PA, CD, SD activities in coherent phases. This lends further credence to problem 508structuring as one way in which problem-solving activities can be structured. With regard to 509transitions, there were no significant transitions that WS groups were more likely to exhibit. 510In contrast, the discussions of IS groups were more likely to exhibit many significant 511transitions (PA-PC, PA-CD, and CD-SD) as well as feedback loops (SE-PA and SE-PC).³ 512

³ It is important to note that in the initial study (Kapur, 2008), LSA analysis was triangulated through an interactional analysis of discussion excerpts explaining the various transitions and feedback loops.

Consistent with the initial study, the discussions of WS groups were more likely marked 513by interactional sequences of PA-PA-PA, CD-CD-CD, SD-SD-SD. Discussions of IS groups, 514by contrast, were more likely marked by sequences such as PA-PC-PC-PC, PC-PC-PC, 515PA-CD, PA-CD-SD, CD-SD, CD-SD-PC, CD-SD-PC-PC, SD-PC-PC, SE-SE-SE, 516SE-PA, SE-PA-PC, SE-PA-PC-PC-PC, SE-PA-CD, SE-PA-CD-SD, SE-PA-CD-SD-PC-PC-517PC-PC, SE-PC, and SE-PC-PC-PC. The greater the number of significant transitions and 518feedback loops, the greater the number of possibilities in which the discussion could unfold 519from any given point in the discussion, in turn, suggesting not only greater interactional 520complexity but also more divergent temporal trajectories. Therefore, the IS group discussions 521seem to exhibit greater divergence and complexity relative to those of WS groups. Of course, 522an intuitive way of understanding this is to realize that the greater the number of interactions 523between the components (functional categories) of a given system (group discussion), the 524greater is its complexity (Holland 1995; Kauffman 1995). Therefore, LSA suggests that the 525greater multiplicity of the solution paths, solutions, and criteria for evaluating solutions 526afforded by ill-structured problems resulted in characteristically different, and more complex 527and divergent interactional sequences, especially in the form of transitions and feedback loops 528(Kapur et al. 2007). In contrast, interactional sequences in the discussions of WS groups were 529comparatively simpler and orderly. 530

Overall, the above mentioned inferences drawn from LSA seem consistent with those found531in the initial study. It is perhaps noteworthy that the above analysis suggests an additional532significant transition: SD-PC transition was found to be significant in this study but not in the533initial study. Therefore, the interactional sequences exhibited by the IS groups in the present534study were arguably even more complex and divergent than those in the initial study.535

Group performance

The measure of group performance was operationalized as the quality of solution produced 537by the group. This was initially problematic because there were no objectively right or 538wrong answers to the problem scenarios. However, in consultation with the teacher experts, 539the strategy adopted was to focus on the extent to which groups were able to support their 540decisions through a synthesis of both qualitative and quantitative arguments, and supporting 541them with justifiable assumptions. The extent to which groups were able to accomplish this 542was rated on a scale from 0 to 4 points in units of 0.5 using a holistic rubric shown in Table 4. 543Recall that we used the same rubric across our pilot and exploratory studies. 544

An ANCOVA, F(1, 56)=4.61, p=.036, partial $\eta^2=.11$, revealed that the quality of solution 545 produced by WS groups, M=2.84, SD=1.26, was on average significantly better than that of IS groups, M=1.29, SD=1.08, controlling for group prior knowledge (see Fig. 2). 547

Discussion of group-level analyses

Differences between groups on the various process and outcome measures can be explained 549in terms of the affordances of well- versus ill-structured problems, and consequently in 550terms of the level of structure imposed on IS and WS groups. Because ill-structured 551problems do not provide a clear problem definition, IS groups spent proportionally greater 552amounts of interactional activity on problem analysis, problem critique, and criteria for 553developing a solution. This was consistent with what was found in the initial study. LSA 554further revealed that this lack of clarity in problem definition perhaps also resulted in 555sustained critiquing of attempts to analyze the problem. The larger and more complex 556solution space afforded by ill-structured problems resulted in sustained evaluation of 557

536

Computer-Supported Collaborative Learning

Quality	Description
0	Solution weakly supported, if at all
1	Solution supported in a limited way relying either on a purely quantitative or a qualitative
	argument with little, if any, discussion and justification of the assumptions made
2	Solution is only partially supported by a mix of both qualitative and quantitative arguments;
	assumptions made are not mentioned, adequately discussed, or justified to support the decision
3	Solution synthesizes both qualitative and quantitative arguments; assumptions made are not
	adequately discussed and justified to support the decision
4	Solution synthesizes both qualitative and quantitative arguments; assumptions made are
	adequately discussed and justified to support the decision

Mid-point scores of .5, 1.5, 2.5, and 3.5 were assigned when the quality of solution was assessed to be t4.8 between the major units 0, 1, 2, 3, and 4

attempts at solution development, which, in turn, fed back into problem analysis and 558critique. Thus, the discussions of IS groups were, on average, more complex, and exhibited 559greater numbers and variety of transitions and feedback loops. Consistent with the findings 560of the initial study, IS groups found it difficult to converge on the causes of the problem, set 561appropriate criteria for a solution, and actually develop a solution, resulting in poor group 562performance. WS groups, on the other hand, solved problems that offered more defined 563problem and solution spaces. Thus, their discussions were, on average, more coherent, less 564complex, and less likely to exhibit complex transitions or feedback loops. WS groups found 565it relatively easier to converge on the causes of the problem, set appropriate criteria, and 566develop a solution, which, in turn, resulted in relatively higher group performance. Thus, on 567the conventional standards of efficiency, accuracy and quality of performance, IS groups 568seemed to have *failed* compared to WS groups. 569

Individual performance

Recall that after solving either WS or IS problems in groups, each student individually 571solved WS problems (WS posttest), followed by IS problems (IS posttest). Fig. 3 shows the 572mean individual performance on the WS and IS posttests by students from WS and IS 573groups. Analyses of individual performance on WS and IS problems follows below. 574

Performance on WS posttest Controlling for the effect of individual prior knowledge, an 575ANCOVA revealed that students from IS groups, M=4.18, SD=2.07, significantly 576



EDTIT 102Rt SoB Row 05F1/2008





outperformed their counterparts from the WS groups, M=2.91, SD=2.33, on the WS 577 posttest, F(1, 159)=18.79, p<.001, partial $\eta^2=.11$. Fifteen students with missing data for 578 either the pretest or the WS posttest were excluded from the analysis. 579

As noted earlier, a limitation of this analysis is that it ignores the error correlations 580between group members. To examine whether these results hold at the group level as well, 581a similar analysis at the group level was necessary with mean WS posttest score for the 582group as the dependent variable (Snijders and Bosker's 1999; Raudenbush and Bryk 2002). 583Controlling for the effect of group prior knowledge and group solution quality, an 584ANCOVA revealed that the average WS posttest performance of IS groups, M=4.24, SD=5851.47, was significantly better than their counterparts from the WS groups, M=2.87, SD=5861.83, on the WS posttest, F(1, 55)=17.26, p<.001, partial $\eta^2=.24$. 587

Taken together, the results seemed stable across the level (individual or group) of analysis, that is, students from IS groups performed significantly better on the WS posttest than those from the WS groups. It is interesting to note that the effect of individual prior knowledge was not significant, F(1, 159)=2.08, p=.151, nor was that of group prior knowledge, F(1, 54)=.03, p=.861.

Performance on IS posttest Controlling for the effect of individual prior knowledge, an593ANCOVA revealed that students from IS groups, M=2.22, SD=1.71, significantly594outperformed their counterparts from the WS groups, M=.94, SD=1.25, on the WS595posttest, F(1, 158)=18.35, p<.001, partial $\eta^2=.10$. Importantly, individual performance on596the WS posttest was a significant predictor of individual performance on IS posttest, F(1, 158)=59730.16, p<.001, partial $\eta^2=.16$.598

Once again, a similar analysis at the group level with mean IS posttest score for the group as the dependent variable was carried out. Controlling for the effect of group prior knowledge and group solution quality, an ANCOVA revealed that the mean WS posttest performance of IS groups, M=2.34, SD=1.37, was significantly better than their counterparts from the WS groups, M=.86, SD=.94, on the WS posttest, F(1, 53)=14.10, p<.001, partial $\eta^2=.21$. Again, mean group performance on the WS posttest was a significant predictor of mean group performance on IS posttest, F(1, 53)=14.50, p<.001, partial $\eta^2=.22$.

Taken together, the results seemed stable across the level of analysis: students from IS606groups performed significantly better on the IS posttest than those from the WS groups.607Further, their performance on the WS posttest was a significant predictor of their608performance on the IS posttest. Again, the effect of individual prior knowledge was not609significant, F(1, 158)=1.90, p=.170, nor was that of group prior knowledge, F(1, 53)=.04,610p=.841.612

Computer-Supported Collaborative Learning

Discussion of individual-level results

As hypothesized, our analyses suggested that contrasting IS followed by WS problems not 614 only helped students make sense of the preceding IS problem better-a retrospective 615transferring effect (Marton 2007)-but also solve the very WS problem that provided the 616 contrast better-a transferring-in effect (Schwartz and Bransford 1998). In turn, this may 617 have helped them become better solvers of IS problems subsequently-a transferring-out 618 effect (Schwartz and Bransford 1998). All this, in the absence of the directional contrast, 619 might have remained unrealized. Therefore, despite the greater struggle, complexity, and 620 divergence in the discussions of IS groups resulting in failure in the short term, students 621from IS groups outperformed those in WS groups on the WS and IS posttests. This 622 demonstrated confirmatory evidence for the productive failure hypothesis. 623

Variation within WS and IS groups

Missing from the initial study as well as our analyses thus far is an explanation for the variance 625 of student performance within the WS and IS groups on the individual transfer measures (i.e., 626 WS and IS posttests). While, on average, IS groups exhibited productive failure, it was also 627 clear that some failed more than others in the shorter term inasmuch as some students from these 628 groups performed better than others on the transfer measures. Likewise, while, on average, WS 629 groups exhibited shorter-term success, some clearly were more successful than others inasmuch 630 as some students from these groups performed better than others on the transfer measures 631 (although, on average, not as well as those from the IS groups). What explains this variation 632 within WS and IS groups? Explaining the variance of student performance within WS and IS 633 groups on the individual transfer measures is an important issue that the initial study was not 634 designed to address, leaving it for future research to do so. 635

There are three possible candidates that could possibly explain this variance, alone or in 636 combination. It could be that some of the variance in individual transfer measures within 637 WS and IS groups could simply be due to preexisting differences in prior knowledge of the 638 students. In other words, it might be the case that students high on prior knowledge tended 639 to do better on the transfer measures than those that were not, regardless of the experimental 640 condition. However, from the analyses reported in the preceding section, variation on the 641measures of individual and group prior knowledge did not explain any variation in WS or 642IS posttest performance. Recall that none of the measures of prior knowledge—at the 643 individual or group level-had a significant effect on WS and IS posttest performance. This 644 suggests that preexisting difference in prior knowledge was not a significant factor in 645explaining variation in student performance on WS and IS posttests. 646

If it is not preexisting differences in prior knowledge, then temporality suggests that 647 either the nature of interactional content and patterns or group performance (or both) were 648 responsible for the variation. Through an analysis of the functional content and LSA 649 reported in the preceding section, we have already established how the nature of the 650 interactional dynamics differed between the two experimental conditions. Therefore, it 651remained to check if a possible source of the variation could be the level of success/failure 652experienced by the groups. In other words, did students from IS groups that were more 653 successful (i.e., produced better quality of group solutions) perform better on the individual 654transfer measures than students from the IS groups that were not as successful. Likewise, 655 did students from WS groups that were more successful perform better on the individual 656 transfer measures than students from the WS groups that were not as successful? 657

667

683

A MANCOVA did not reveal a significant effect of group performance on mean WS and 658IS posttest performance of students, F(2, 54)=1.51, p=.231, across the two conditions 659controlling for any differences in group prior knowledge. Univariate analyses further 660 revealed that the effect of group performance on mean WS posttest score was not 661 significant, F(1, 55)=.074, p=.787. Likewise, the effect of group performance on mean IS 662 posttest score was also not significant, F(1, 55)=2.74, p=.103. This suggests that students 663 from groups that produced a better quality solution did not necessarily perform better on the 664 WS and IS posttests, regardless of the experimental condition, suggesting, in turn, that it 665 was the nature and complexity of the interactional dynamics that was the important factor. 666

General discussion

This study was designed to provide confirmatory evidence for our exploratory work on 668 productive failure. The data provided strong evidence for this: first, students from IS groups 669 performed significantly better on both the WS and IS posttests compared to their 670 counterparts from the WS groups; and second, their performance on the WS posttest was 671 a significant predictor of their subsequent performance on the IS posttest. Therefore, despite 672the greater struggle, complexity, and divergence in the discussions of IS groups seemingly 673 resulting in failure in the shorter term, students from IS groups outperformed those in WS 674 groups on both the WS and IS posttests, in turn, demonstrating a confirmatory proof for 675 productive failure. 676

The rest of the discussion is organized around two major areas. First, notwithstanding 677 the fact that we sought a confirmatory study of productive failure in the work reported here, 678 we put forward a plausible explanation for the productive failure effect. Second, we 679 highlight some limitations of our study as well as the analyses. These limitations naturally 680 constrain and bound our claims but they also open up exciting lines of inquiry for future 682 research.

Explaining productive failure

At a general level, the explanation for productive failure is fairly straightforward: Learner-684 generated processes that may initially seem to fail vis-à-vis conventional standards 685 efficiency, accuracy, and performance quality may well have a hidden efficacy about them. 686 Indeed, what separated the interactional dynamics of IS from those of the WS groups was a 687 focus on problem analysis and criteria development, as well as sustained problem critique 688 and solution evaluation with a number of transitions and feedback loops. Although 689 seemingly unproductive and leading to failure in the shorter term, a more complex and 690 divergent exploration of the problem and solution spaces as evidenced by the emergence of 691 a diversity of interactional sequences was what differentiated the interactional dynamics of 692 IS groups from those of WS groups. Recall that our analyses ruled out preexisting 693 differences in prior knowledge and group performance as sources of variation of student 694 performance on the transfer measures. Instead, it was the content and interactional patterns 695 that seemed to be of primary importance. Of course, one could always argue that the 696 interactional sequences in IS groups were, in itself, a structure for students in these groups. 697 However, this structure seemed to have emerged from within as opposed to being imposed 698 from an external source (Schwartz 1995; Stahl 2007), which is precisely the point of 699 designing for productive failure. It also means that what seemed to be failure from the 700 conventional lenses of efficiency, accuracy, and quality of performance may not always be 701 Computer-Supported Collaborative Learning

an adequate gauge of learning. From an alternative lens, it is quite reasonable to suggest that the IS groups actually did not really fail, which is precisely the point of productive failure. 703

But, this still does not answer the question, "what was actually learned by IS group 704 students that WS group students did not?" A reasonable explanation comes from the notion 705of knowledge assembly (Schwartz and Bransford 1998), that is, the idea that students from 706 IS groups managed to assemble key ideas and concepts as well as relations between them 707 better than those from WS groups, especially upon solving the well-structured problems. 708 Perhaps what was happening in the complex, divergent exploration of the problem and 709 solution space was that learners were seeking to assemble or structure key ideas and concepts. 710 They did not quite accomplish this in the short term, but it possibly engendered sufficient 711knowledge differentiation that prepared them to better accomplish this upon receiving the 712 contrast in the form of a well-structured problem. How so? Answering this question brings us 713to the notion of discernability (Marton 2007; Schwartz and Bransford 1998). 714

According to Schwartz and Bransford (1998), the notion of discernability is critical to 715knowledge differentiation because "individuals learn well when they have generatively 716discerned features and structures that differentiate relevant aspects of the world" (p. 493). 717 From this perspective, the explanation that perhaps what students in IS groups learned was 718 how to structure an ill-structured problem seems most plausible. As hypothesized, solving 719the ill-structured problem influenced how they dealt with and learned from well-structured 720problems, which, in turn, helped them discern critical and relevant aspects of an ill-721 structured problem, both retrospectively and prospectively (Marton 2007). Indeed, as one of 722 the anonymous reviewers of this manuscript pointed out, the contrasting experience 723 rendered features of both well- and ill-structured problems discernable; features that 724perhaps remained indiscernible to students who had not struggled to structure problems 725 themselves. Therefore, the learning happened when the students moved from one context to 726the next (i.e., the contrast of solving ill-structured problems followed by well-structured 727 problems), not simply in the experience of what seemed to be failure. Seen this way, our 728 findings are consistent with past research that there may very well be a hidden efficacy of 729 processes that rely on generativity on the part of the learners (Mestre 2005); processes that 730seemingly lead to failure in the shorter term but engender a productive preparation for 731future learning in the longer term (Schwartz and Bransford 1998; Schwartz and Martin 732 2004). 733

Finally, the ability to perceive and structure a complex, ill-structured problem is a critical 734 dimension that seems to differentiate experts from novices, and a substantial amount of 735 research speaks to this (e.g., de Groot 1965; Chi et al. 1981; Sandberg 1994). Because 736 physics is the content domain in our two experiments, Chi et al.'s (1981) research showing 737 that physics problems are perceived differently, and consequently, are categorized and 738 represented differently by expert physicists than novice physics students, adds further 740 weight to this explanation.

Limitations and implications for future research

741

Despite having carried out a confirmatory study of productive failure in CSCL groups, it is much too early to attempt any generalization of the claims; the scope of inference technically holds only under the conditions and settings of the respective study and is, thus, circumscribed by the content domain, communication modality, age group, and cultural dispositions, among other factors. Future extensions across content topics, age groups, and cultural contexts will only add greater richness, diversity, and value to the wider ecology of CSCL research. Thus, it becomes all the more important to delimit our findings carefully. 748

Most importantly, there is no suggestion here that productive failure will hold for 749contrasts across activities that employ a different operationalization of structure. The 750experimental manipulation in our studies has been problem structuredness. The collabo-751rative problem-solving process could be structured in a myriad of other ways. For example, 752whether the productive failure effect will be applicable when structure is operationalized as 753micro and macro scripting of CSCL interactions remains an open question. Indeed, future 754research would do well to address it. In many ways, the contrast of having students solve 755 ill- followed by well-structured problems is also a kind of a script; perhaps not a micro or a 756 macro script (which are administered *within* problem-solving activities) but a design script 757 (which is administered across problem-solving activities). Additionally, one could seek 758extensions to and connections with research on scaffolding by keeping the level of structure 759in the problems the same while varying the structure in the problem-solving process 760through the provision of process scaffolds, argumentation tools, representational tools, and 761 so on. Carrying out a comparative study similar to the ones reported herein but with a 762 different form of providing structure constitutes a natural and immediate extension. Such 763 replications and extensions of productive failure research with different types of structuring 764form the focus of our current research program, specifically in the domain areas of physics 765and mathematics (e.g., Kapur et al. 2008). 766

Another limitation stems from the scope of analyses that we have focused on thus far. In 767 our efforts to demonstrate a confirmatory study of productive failure, we have mainly 768 focused on broad, macro patterns in group and individual performance. Even though we 769 carried out analysis of group interactions in terms of the functional content and emergent 770 interactional sequences, we have not quite focused just yet on in-depth, microgenetic 771 accounts of group discussions. Such microgenetic analyses, when combined with the macro 772 analyses we have already carried out, would only further unpack the complexities of the 773 phenomenon we seek to understand and explain. Examining the nature and content of 774interactional behaviors and relating them to eventual variation in and content of group and 775 individual performance-qualitatively and quantitatively-would be most insightful. For 776 example, interactional analysis might shed light on how collaborative problem solving 777 influences the development of problem-solving skills. Indeed, a plausible inference one 778 might draw from our macro analyses is that some of these skills were possibly appropriated 779 by the individual students, as evidenced on the posttest performance measures. A 780 particularly important contribution to CSCL research would be an analysis of the processes 781 of appropriation; an effort that would require qualitative analysis of the group discussions. 782However, such microgenetic analysis is only in its infancy and much more work has to be 783 accomplished before any meaningful findings emerge. 784

Another avenue for future research and analyses stems from the finding that preexisting 785differences in individual prior knowledge did not explain variation in individual 786 performance. Why is this significant? A frequent argument we have encountered is that 787 productive failure instructional designs may perhaps be more suited for higher-ability 788 learners. Indeed, such an argument finds support in a substantial amount of aptitude-789 treatment interaction (ATI) research (e.g., Cronbach and Snow 1977; Kyllonen and Lajoie 790 2003; McNamara 2001). For example, Cronbach and Snow (1977) compared the effect of 791 high versus low structure (in the form of explicit organization of information and provision 792 of learning support) in instructional presentations to find that highly-structured instruction 793 tended to benefit learners low on prior knowledge on the subject matter whereas relatively 794 less-structured instruction tended to benefit learners higher on prior knowledge. Findings 795 from our study seem to be inconsistent with past research, suggesting instead that both low 796 and high knowledge can learn from the productive failure instructional approach. That said, 797 it may just as well be that the method for modeling the effect of prior knowledge is not congruent with the nature of the effect. In other words, it is a reasonable proposition that the effect of prior knowledge may be a threshold function instead of a linear function, meaning that beyond a certain threshold, variation in prior knowledge no longer explains the effect of productive failure kinds of instructional designs. There is clearly a need to rethink the nature of the effect of prior knowledge and its interaction with instructional designs. 803

Conclusion

804

Taken together, findings from the initial and confirmatory studies suggest an important 805 implication for CSCL research; the implication being that by not overly structuring the 806 collaborative problem-solving interactions of learners and leaving them to struggle and 807 possibly even fail while solving complex, ill-structured problems can be a productive 808 exercise in failure. Our argument, however, should not be mistaken as an argument against 809 structuring the collaborative interactions of CSCL groups. Of course, and as we have said at 810 the outset, believing in the efficacy of structuring what might otherwise be a complex, 811 divergent, and unproductive process is well-placed and supported by research (Fischer et al. 812 2007; Kirschner et al. 2006). We do argue, however, that we must allow for the concomitant 813 possibility that under certain conditions even ill-structured, complex, divergent, and 814 seemingly unproductive processes have a hidden efficacy about them. Only then can CSCL 815 research systematically examine conditions under which this possibility can be realized. We 816 are not alone in advocating this, and research on productive failure only adds weight to the 817 growing number of voices—some louder than others—that have alluded to resisting an all-818 so-common, efficiency-dominant rush to overly structure CSCL interactions (Dillenbourg 819 2002; Scardamalia and Bereiter 2003; Schwartz 1995; Stahl 2007; and more generally, 820 Schwartz et al. 2005; Hatano and Inagaki 1986; Petroski 2006). Just as it is important to 821 investigate the conditions under which ill-structured problem-solving activities can be 822 structured so that they lead to productive success, it is perhaps equally important to 823 investigate the conditions under which ill-structured problem-solving activities lead to 824 productive failure in the absence of support structures. Together, the two complementary 825 lines of inquiries stand to advance the field in ways that neither of them alone can. 826

AcknowledgementsThe research reported in this paper was funded in part by the Spencer Research827Training Grant and the Education Policy Research Fellowship from Teachers College, Columbia University828to the first author. The authors would like to thank the students, teachers, and principals of the participating829schools for their support for this project. We are also grateful to David Hung, Donald J. Cunningham,830Katerine Bielaczyc, Katherine Anderson, Liam Rourke, Michael Jacobson, Rebecca Mancy, Rogers Hall,831Sarah Davis, Steven Zuiker, and John Voiklis for their insightful comments and suggestions.832

Appendix A-Sample Items from the 25-item MCQ Pretest

Two cars having different weights are traveling on a level surface with different but834constant velocities. Within the same distance, greater force will always be required to stop835the car with the greater836

(A) weight (B) velocity

(C) kinetic energy (D) momentum

A 5 kg block is resting on a rough horizontal plane. The coefficient of friction between 839 the block and the plane is 0.8. A 50 N force parallel to the plane is applied on the block for 840

🖄 Springer

833

837

845

850

851

852

853

857

10 s and then removed. The block eventually comes to a stop. Assuming $g=10 \text{ ms}^{-2}$ and 841 that the coefficient of friction does not change, the total distance traveled by the block 842 equals 843 844

(A) 20 m (B) 25 m

(C) 100 m (D) 125 m

A car starts moving from rest in a straight line with a constant acceleration of 5 ms⁻², 846 then at constant velocity, and finally decelerating at the rate of 5 ms^{-2} before coming to a 847 stop. If the total time of motion equals 5 s and the average speed for the entire motion 848 equals 4 ms⁻¹, how long does the car move at constant velocity? 849

(A) 1 s (B) 2 s

(C) 3 s (D) 4 s

Appendix B-Collaborative Phase Problem Scenarios

Ill-structured Problem 1

You have recently been hired as a lawyer for a prestigious law firm. On your first day, you 854 are sent to meet with an important client who has been fined for speeding. Opening your 855 work file, you find your assignment: 856

Dear new lawyer,

This morning, I received a call from Mr. Gupta asking me for help. According to him, he 858 almost ran over a small boy this morning in downtown Ghaziabad and was fined for 859 speeding. He insists that he was not. He says that the boy suddenly ran on to the road and 860 he braked very hard and managed to avoid an accident. However, this was enough for a 861 policeman who happened to be there to fine him Rs. 20,000 for speeding. Mr. Gupta is a 862 very important client of our firm and we must do our best to help him. I trust you will give 863 this case your best effort. I am attaching his file for your reference. 864

I am meeting with Mr. Gupta later this evening. So, I need you to investigate this 865 case and submit your report to me with your analyses and recommendation by today. 866

Sincerely, Nitin Sharma 867 Senior Partner 868 PS-Please note that the word of law is very clear on this. A person is speeding if and 869 only if he is driving above the legal speed limit of the road. No exceptions. 870

CLIENT FILE	871
Name: Mr. Amit Gupta	872
Age: 52 years	873
Driving Experience: 34 years	874
Prior Traffic Violations: 1981 (Fined for speeding, Rs. 500),	875
1993 (Fined for drunk driving, Rs. 10,000)	876

To carry out your investigation, you go through a number of steps such as a) 877 interviewing an eyewitness, b) analyzing the incident report filed by traffic police, c) 878 accessing the medical examination reports, and d) interviewing the mechanic who inspected 879 the car after the incident. 880

EYEWITNESS' ACCOUNT

"I was walking on the roadside pavement. I don't recall the traffic on the road to be 882 particularly heavy. Suddenly, I noticed a small boy run out on to the road chasing a cricket 883 ball. The next thing I heard was a loud screeching sound. I realized that it came from an 884 Ambassador car skidding to a stop in order to avoid running the boy over. The boy was 885

🖉 Springer

Computer-Supported Collaborative Learning

very lucky to have escaped any injury. I think the boy took about 3 s to cross the road, but I don't think he looked at the traffic before crossing the road. He was just chasing the ball!" TRAFFIC POLICE INCIDENT REPORT	886 887 888
Traffic conditions: Normal	889
• Weather conditions: <i>Bright and sunny: dry road</i>	890
• No evidence of a collision between the car and the boy.	891
• Number of passengers in the car besides the driver: <i>None</i>	892
• Evidence of skid marks: <i>about 15</i> m	893
• Speed limit on the road: 55kmph	894
• Width of the road: <i>about 4.5</i> m	895
MEDICAL EXAMINATION REPORT	896
General Comments:	897
Neither the driver nor the boy sustained any physical injury whatsoever.	898
Results of the car driver's medical tests	899
• BP (Blood Pressure)=110/80	900
• HR (Heart Rate)=80	901
• Weight=75 kg	902
• Reaction Time=0.8 s on an average	903
Drug/Alcohol Screen=Negative	904
MECHANIC	905
You: What can you say about the condition of the car from your inspection?	906
Mechanic: Well, this is a heavy car weighing about 1,570 kg and I can clearly see some	907
wear and tear of the tires and the braking system. The braking fluid is also running out. As	908
a result, the traction between the tires and the road does not seem to be as good as it can be.	909
You: Oh! Does this mean the car was not maintained properly?	910
Mechanic: Not really. You see, the traction also depends on the condition of the road.	911
The coefficient of friction between the car's tires and the road is usually between 0.6 and	912
0.7. So, given the city's roads, the level of traction not being as good is quite	913
understandable.	914
You: So, what are you saying?	915
Mechanic: What I'm saying is that although the traction is not as good as it could have	916
been, this is quite normal in Ghaziabad. Also, it is hard to tell how much of the wear and	917
tear happened during the skilding itself.	918
<i>You:</i> OK. Thank you for your time.	919
Well-structured version of ill-structured problem 1	920
You are a lawyer in a prestigious law firm. You've been assigned the following case:	921
A man was driving his car when, suddenly, a small boy ran out on to the road chasing a	922
ball. He slammed on the brakes and skidded to a stop, leaving a 15 m long skid mark on the	923
road. Luckily the boy was not hurt, but a policeman watching from the sidewalk walked	924
over and fined the man for speeding. An investigation found out that the speed limit on the	925
	000

road is 55kmph. It also determined that the coefficient of friction between the tires and the 926road was 0.6. The man's mass was 75 kg and his reaction time, on average, was found to be 927 about 0.8 s. The car's information manual indicated the mass of the car to be 1,570 kg. 928 Witnesses say that the boy took about 3 s to cross the 4.5 m wide road. 929

As the man's lawyer, will you fight the fine in court? Present your case as best you can. 930

References 931932 Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. The Journal of the Learning Sciences, 13(1), 1–14. 933Barron, B. (2003). When smart groups fail. The Journal of the Learning Sciences, 12(3), 307–359. 934935 Bakeman, R., & Gottman, J. M. (1997). Observing interaction: An introduction to sequential analysis. New York: Cambridge University Press. 936 Bielaczyc, K. (2006). Designing social infrastructure: Critical issues in creating learning environments with 937 938 technology. The Journal of the Learning Sciences, 15(3), 301-329. Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple 939 implications. In A. Iran-Nejad, & P. D. Pearson (Eds.), Review of research in education, 24 (pp. 61-940101). Washington, DC: American Educational Research Association. 941 942Bromme, R., Hesse, F. W., & Spada, H. (2005). Barriers and biases in computer-mediated knowledge communication-and how they may be overcome. New York, NY: Springer. 943Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational 944Researcher, 18(1), 32-42. 945946 Chatterji, M. (2003). Designing and using tools for educational assessment. Boston: Allyn & Bacon. 947 Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. The Journal of the 948 Learning Sciences, 6(3), 271–315. Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by 949experts and novices. Cognitive Science, 5, 121-152. 950Cho, K. L., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem 951solving. Educational Technology, Research and Development, 50(3), 5-22. 952953 Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press. Collins, H. (1985). Changing order. London: Sage. 954955Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods: A handbook for research on 956 interactions. New York: Irvington. 957 de Groot, A. D. (1965). Thought and choice in chess. The Hague, NL: Mouton. Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional 958design. In P. A. Kirschner (Ed.), Three worlds of CSCL. Can we support CSCL (pp. 61-91). Heerlen, 959 NL: Open Universiteit Nederland. 960 Dillenbourg, P., & Jermann, P. (2007). Designing integrative scripts. In F. Fischer, I. Kollar, H. Mandl, & J. 961962M. Haake (Eds.), Scripting computer-supported collaborative learning (pp. 275–302). New York, NY: Springer. 963 Erkens, G., Kanselaar, G., Prangsma, M., & Jaspers, J. (2003). Computer support for collaborative and 964argumentative writing. In E. De Corte, L. Verschaffel, N. Entwistle, & J. van Merrienboer (Eds.), 965Powerful learning environments: Unravelling basic components and dimensions (pp. 157–176). 966 967 Amsterdam: Elsevier Science. Ertl, B., Kopp, B., & Mandl, H. (2007). Supporting collaborative learning in videoconferencing using 968 969 collaboration scripts and content schemes. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), 970 Scripting computer-supported collaborative learning (pp. 213–236). New York, NY: Springer. 971Fischer, F., Kollar, I., Mandl, H., & Haake, J. (2007). Perspectives on collaboration scripts. In F. Fischer, H. Mandl, J. Haake, & I. Kollar (Eds.), Scripting computer-supported collaborative learning (pp. 1–10). 972New York, NY: Springer. 973Fischer, F., & Mandl, H. (2005). Knowledge convergence in computer-supported collaborative learning: The 974975 role of external representation tools. The Journal of the Learning Sciences, 14(3), 405-441. Ge, X., & Land, S. M. (2003). Scaffolding students' problem-solving processes in an ill-structured task using 976 question prompts and peer interactions. Educational Technology, Research and Development, 51(1), 21–38. 977 Giles, J. (2006). The trouble with replication. Nature, 442, 344-347. 978Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. Cognitive Science, 16, 395-429. 979Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), 980 Child Development and Education in Japan (pp. 262-272). New York: Freeman. 981Hewitt, J. (2005). Towards an understanding of how threads die in asynchronous computer conferences. The 982983 Journal of the Learning Sciences, 14(4), 567–589. 984Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? Educational 985Psychology Review, 16(3), 235–266. 986

Holland, J. H. (1995). Hidden order: How adaptation builds complexity. New York: Addison-Wesley.

Jonassen, D. H. (2000). Towards a design theory of problem solving. Educational Technology, Research and 987 Development, 48(4), 63-85. 988

Computer-Supported Collaborative Learning

Jonassen, D. H., & Kwon, H. I. (2001). Communication patterns in computer-mediated vs. face-to-face group	989
problem solving. Educational Technology, Research and Development, 49(1), 35-52.	990
Kapur, M. (2008). Productive failure. Cognition and Instruction, 26(3), 379–424.	991
Kapur, M., Dickson, L., & Toh, P. Y. (2008). Productive failure in mathematical problem solving. In B. C.	992
Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive	993
Science Society (pp. 1/1/–1/22). Austin, 1X: Cognitive Science Society.	994
Kapur, M., & Kinzer, C. (2007). The effect of problem type on collaborative problem solving in a	995 006
synchronous computer-mediated environment. <i>Laucational Technology, Research and Development, 55</i>	990
(J), +39-+39.	997
supported collaborative learning (CSCL) groups Computers and Education 51 54-66	999
Kauffman S (1995) At home in the universe New York: Oxford University Press	1000
King, A. (2007). Scripting collaborative learning processes: A cognitive perspective. In F. Fischer, I. Kollar,	1001
H. Mandl, & J. M. Haake (Eds.), Scripting computer-supported collaborative learning (pp. 13–38). New	1002
York, NY: Springer.	1003
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work:	1004
An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based	1005
teaching. Educational Psychologist, 41(2), 75–86.	1006
Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hamalainen, R., Hakkinen, P., & Fischer, F. (2007).	1007
Specifying computer-supported collaboration scripts. International Journal of Computer-Supported	1008
Collaborative Learning, 2(2–3), 211–224.	1009
Kylionen, P. C., & Lajoie, S. P. (2003). Reassessing aphrtude: introduction to a special issue in nonor of Biologie E. Gray, Educational Danabalanist 29, 70, 82	1010
Lin X Hmelo C Kinzer C & Secules T I (1900) Designing technology to support reflection	$1011 \\ 1012$
Educational Technology Research and Development 47(3) 43-62	1012 1013
Lund K Molinari G Sejourne A & Baker M (2007) How do argumentation diagrams compare when	1010
students pairs use them as a means for debate or as a tool for representing debate? International Journal	1015
of Computer-Supported Collaborative Learning, 2(2–3), 273–296.	1016
Marton, F. (2007). Sameness and difference in transfer. The Journal of the Learning Sciences, 15(4), 499–535.	1017
McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence	1018
and prior knowledge. Canadian Journal of Experimental Psychology, 55(1), 51-62.	1019
McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better?	1020
Interactions of text coherence, background knowledge, and levels of understanding in learning from text.	1021
Cognition and Instruction, 14(1), 1–43.	1022
Mestre, J. P. (2005). Transfer of learning from a modern multidisciplinary perspective. Greenwich, C1:	1023
Information Age. Mirza N. M. Tartas V. Parrat Clarmont A. & da Piatro I. (2007). Using graphical tools in a phasad	$1024 \\ 1025$
activity for enhancing diabatical skills: An example with Digalo International Journal of Computer-	1025
autivity for emained and statis from example with Digard. International Sournal of Computer Supported Collaborative Learning (iiCSCL) 2(2–3) 247–272	1020 1027
Petroski, H. (2006). Success through failure: The paradox of design. Princeton, NJ: Princeton University	1028
Press.	1029
Poole, M. S., & Holmes, M. E. (1995). Decision development in computer-assisted group decision making.	1030
Human Communications Research, 22(1), 90–127.	1031
Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical Linear Models. Thousand Oaks, CA: Sage	1032
Publications.	1033
Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing	1034
student work. The Journal of the Learning Sciences, 13(3), 423–451.	1035
Rummel, N., & Spada, H. (2007). Can people learn in computer-mediated collaboration by following a	1030
collaborative learning (pp. 39–56). New York, NY: Springer.	$1037 \\ 1038$
Sandberg, I. (1994). Human competence at work: An interpretative approach. Göteborg, Sweden: BAS.	1039
Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific	1040
explanations. Cognition and Instruction, 23(1), 23–55.	1041
Scardamalia, M., & Bereiter, C. (2003). Knowledge building. In J. W. Guthrie (Ed.), <i>Encyclopedia of</i>	1042
<i>Education</i> . New York, NY: Macmillan Reference.	1043
improve knowledge construction in asynchronous discussion groups? International Journal of	1044
Computer-Supported Collaborative Learning (iiCSCL) 2(2-3) 225–246	1046
Schwartz, D. L. (1995). The emergence of abstract dvad representations in dvad problem solving. The	1047
Journal of the Learning Sciences, 4(3), 321–354.	1048

 Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. <i>Cognition and Instruction</i>, 16(4), 475–522. Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), <i>Transfer of learning from a modern multidisciplinary perspective</i> (pp. 1–52). Greenwich, CT: Information Age Publishing. Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. <i>Cognition and Instruction</i>, 22(2), 129– 	$1049 \\ 1050 \\ 1051 \\ 1052 \\ 1053 \\ 1054$
184.	1055
Snijders, T. A. B., & Bosker, R. J. (1999). Multilevel analysis. London: Sage Publications.	1056
Spiro, R. J., Feltovich, R. P., Jacobson, M. J., & Coulson, R. L. (1992). Cognitive flexibility, constructivism,	1057
and hypertext. In T. M. Duffy, & D. H. Jonassen (Eds.), Constructivism and the technology of	1058
instruction: A conversation (pp. 1–5). Hillsdale, NJ : Erlbaum.	1059
Stahl, G. (2005). Group cognition in computer-assisted collaborative learning. Journal of Computer Assisted	1060
Learning, 21, 79–90.	1061
Stahl, G. (2007). Scripting group cognition. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.),	1062
Scripting computer-supported collaborative learning (pp. 327–336). New York, NY: Springer.	1063
Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for	1064
CSCL. International Journal of Computer-Supported Collaborative Learning (ijCSCL), 1(3), 315–337.	1065
Suthers, D. D., & Hundhausen, C. (2003). An empirical study of the effects of representational guidance on	1066
collaborative learning. The Journal of the Learning Sciences, 12(2), 183–219.	1067
VanLehn, K. (1999). Rule learning events in the acquisition of a complex skill: An evaluation of cascade.	1068
The Journal of the Learning Sciences, 8(1), /1–125.	1009
Valletini, K., Sher, S., Mulray, C., Taniauchi, I., & Baggeut, W. B. (2005). Why do only some events cause	1070
Vos I F (1088) Problem solving and reasoning in ill-structured domains. In C Antaki (Ed.) Analyzing	$1071 \\ 1072$
everyday explanation: A casebook of methods pn 74–93 London: Save Publications	1072
Voss I (2005) Toulmin's model and the solving of ill-structured problems <i>Argumentation</i> 19 321–329	1074
Wampold, B. E. (1992). The intensive examination of social interaction. In T.R. Kratochwill, & J.R. Levin	1075
(Eds.), Single-case research design and analysis: New directions for psychology and education (pp. 93–	1076
131). Hillsdale, NJ: Erlbaum.	1077
Weinberger, A., Stegmann, K., Fischer, F., & Mandl, H. (2007). Scripting argumentative knowledge	1078
construction in computer-supported learning environments. In F. Fischer, I. Kollar, H. Mandl, & J. M.	1079
Haake (Eds.), Scripting computer-supported collaborative learning (pp. 191-212). New York, NY:	1080
Springer.	1081
Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child	1082
Psychology and Psychiatry and Allied Disciplines, 17, 89–100.	1083
	1084