Assessment of (Computer-Supported) Collaborative Learning

Jan-Willem Strijbos

Abstract—Within the (Computer-Supported) Collaborative Learning (CS)CL research community, there has been an extensive dialogue on theories and perspectives on learning from collaboration, approaches to scaffold (script) the collaborative process, and most recently research methodology. In contrast, the issue of assessment of collaborative learning has received much less attention. This article discusses how assessment of collaborative learning has been addressed, provides a perspective on what could be assessed, and highlights limitations of current approaches. Since assessment of collaborative learning is a demanding experience for teachers and students alike, they require adequate computer-supported and intelligent tools for monitoring and assessment. A roadmap for the role and application of intelligent tools for assessment of (CS)CL is presented.

Index Terms—Assessment, collaborative learning, computers and education, psychology.

1 INTRODUCTION

Collaborative learning (CL) is nowadays a common practice at all levels of education. Since the 1970s small group dynamics have been intensively studied in education (for a detailed historical overview, see [1]). Research in the 1980s focused initially on face-to-face student-student interaction in primary education, but it was soon extended to secondary and higher education. Rapid developments in computer-mediated communication in the late 1980s led to a new discipline in the 1990s now referred to as Computer-Supported Collaborative Learning (CSCL).

The introduction of computers in collaborative learning induced a shift in the behaviors being studied. Initially research focused on individual cognitive learning gain (1970-1990), whereas the rise of computer technology stimulated a shift in studying how group process affects individual and group cognition (1990-present). At present, there are roughly two dominant foci in (CS)CL research: 1) understanding (successful) CL practices, and 2) determining effective conditions for (successful) CL [2], [3].

Although it is common knowledge that assessment can strongly influence learning [4], interestingly assessment of CL has thus far remained an implicit issue—despite several publications highlighting the diverse theoretical and methodological positions in (CS)CL [2], [5], [6]. In essence, what are considered relevant CL outcomes governs their 1) operationalization and 2) subsequent measurement. The assessment of CL is directly shaped by what is measured; however, it is more than merely measurement because it also contains a statement on the quality of the CL process or product in relation to presupposed criteria.

Assessment criteria, in turn, are shaped by the purpose of assessment. Broadly two purposes are distinguished: summative and formative [7], [8]. Summative assessment (also referred to as “assessment of learning”) is decontextualized and individualistic, it is isolated from the learning process, and it takes place only at the end of a course to judge how well a student performed. Summative assessment focuses strongly on the cognitive aspects of learning, often applies a single performance score, and it is designed and conducted by the teacher. Formative assessment (also referred to as “assessment for learning”) is contextualized and aims to build a comprehensive picture of learners’ characteristics. It is an integral part of a learning process and takes place several times during a course rather than only at the end. Formative assessment focuses on cognitive, social, and motivational aspects of learning, often applies a multi-method approach and it leads to a profile instead of a single score. Although distinguishing both purposes can be useful, it should be kept in mind that any assessment—when communicated to a student—involves the use of feedback information and whether the use of this information is more summative or formative is an issue of interpretation rather than one of absolutes [7], [9], [10].

Alike any other individual learning context, key to the assessment of CL is the operationalization of relevant CL outcomes, i.e., “why and what to assess.” The next section first reviews assessment in CL approaches developed in the 1970s and 1980s followed by common approaches to assessment of CL. Subsequently, four metaphors on learning in CL are discussed, and it concludes with three major challenges for assessment of CL. The third section presents an alternative framework of CL that explicitly addresses these challenges. The final section will discuss the need to align CL and its assessment, the teacher’s role in assessment of CL, the students’ role, and how technology can offer them (intelligent) tools for monitoring and assessment of CL.

2 WHY ASSESS AND WHAT TO ASSESS?

Presently, CL researchers widely agree that it is crucial for collaborative learning that all group members perform a fair...
productive interaction in any CL approach [21]. Moreover, dependence are both crucial preconditions for potential agreement that individual accountability and positive interdependence are two often voiced complaints regarding unsatisfactory assessment of CL.

2.1 Assessment and Accountability
Understanding the current CL assessment practices requires revisiting the theoretical origins of the approaches developed in the 1970s and 1980s—at the time referred to as “cooperative learning.” Two basic CL principles were introduced early-on to ensure that all members contribute to the group: individual accountability [13] and positive (goal) interdependence [14]. With respect to assessment of CL Slavin’s [15], [16] original operationalization of individual accountability is of particular interest. Slavin’s view emphasizes reward and goal structures under which students operate. More importantly, it emphasizes that group rewards enhance student learning only if group rewards are based on the individual learning of all group members. In most of Slavin’s CL methods (e.g., Student Teams Achievement Divisions (STAD), students take individual quizzes on the learning material and teams obtain a certificate based on the improvement of each group member over their past performance. In other words, the original operationalization of individual accountability was achieved through assessment (reward interdependence). More precisely, summative assessment using (standardized) tests or quizzes; and this practice reflects the assumption that all students must individually master the same learning material or skill. Hence, the issue of CL assessment is not new, but it has been historically described in terms of “rewards” and positioned as a motivational perspective on CL [15], [16]—albeit an instrumental and extrinsic perspective on motivation [17].

Whereas Slavin emphasized the reward structure and that individual students help group members because it is in their own interest, Johnson and Johnson [18], Cohen [19] and Sharan and Sharan [20]—contemporary researchers with Slavin—emphasized goal interdependence over reward interdependence, that is, students help group members because they care about the group. Nevertheless, common to all approaches developed in the 1970s and 1980s is that achievement is the principal outcome variable and typically measured in terms of “scores” on (standardized) individual tests or quizzes. It should be kept in mind that the CL methods developed in the 1970s and 1980s originally targeted primary education and were later adapted for secondary and higher education.

Cohen’s [19] seminal review reoriented CL research stressing the role of tasks and interaction over the “reward versus goal interdependence” debate. Nowadays, it is agreed that individual accountability and positive interdependence are both crucial preconditions for potential productive interaction in any CL approach [21]. Moreover, most current CL methods achieve individual accountability by other means than reward interdependence, for example, through the use of roles [22], [23], [24], [25], rather than through assessment (reward structures).

2.2 Common Approaches to Assessment of CL
Although assessment of CL gained attention in the past decade in face-to-face [26], [27] and online [29], [30], [31], [32], [33], [34] CL contexts, it is currently still

1. mostly summative,
2. designed and conducted by the teacher,
3. comprised of individual tasks (i.e., tests, quizzes, essays, etc.), a group task with each student receiving the same grade or a combination of group and individual tasks, and
4. nearly exclusively focused on cognitive outcomes (achievement).

Especially group grades and/or grading based on a mix of group and individual tasks can be problematic in a practical as well as theoretical sense.

Kagan [35] argues that group grades should never be used, despite the appeal for grading efficiency, because

1. they violate individual accountability and invite free-riding or that the most able group member conducts most (or all) of the work (“sucker effect,” see [36], [37]),
2. an individual student typically has little influence on group formation and due to the coincidental presence of high or low achieving students or a free-rider, a group grade over- or underspecifies an individual students’ competence, and equally active students may receive different grades only because of group formation outside of their control,
3. low- and medium-ability students generally profit more from group grades than their high-ability counterparts [28], and
4. unsatisfactory experiences with group grades often result in a reluctance for CL among students (and parents).

A group task combined (or supplemented) with one or more individual tasks becomes problematic when the final grade consists of the average with a weighting factor applied. In these cases, the individual tasks are used to “correct” group grades for potential social loafing effects [38] assuming that performance on individual tasks 1) directly reflects that individuals’ contribution to the group task and 2) validly compensates for a possible lack of individual effort and quality of contributions during the group task. Moreover, the approaches to such weighting vary and no clear guidelines exist. In fact, the percentage of the final grade contributed by group formation and due to the coincidental presence of high or low achieving students or a free-rider, a group grade over- or underspecifies an individual students’ competence, and equally active students may receive different grades only because of group formation outside of their control.
which a group task contributes to the final grade affects students’ preferences [38], with low-ability students being more prone to prefer a course where the group task contributes stronger to the final grade compared to the individual tasks. In addition, higher education students’ experiences with group grades affect their perceptions, i.e., students with less group grade experiences were more likely to agree that all members deserve the same grade and part-time students were more likely to consider a group grade as fair compared to full-time students [40].

Overall assessment of CL is typically conducted after the collaboration [41] and regrettably disconnected from the instructional setting, i.e., a lack of “constructive alignment” exists [42] meaning that processes and outcomes associated with CL should feature in assessment of CL [39]. However, some recent assessment formats hold a clear potential for making CL assessment more interactive and also increase student involvement, for example, 1) peer- and/or self-assessment of each individuals’ contribution to a group product, process, or both [27], [28], [43], [44] and 2) individual student portfolios signifying the most influential contributions to a knowledge building community [45], [46]. However, irrespective of a lack of constructive alignment and variability in assessment formats used, the assessment of learning in CL contexts is predominantly focused on cognitive outcomes (achievement), and not so much on social and motivational outcomes [41]. Hence, the next section reviews four metaphors on learning within (CS)CL on their 1) assumptions regarding cognitive outcomes, and 2) implicit assumptions regarding assessment.

### 2.3 Four Metaphors on Learning in CL

Following Section 1, which argued that what are being considered relevant outcomes directly governs what is assessed, the conceptualization of “learning” in CL becomes crucial. Over the last decade, four metaphors emerged in relation to learning and CL. Sfard [47] distinguished the acquisition metaphor and the participation metaphor in thinking about learning. Lipponen et al. [5] added the knowledge creation metaphor in their review of practices and orientations to CSCL. Stahl [48] subsequently proposed a socio-cultural-based group cognition metaphor (for a detailed discussion of group cognition perspectives, see [49]).

Within the acquisition metaphor learning is considered as the accumulation of knowledge in an individual mind (cognitivism; see [47]). Following this metaphor, learning is typically assessed in terms of individual knowledge gain. The assessment practice in the cooperative learning tradition (Section 2.1) is typical for this perspective given the emphasis on achievement and application of (standardized) tests to determine each individual’s improvement over their past performance (for an overview on computer-based cooperative learning, see [50]). Although uncommon in CSCL, some examples are [51], [52].

The participation metaphor stresses that learning is not solely located in an individual’s mind, but situated in a sociocultural context and constructed through interaction, discourse and participation within a community of practice [53], [54]. For some typical CSCL examples, see [55], [56], [57], [58]. Assessment of learning within this perspective focuses on how community members participate in a collaborative process of sharing and distributing expertise—without the assumption that all members must (or will) acquire the same knowledge.

Lipponen et al. [5] proposed the knowledge creation metaphor as a first attempt to rise above the acquisition and participation perspectives: “(...) the knowledge creation metaphor stresses collaborative activities [but] individual activities are also emphasized; not individuals separately, but as part of a social stream of activities” (p. 46). Furthermore, the knowledge creation metaphor stresses the advancement of understanding (transformation through knowledge building) mediated by the shared objects (artifacts) on which participants are working [5]. For some typical CSCL examples, see [45], [46], [59], [60], [61], [62].

Finally, Stahl’s [48] proposal for a socio-cultural-based group cognition metaphor “(...) may be considered as a strong form of distributed cognition” (p. 459). Whereas Salomon [63] states that “The product of the intellectual partnership that results from the distribution of cognitions across individuals or between individuals and cultural artifacts is a joint one; it can not be attributed solely to one or another partner” (emphasis added) (p. 112), Stahl [48] stresses that “The collective discourse is primary, and the individual appears as a node in the group, a contributor of pieces of the whole, and an interpreter of shared meaning” (p. 460). Stahl critiques Salomon for assuming that understanding can only be stored as a representation, thus implying a central role for the individual mind, and proposes that “(...) we talk about meaning making taking place at the small-group level, based on interpretations of the shared meaning that take place at the individual level” (p. 437). However, even though Stahl’s perspective is socio-culturally based and Salomon’s perspective is cognitive-based, both perspectives emphasize 1) the emergent property of the product of interaction, 2) that it cannot be reduced to the individual agents involved, and 3) that the emergent product is interpreted/internalized by individual agents. The difference between Stahl’s and Salomon’s perspective, however, is whether the group level or individual level is considered leading—in other words, whether the “engine for learning” is located at the group level or individual level. For some typical CSCL examples of a socio-cultural group cognition perspective, see [64], [65], [66].

### 2.4 Three Challenges for Assessment of CL

Reviewing the four metaphors on learning in (CS)CL reveals three major challenges for assessment of CL. The first challenge concerns the issue whether the assessment should focus on the individual level or the group level. The second challenge concerns whether cognitive outcomes should be operationalized in terms of students’ having the same knowledge—also referred to as convergence in (CS)CL—equivalent knowledge, similar knowledge, and/or even the extent of divergent knowledge. The third challenge is whether assessment of CL should focus solely on cognitive outcomes. Each of these challenges will be elaborated in more detail.

#### 2.4.1 Challenge 1: Individual Level versus Group Level

Since the CL approaches developed in the 1970s-1980s, the topic of whether the individual level or the group level should be assessed, that is, individual grades based on
testscores versus group grades for group projects, is debated [39]. Obviously, this is connected to the learning metaphor adopted, but it would be an oversimplification to argue that proponents of the acquisition metaphor only stress individual-level assessment and that proponents of the participation, knowledge creation, and group cognition metaphor only stress group-level assessment.

Akkerman et al. [49] identified in their review of group cognition perspectives, three boundary studies that used concepts from both cognitive and socio-cultural approaches. Yet, all of them used the individual level as their starting point. Stahl’s [48] view on group cognition takes the group level as the starting point, however, whereas Stahl views the individual as subsidiary to the group, Akkerman and colleagues call for more studies that aim to integrate both perspectives: “By either extending cognitive conceptualizations of group cognition to include social accounts of cognition or by extending socio-cultural conceptualizations of group cognition to include more precise accounts of individual cognition, we believe the studies on group cognition would be more complete” (p. 56).

Furthermore, Webb [67] concludes in discussing individual-level versus group-level assessments that the “students’ competence may be misestimated in group assessment, but some groups may produce more misestimation than others” (p. 147) and provides a powerful argument for considering both the individual and group levels: “(...) scores on work submitted from group assessment should not be used to make inferences about the competence of individual students. Without data on group processes, scores from group assessment are better interpreted as what students can produce when working with others” (p. 149). Nevertheless, evaluation criteria and group functioning appear to affect group performance, which, in turn, may positively predict individual essay performance [26].

Finally, irrespective of whether the individual level or group level is the focus of assessment, it is important to consider that most educational groups are typically “ad hoc groups”: established in relation to a specific task, series of tasks or course. To some degree, the group level is the focus of assessment, it is important to....

2.4.2 Challenge 2: Convergence versus Similarity
There is a vast array of terminology to refer to the anticipated outcome of group-level cognition. Commonly used nouns to refer to cognitive outcomes are knowledge, understanding, mental model, cognition, vision, representation, and meaning, which are commonly further specified by the following adjectives: convergent, same, shared, mutual, consensus, common, collective, equivalent, similar, and divergent [49], [68]. Since the inception of CSCL convergence has been a dominant measure and regarded as the most relevant outcome, that is, the more knowledge students have in common during or after a CL event, the more it can be assumed that CL resulted in cognitive benefits [68], [69], [70], [71]. Convergence operationalized as held in common after CL implies that students’ individual knowledge must be the same.

However, it is much more likely that what individuals learn from CL is similar rather than the same. Consider the familiar example of a group vacation. When visiting the Acropolis in Athens, the entire group shares the event. Yet, when all photos by all group members are compared afterward, some photos by different group members will be exactly the same (copying or alerting each other), some of them will be similar (same object, but different perspective), other photos will be completely different (everybody aims for their unique scoop), and it is even possible that group members have no photos in common with one or more other group members. This everyday example clearly illustrates that a shared event need not result in the same individual experience of that event. In a similar vein, a CL event need not be experienced the same by each individual.

Fischer and Mandl [69] concluded that much less convergence (in terms of shared knowledge) was observed in a CSCL context in comparison to the potential for convergence given each participants’ individual knowledge. Convergence operationalized as held in common/the same—irrespective of whether it is located in individual minds or in the interaction (discourse)—appears to be too strict as a criterion to determine whether a CL event results in learning. This led Weinberger et al. [68] to distinguish “convergence” from “equivalence,” and they defined the latter as “(...) learners becoming more similar to their learning partners with regard to the extent of their individual knowledge” (p. 417). However, “equivalent” is defined in [72] as “equal to another in status, achievement or value [emphasis added],” whereas “similar” is defined as “having qualities in common.” Hence, similarity appears more adequate because it does not imply equality in the end state or result, for example, “achievement.” Moreover, divergent processes are also in operation [70]. Each individual student may develop an internal cognitive representation different from the one they achieved as a group [73], [74]. Miyake [75] concludes with respect to individual knowledge construction from CL that: “(...) even during a highly collaborative comprehension activity, social sharing of the situation does not impede each participant from pursuing individualistic knowledge construction. Rather, the interactive process supported each to realize different perspectives to check and modify their understandings by making explicit the different perspectives, which are not within their individual repertoire” (p. 255).

In other words, knowledge that emerged during the CL event is internalized differently, given each participants prior perspective. Hence, individual students’ knowledge may become more similar due to a CL event, but simultaneously their individual knowledge may differ to a large extent “because each participant works from a different starting schema, what is obvious and natural to one may not be so to the other” (p. 464) [74].

To conclude, convergence can best be construed as the pinnacle of CL. It is the most extreme instance of similarity, or in other words, the positive extreme of the “shared knowledge” dimension ranging from convergence (same; equal; shared), via similarity (analogue; parallel; partially shared), to divergence (different; disparate; unshared). Key to assessment of CL, then, becomes to determine when the observed degree of similarity can still be attributed to CL. One approach could be to determine the degree of...
transactivity, that is, the extent to which students refer and build on each others’ contributions [68], [76] reflected in collaborative dialogue or individual products, or the extent to which students transform a shared artifact (e.g., a group report). In other words, it becomes essential to determine a situation dependent lower bound (threshold) for learning induced by a specific CL event.

2.4.3 Challenge 3: It’s Not All about Cognition

Reviewing the literature, it is apparent that cognitive outcomes are central to the assessment of learning in past and present (CS)CL studies [41], however, cognitive outcomes are not the only outcomes of CL. Slavin [16] already identified three major perspectives in cooperative learning research—the motivational, social (cohesion), and cognitive—and stated that they “(...) may be seen as complementary, not contradictory” (p. 52) and that there are many other outcomes like “(...) intergroup relations, self-esteem, acceptance of mainstreamed classmates, pro-social norms, and so on” (p. 64). Social (cohesion) aspects, such as intergroup relations, are typically emphasized in the “Learning Together” approach [18] and the “Group Investigation” approach [20].

In the context of Group Investigation, there also appear to be positive effects in relation to aspects commonly associated with intrinsic motivation, such as interest, enjoyment, and (mutual) encouragement [17].

To a certain degree, the social dimension and associated outcomes are also considered in recent literature [27], [77], [78]. Kumpulainen and Mutanen [79], for example, distinguish three dimensions in their analysis of peer group interaction: 1) functional analysis (characterizes communicative strategies used by participants in social activity), 2) cognitive processing (examines ways in which students approach and process learning tasks in their social activity), and 3) social processing (focuses on the nature of social relationships that are developed in students’ social activity). The social processing dimension describes peer group interaction in terms of collaborative, tutoring, argumentative, individualistic, domineering, conflict, and confusion. Tolmie et al. [80] recently studied the social effects of CL among 575 primary schools students (aged 9-12) which revealed that

1. CL leads to a dual impact in terms of cognitive and social gains,
2. students’ collaborative skills improve alongside understanding and optimal social relations need not be in place prior to collaboration,
3. social context (rural versus urban schools) did not affect cognitive or social gains; rather engagement in CL raises both cognitive and social gains counteracting prior social differences, and
4. convergence over time between transactive dialogue and collaborative skills (in terms of work relations) suggests that “(...) cognitive and social gains would appear to be interlinked, if distinguishable outcomes” (p. 188). In the context of CSCL, however, social interaction is still 1) often taken for granted, or 2) restricted to cognitive processes [81].

The motivational dimension and associated outcomes have also received increased attention in recent literature [82], [83], [84], [85], [86]. In contrast to an extrinsic operationalization of motivation during CL in terms of rewards (see Section 2.1), present motivational perspectives—e.g., the “dual processing self-regulation model” [87], “self-determination theory” [88], and “person-object theory” [89]—share the premise that students have multiple goals with their subsequent motivations, actions, and affective responses. Likewise, students have multiple goals and motivations in the context of CL [84], [90]. Hijzen et al. [84], for example, found that mastery goals (“I want to learn new things”) and social responsibility goals (“I want help my peers”) prevail in effective CL groups. Furthermore, belongingness goals (e.g., “I want my peers to like me”) were more important than mastery goals in ineffective CL groups, whereas the opposite was observed for effective groups. Finally, students in effective CL groups were more aware of their goals compared to students in ineffective groups [84]. In CSCL research studies on motivational processes (including the role of emotions) are under-represented although motivation research is gaining interest in CSCL research [91], [92].

Consideration of outcomes other than cognition has direct implications for the design of CL assessment [39]: “If the emphasis is on using peer learning to improve subject-matter learning, it will lead to one kind of assessment design. If the emphasis is on promoting teamwork then design for assessment will need to be quite different” (p. 419). In fact, akin to a probabilistic view on the design of CL [21], Strijbos et al. [93] argue that the design of “(...) assessment could be probabilistic as well, i.e., it should not focus on the specific learning goal X, but X’ and X” or even something unforeseen ‘U’ are equally probable collaboration outcomes that signal learning—intended or not” (p. 247). This unforeseen outcome “U” could be cognitive, but it could equally likely be social or motivational.

2.5 Conceptualizing CL in Function of Assessment

Assessment of CL has predominantly focused on cognitive outcomes in terms of achievement or knowledge convergence. Despite the consensus on the need for individual accountability and positive interdependence during CL, this is not systematically reflected in assessment of CL irrespective of available innovative assessment formats allowing for a closer alignment of CL assessment with CL processes and products. Reviewing the four metaphors for “learning” presently adopted in diverse (CS)CL studies revealed three challenges for assessment of CL: 1) the level of assessment (individual versus group), 2) operationalization of cognitive outcome (convergence versus similarity), and 3) considering other possible outcomes apart from cognitive outcomes. These three challenges for CL assessment reveal that the present metaphors for learning in CL (see Section 2.3) each emphasize 1) different aspects of a CL event, and 2) appear incomplete given their sole focus on cognitive outcomes. In response to the call by Akkerman and colleagues [49] for reconciliation between the cognitive and socio-cultural perspective, the next section introduces Group Experience as a metaphor for CL.

3 GROUP EXPERIENCE AS A METAPHOR FOR CL

Central to the proposed metaphor is the notion of learning as experiences. The role of experience for learning is historically connected to Kolb’s [94] experiential learning theory. Kolb
regards experience as the source of learning and development and defines learning as “(...) a process whereby concepts are derived from and continuously modified by experience” (p. 26). Although recent modifications of the experiential learning theory also incorporate “conversation” as a process through which the experiences can be refined and transformed [95], the actual learning from experience is still confined to an interpretative act by an individual. In contrast to Kolb’s [94] cognitive notion, Ellis and Goodyear’s [96] ecological perspective on education posits that the central role of “experience” is grounded in a relational perspective on “learning,” that is, learning is shaped through relations between the psychological (individual mind) and social (socio-cultural practices):

“An experience has to be somebody’s experience. That is, an experience involves a relationship between a person and a phenomenon. It is neither mental nor physical but a relationship between the subject and the world (...) What is experienced is important, and the characteristics of who is doing the experiencing are also important in shaping the experience. (You and I will differ in how we make sense of things. But I also act and experience things differently in different contexts)” (p. 7).

Moreover, learning as experiences implies that 1) what is learned need not be solely cognitive [31], [96], 2) what emerges as the experience is interpreted differently by who is experiencing, and 3) the experience is not confined to the individual. The next sections elaborate on the theoretical foundation of the Group Experience (GE) metaphor followed by an illustration of the metaphor.

### 3.1 Theoretical Foundations

Strijbos [97], [98] (see also [99], [100]) positions the GE metaphor as an ecological perspective on (CS)CL (for similar views, see [77], [83], [101]). It is inspired by the notions of competency-based education [102], human ecology [103], and distributed cognition [63]. In the following paragraphs each of these notions, as well as how they serve as foundations for an ecological perspective on (CS)CL and the GE metaphor, will be made explicit.

Competency-based education has become a central topic in vocational, higher, and post-secondary education [104], [105], [106] influenced by a shift toward education that more closely resembles “work” in a professional context, also referred to as learning with real events [107]. Although competencies are conceptualized in many ways [108], they include the ability to operate in ill-defined and ever-changing environments, deal with nonroutine and abstract work processes, handle decisions and responsibilities, understand dynamic systems, operate within expanding geographical horizons, and to work in groups. In general, competencies can be considered an arrangement of knowledge, skills and attitudes; or inextricably intertwined cognitive, social, and motivational processes.

Although it is widely agreed that cognitive, social, and motivational processes are important for learning in general (see Section 2.4.3 in relation to CL), there is, however, no agreement in the context of (CS)CL whether—and to what extent—they should be considered as learning enhancers, outcomes, or both. In line with Slavin’s [16] statement that these processes may be complementary, Strijbos [97], [98] (see also [99], [100]) argues that CL resides in the complex combination of cognitive, social, and motivational processes. Motivational processes, for example, may increase persistence and subsequent cognitive gains. Likewise, social processes may affect cohesion and responsibility which, in turn, facilitate cognitive gains. Finally, cognitive gains might increase a students’ sense of competence and stimulate his/her motivation. In fact, Crook [83] posits that “The quality of a collaborative encounter may depend just as much on the participants’ enthusiasm for engagement [motivational processes] as it does on their harmony of knowledge [cognitive processes], or their experience at resolving cognitive conflict [social processes].” [Emphasis added] (p. 163). Crook’s claim is supported by the fact that 1) social processes have always been an important component of the “Learning Together” approach [18], and 2) relationships between cognitive, social, and motivational processes are increasingly studied in (CS)CL research—although most studies consider only two of the three processes (see for example [76], [79], [80], [109], [110]).

An individual’s experience in terms of cognitive, social, and motivational processes, however, is not static. These processes develop over time as well as reciprocal relationships between them. The GE metaphor reflects a dynamic view on learning (and CL in particular) which is rooted in Bronfenbrenner’s [103] view of dynamic social systems (human ecology), and Salomon’s [63] conceptualization of distributed cognition.

Human ecology [103] distinguishes four social systems in which humans function: the microsystem (e.g., a small group, or a student and a teacher interacting), the mesosystem (e.g., a local school), the exosystem (e.g., the work environments of a students’ parents, which can indirectly influence the mesosystem or microsystems) and the macrosystem (e.g., the American society). Interaction between agents in each system not only affects any of the agents present, but may indirectly affect the other agents’ behaviors in other systems (see also [96]).

Hatch and Gardner [111] build on Bronfenbrenners’ view in their contextual influence model. Instead of four systems, they distinguish three levels of influence: personal, local, and cultural forces. Personal forces consist of the individuals’ abilities and experiences within a given culture. Personal forces can be affected by local forces, which are construed as the resources and people within a specific local setting, such as peers, home, parents, school, and work. Finally, cultural forces—such as institutions, practices, and beliefs—influence the local and personal forces (schooling, child rearing, language, etc.). Moreover, these forces are interdependent: “(...) no single formula determines which skills a person will use and develop. The effects of this triad of forces depend on their composition and alignment at a given site and at a given point in time” (p. 171) [111].

An ecological perspective on CL acknowledges the potential influence on the CL process and product by events outside the CL setting. In distance education, for example, students 1) vary in their educational and personal background and 2) typically have a job and a family. Thus, the impact of family and/or work events is not only imaginable but rather inevitable [77]. In the contexts of
primary or secondary education, a parents’ job loss or even divorce can have strong impact on a students’ motivation and subsequently affect their individual learning or contribution to CL. Moreover, an ecological perspective frames CL as a dynamic microsystem. Fischer and Granott [112], for example, argue that collaborative interaction unfolds at multiple levels and separate concurrent nonlinear and dynamic strands: each activity thread showing a distinctive interactive pattern different from the others.

Interestingly, distributed cognition is defined by Salomon [63] as “(...) a system that comprises an individual and peers, teachers or culturally provided tools” [emphasis added] (p. 112). The components of this definition are parallel to Hatch and Gardner’s [111] forces: “individual” (personal forces), “peers and teachers” (local forces), and “culturally provided tools” (cultural forces). Salomon stresses that a distributed cognition should not be reduced to the sum of individual cognitions, but that it emerges as a result of the interaction (see p. 112 quote in Section 2.3). More importantly, Salomon continues stating that “Each partner can still be seen as having qualities his or her own, some of which enter the distributed partnership and are affected by it reciprocally, while other qualities may not be so influenced” (p. 121). This quote conveys three important ideas for an ecological perspective on CL. First, the word “qualities” opens the possibility that Salomon perhaps also implied social and motivational processes in addition to cognition. Second, the phrase “some of which” communicates that collaborating partners do not necessarily contribute the same qualities to the interaction. Third, qualities entering the distributed partnership “are affected by it reciprocally, while other qualities may not be so influenced,” which reflects that each individual’s interpretation of a distributed cognition is different because their prior individual cognitions were different. In other words, each individual’s experience of the distributed cognition during interaction—or rather the distributed emergent experience—is interpreted differently given their prior experiences.

3.2 Group Experience (GE) as a Metaphor for CL

Fig. 1 illustrates the GE metaphor and how each interacting student subsequently internalizes his/her “distributed emergent experience” (X) differently. In general, a CL event starts at time 1 (T1) with each student contributing activities (A1 and A2) based on his/her prior respective experiences (E1 and E2). Experiences consist of cognitive (C), social (S), and motivational (M) components. As a consequence, activities that both students employ during the CL event will contain traces of the components. A CL event generates a “distributed emergent experience” (X) with cognitive,
social, and motivational features. The features of the distributed emergent experience will be internalized by each collaborating partner at time 2 (T2). However, given their individual prior experiences, all features of the distributed experience will be internalized differently: student 1 internalizes the experience as X’ with features C’, S’ and M’ and student 2 as X” with features C”, S”, and M” (for a similar systemic representation—albeit solely a cognitive account—see Cress and Kimmerle [113]).

In sum, central to the GE metaphor for CL is that

1. learning from CL occurs through distributed emergent experiences,
2. what each individual learns is more likely to be similar than exactly the same (although the latter is not excluded),
3. it can affect different processes (cognitive, social, or motivational) for each individual, and
4. the distributed emergent experience is internalized differently by each individual given his/her prior experiences—even if the same process is affected! for “(…) the experiences evoked in collaboration are highly variable and such variability is likely implicated in learning outcomes” (p. 162) [83].

The following section outlines why the GE metaphor is a more inclusive account for CL and the wide variety of CL practices as well as a more inclusive account for assessment of CL.

3.3 Benefits of the GE Metaphor for CL

There are both direct and indirect benefits of the GE metaphor. Direct benefits are that 1) the terminology of “experience” and “activities” might be acceptable to both cognitive and socio-cultural approaches, 2) there are other possible outcomes apart from cognition, and 3) outcome similarity is more likely than convergence due to each individuals’ interpretation of the distributed emergent experience. Indirect benefits relate to observations and processes that, apart from their sole focus on cognition, could not be well explained by previous metaphors.

1. Internal scripts. Conceptualizing CL as a GE implies that students’ prior experiences will affect the collaborative process, for example, via internal collaboration scripts [114], which, in turn, may interact with the external scripts (e.g., instructions) given. Carmien et al. [115] ground this interplay of internal and external scripts in a distributed cognition perspective.

2. Time. The dynamic nature of the GE metaphor explicitly includes the notion of time, allowing for a wider scope in CL research, which has typically “(…) dwell on situations that have no history” [83]. Presently, the notion of time is receiving more attention in (CS)CL in the sense of a) the sequence of actions during a CL event [116], [117] as well as b) longer term perspectives (e.g., day, week or month) [55], [56], [65], [118], [119].

3. Bridging. Sarmiento-Klapper [120] recently introduced the concept of “bridging.” Bridging activity aims at “(…) overcoming discontinuities emerging over the multiple episodes of interaction” (p. 91). In the context of Sarmiento-Klapper’s work the bridging activities involve how students cope with changes in group constellation (one or several members are missing or substituted by other students not present during previous episodes) and how these changes affect their participation. Depending on the constellation students may position themselves differently [65], for example, the absence of a dominant group member may result in a previously quiet group member taking the lead. In view of the GE metaphor, the concept of bridging provides support for the role of the individual. The individual’s interpretation of a distributed emergent experience constitutes what can be “bridged” between the CL events (or episodes). When a group performs a set of tasks in multiple constellations, the artifact(s) created (a dialogue, report) may still be available (short term), however, in the context of consecutive courses or over the span of a year (long term)—in which students collaborate in diverse and ad hoc groups—the artifacts are not easily available (a report is more likely retained as an accessible artifact than a dialogue) and unlikely to be accessed by students. Moreover, a report may contain the outcome of the CL event to some extent, but it will not convey how an individual experienced the input of fellow group members—other than what was internalized and available for recall. Hence, in long term settings participants are more likely to rely on internalized distributed emergent experiences. In sum, the concept of “bridging” helps to account for what Ellis and Goodyear [96] call “(...) the characteristics and capabilities that someone can ‘carry with them‘ from one situation to another” (p. 7).

4. (Cross-age) peer tutoring. Whereas convergence has been useful (to some extent) for the analysis of CL outcomes, it is not applicable to peer tutoring (PT). In reflecting that CL outcomes can also be divergent (knowledge shared at the pretest is not necessarily shared at the posttest) Jeong and Chi [70] refer to prior work on tutoring in which it was difficult to find traces of convergence, which could be due to the diverse roles. The limitation of the convergence measure is particularly prevalent in relation to cross-age PT, where the tutee is assumed to acquire the skills being taught by the tutor—and not the other way around. Learning for the tutee can still be conceptualized in terms of similarity or convergence, but the tutor is (in most cases) assumed to learn something completely different, for example, the skills on how to teach or explain particular subject matter to a younger student. Hence, cross-age PT implies that each student’s learning from the interaction will not be convergent, maybe similar, but most likely divergent. Moreover, as “it is common in peer learning activities for students to have differentiated roles. Their assessable products may not be the same” (p. 422) [39].

5. Help seeking. The GE metaphor is also able to account for help-seeking practices, where the social relationship (i.e., trust in the help giver’s competence) with the more able student, teacher, or parent might
increase task persistence and subsequent cognitive gain for the help seeker [121]. In terms of learning as experience, the distributed emergent experience (help seeking/receiving) is internalized in terms of social skills by the more able partner (providing help) and by the help seeker in terms of cognitive skills (solving a problem).

6. **Communities.** Community perspectives on (CS)CL and Networked Learning (NL) typically consider other processes and outcomes of CL in addition to the cognitive benefits. Membership of a particular community of learners can, for example, have a significant impact on a participants’ motivation to learn, and, in turn, this may result in the participant becoming an active contributor to the community and even helping others to learn [31], [55], [56], [65], [122]. Finally, although the GE metaphor describes a small group setting, it is precisely the fact that experiences of CL events are embedded in a larger local and cultural context that aligns the GE metaphor with community and NL approaches. Although socio-cultural perspectives are prevalent in the community-oriented studies, the terminology of “experiences” and “activities” of the GE metaphor, may help in reconciling the cognitive and socio-cultural perspectives.

The overview of indirect effects (although not intended to be exhaustive) clearly illustrates that the GE metaphor accounts for several observations and CL practices, which have traditionally been difficult to underpin using existing metaphors for learning during CL.

Conceptualizing CL as a GE might reconcile the cognitive and socio-cultural perspectives in CL research. The individual and group levels are equally important in explaining learning benefits (cognitive, social, and motivational) of CL—both in relation to a specific context (e.g., a course, task, or series of tasks) as well as in relation to long-term trajectories (e.g., how distributed emergent experiences in different courses leave a trace in each individual’s experience of a CL event). Hence, three needs for assessment of CL can be derived: 1) assess the individual and the group levels, 2) assess transformation over time—before, during, and after CL, and 3) assess multiple concurrent processes and outcomes (cognitive, social, and motivational).

The next section discusses how CL research and practices might respond to these three needs through better alignment between CL and its assessment and the application of intelligent support tools to assist teachers and students in such complex multifaceted assessment.

## 4 Intelligent Support for CL Assessment

In general, the assessment of (CS)CL focuses on individual tasks, a group task, or a combination of group and individuals tasks (Section 2.2), criteria are predefined by the teacher and the assessment is conducted by the teacher—treating the assessment process as a “black box” without any involvement of the learners [123]. Key in responding to the three needs for assessment of CL is to achieve a better alignment between the CL event and its assessment [124]. A course involving CL should apply an assessment of CL that

1. There is no generic set of agreed-upon CL indicators that can be used for assessment of CL,
2. The availability of teacher and student tools for monitoring and assessment of CL processes and products is limited [41], [128], [129],
3. If available, the information collected by these tools (e.g., most systems collect some type of logfile data) is commonly not applied for teacher and student assessment of CL, and
4. Actual teacher practices of CL assessment are sporadically investigated.

Whereas there has always been a large role for a teacher in cooperative learning, in (CS)CL research the role of the teacher was initially limited to that of a “guide on the side.” Recently, however, the role of the teacher has regained interest in relation to CL design [21], [130], [131], [132] and monitoring [41], [133], [134], [135]. In relation to assessment of CL, there are some initiatives to apply complex coding schemes developed for research to assessment of CL [136], which clearly signifies a need for transformation of research instruments into easy accessible and manageable assessment tools for teachers and students. As argued in Section 1, **assessment is measurement**—with the additional challenge that it also contains a statement on the quality of a CL process or product.

However, determining 1) what information to collect, 2) when to collect that information, and 3) how to make the information accessible to the teacher and students are merely the initial steps. Subsequent analysis of collected information, insightful representations, and effective archival and retrieval is even more challenging. This demanding task could be facilitated with intelligent support, however, in line with Salomon [137], we should “(...) let technology show us what can be done, and let educational considerations determine what will be done.” (If it ain’t technology, what is it then?). The next section presents a roadmap for
the role and application of intelligent support to assist teachers and students in the assessment of CL.

4.1 CL Mining, CL Analysis, and CL Display

Many software and online learning environments provide user interaction data as logfiles (state data) and periodical digests (change reports) [138]. The available data are typically used for research purposes and not for monitoring and assessment by the teacher or students. However, significant developments over the past decade, such as visualizing recorded interactions after CL to the students and the dynamic tracking of interactions (see for an overview [129]), enable the inclusion of more and different kinds of data for monitoring and assessment. Intelligent support will be especially useful to assist teachers in handling the huge amount of available CL data and help them to decide whether or not to intervene in a group or address an individual group member.

In relation to the application of intelligent support for monitoring and assessment of CL, three clusters can be distinguished: 1) CL mining, 2) CL analysis, and 3) CL display. Each of these clusters will be discussed in more detail, but it should be noted that this is intended as an illustrative roadmap, rather than an exhaustive review. Current practices will be addressed as well as directions and areas for further development of intelligent support for monitoring and assessment of CL—by both the teacher and the students.

4.1.1 CL Mining

1. *Mine access to system objects and student artifacts.* This refers to a wide variety of standard collaboration objects in most software and online environments (such as a discussion forum, chat, and group space). In addition, the artifacts created by students during CL can be mined, such as
   a. individual or group reports or assignments, wikipages, blogs etc.,
   b. self-, peer, and teacher ratings of the artifacts,
   c. peer and teacher feedbacks on artifacts, and
d. coconstructed whiteboard drawings, etc.

   In short, mining of artifacts in its widest sense.

2. *Mine student discourse and actions.* Whether face-to-face, computer-supported or online, CL generates enormous amounts of process data—such as written text, discourse transcripts, video, audio, spatial movement (e.g., manipulation of 3D projections, interactive surfaces such as tabletops, whiteboards, touch objects, etc.)—that can be mined for analysis. Intelligent support could assist the selection of relevant indicators, as well as automated tagging once mined. Especially for dynamic tracking of transformations (trace data; see [129]). Examples of discourse mining tools are the Analytic Toolkit in KnowledgeForum [45], and the ForumManager developed by Jeong [139] which consists of an Excel file in which threaded Blackboard discussions (version 6.x and higher) can be imported and analyzed for descriptive information (e.g., number of posting by students, number and length of threads, etc.).

3. *Mine system or instructional scripts/agents.* Scripts consist of at least five components: learning objectives, type of activities, sequencing, role distribution and type of representation [140]. Each component can be used as an indicator for automated tagging and assessment (similar to conventional structured dialogue interfaces; [141], [142], [143]). Activities associated with a specific role, for example, peer assessor or peer asesseee, can be associated with target behavior of interest. Finally, embedded agents (e.g., help prompts) can be mined for monitoring and assessment.

4.1.2 CL Analysis

1. *Integrate multiple data sources.* This is one of most challenging aspects for assessment of CL and for which intelligent support will be most welcome. Given the large (and ever expanding) variety of data that can be mined, the integration of multiple data sources is a rapidly growing necessity [144]. Dyke, Lund, and Giradot [145] developed “Trace Analysis Tool for Interaction Analysts” (Tatiana), which affords the analysis of (CS)CL multimodal process data (written discourse as well as video; including the synchronization of video and transcript) and also generates visualizations. In line with the GE metaphor, assessment of CL not only involves cognitive processes but also social and motivational processes. The latter two are typically more difficult to derive from discourse. In addition to objects, artifacts, discourse, scripts, and agents, self-report questionnaires can be used to assess students’ perception of their current CL social or motivational state during collaboration (e.g., the Quality of Working in Groups (QWIGI) inventory [82]).

2. *Analyze multiple levels simultaneously.* The analysis of individual and group-level data is inevitable for assessment of CL. In addition, the relations between both levels should be analyzed. This includes basic analyses of group participation, for example, via Social Network Analysis (SNA) to identify patterns and community member roles [56], [146]. In relation to assessment of CL, the transformation of individual experiences through internalization of group experiences (which also includes social and motivational components in addition to cognitive components) is of particular interest. The application of multilevel modeling is particularly relevant to the analysis of CL processes and products [147], [148].

3. *Analyze sequentiality and transformations over time.* At a basic level generic logfiles, trace data, and digests can be used to determine the degree of interactivity at various stages during a CL event or entire course [149]. Jeong [139], for example, also developed a Data Analysis Tool as an extension to ForumManager for sequential analysis and identification and visualization of patterns. Alternative approaches to the analysis of sequentiality and transformation are, for example, user defined start-and-stop sequence indicators [117], latent growth curves [150], dynamic...
multilevel analysis [151], event-centered analysis [119]), and uptake contingency graphs [116]. Finally, recent developments in automated coding [152], [153] offer directions for natural language processing applications for CL monitoring and assessment.

4.1.3 CL Display

1. Awareness displays. Research on awareness originated in the area of Computer-Supported Cooperative Work (CSCW). For any type of collaboration, it is essential to know “who has done what and when” to coordinate group efforts. Over the past decade research on group awareness has rapidly expanded in CSCL. Hesse [154] distinguishes three broad types of awareness information: group awareness, situation awareness, and history awareness. Various types of group awareness information and visualizations are distinguished: social awareness (“who is around?” [77]), activity awareness (“who has done what?” [155], [156]), group performance and member contribution awareness (“where do we stand as a group?” and “how have individual members contributed to the group?” [126], [157], [158] [159]), and knowledge awareness (“who knows what?” [160]); as well as combinations of several types of awareness information [126], [161]. The information on the functioning of the group and individual members can be easily obtained through self- and peer assessments to assess cognitive (e.g., productivity) and social (e.g., reliability) aspects of CL, which can subsequently be displayed and applied to further individual and group reflections (see for example [126]). Although group awareness is the most common type of awareness in CSCL, situation awareness (e.g., representation of the CL task and the available scripts), and history awareness (e.g., prior CL experiences of one self and/or fellow group members) can be powerful tools for CL assessment by the teacher and the students.

2. Dynamic monitoring and assessment display. Teachers are typically confronted with the task to simultaneously monitor multiple groups, and sometimes these groups work in different or multiple environments. What teachers need, for example, is a dynamic display to monitor and assess progress of all groups or only a subset of all groups. To this end, RSS feeds could be applied. RSS feeds are popular in browsers and blogs, but these are not yet applied to dynamically allow a teacher or student to receive updates on specific parts of the CL environment (e.g., to only receive updates from additions to a specific forum, or only updates on new documents in a specific group space). The RSS feeds could signal the teacher of a potential collaboration breakdown, or help the teacher to monitor a group s/he suspects of being close to a breakdown. Subsequently, the teacher can activate a RSS feed tied to the discussion forum (or to the entire group space) in order to receive regular updates. Obviously, dynamic addition or removal of new feeds is required and will enable the teacher to monitor a forum in group A and the shared document space in group B. Once a group functions well (again) the feed can be removed. Ideally, a monitoring panel should allow for import from various platforms, e.g., Blackboard and Moodle. Alternatively, the teacher could activate “footprint” widgets and monitor student or group activity as they move through the online environment (trace data). Finally, whereas most generic systems provide information on actions performed, for example, which students or groups submitted their assignment, information on those students or groups who did not submit their individual contribution or group assignment is equally important—with respect to these students or groups, a teacher may want close(r) monitoring or even intervene. One endeavor in the direction of dynamic “mash up” tools is the Overherd project that specifically aims to alleviate the assessment and monitoring challenges associated with large lectures and multiple forums through automated analysis and multiple visualizations of user data [162].

3. User group adaptable displays. Obviously, different user groups have different requirements for displays. For example, different visualizations might be more appropriate for different educational levels (primary, secondary, and higher education). Intelligent support could help to extract information for CL displays depending on the specific user group [162]. As for the extent of student involvement in monitoring and assessment, the purpose—i.e., assessment at the end of CL (summative) versus monitoring during CL (formative)—has strong implications for the type of information that can be displayed.

Clearly, significant strides have been made in relation to intelligent support for (CS)CL and the current state-of-the-art holds a strong potential for the future—especially for assessment of (CS)CL (for a similar argument, see [128]). However, Salomon’s [137] cautionary statement on technology—including intelligent support for CL assessment—should be kept in mind; regardless whether the CL event is face-to-face, computer-supported, or online. First and foremost, researchers will have to determine teachers’ and students’ needs for monitoring and assessment of their CL practices. This includes their perceptions of easy to navigate and manipulate displays and archives. This requires not only design of intelligent support and associated applications, but also sound interaction design and educational design. One point for departure could be the KnowledgeForum environment, which of all CSCL technologies is the most widely used by teachers in classrooms. However, teachers working with generic platforms (e.g., Blackboard or Moodle) or in face-to-face classrooms should also be queried about their needs and requirements for tools that would help them and their students in monitoring and assessment of CL.

5 CONCLUSION

Assessment of (CS)CL is a challenging endeavor. It involves multiple processes and outcomes at multiple levels by
multiple agents. Intelligent support tools can make a huge contribution to (CS)CL research and our understanding of CL practices and outcomes. In fact, these tools are crucial for a successful response to the challenges. This paper provides a roadmap for the role and application of intelligent tools for monitoring and assessment of (CS)CL. It has made no attempt to sugarcoat the challenges ahead, and given the complexity, this paper might be conceived as premature. Nevertheless, assessment is an important everyday and demanding part of the learning environment for teachers and students. There are many available, yet often unused, tools and many tools waiting to be discovered. As a research community—i.e., researchers from cooperative learning, CSCL, assessment, learning sciences, computer sciences—we can make the assessment of CL less complex, less demanding, and more worthwhile for teachers and students. We have the rudimentary components, all that is left is to figure out how they can be applied, adapted, and further developed to meet the needs of teachers and students!

REFERENCES


Jan-Willem Strijbos received the MA degree from the Radboud University Nijmegen in 1999 and the PhD degree (with honors) from the Open University of the Netherlands in 2004. From 2005 to 2009, he was a postdoctoral researcher in the Institute for Child and Education Studies at the Leiden University, and he presently works there as an assistant professor. He is a member of the Consulting Board for Computers in Human Behavior and edited special issues on topics such as “CSCL methodology” (Learning and Instruction, 2007), “peer assessment” (Learning and Instruction, 2010), and “roles in CSCL” (Computers in Human Behavior, 2010). He also edited the third volume in the Springer CSCL series, “What We Know about CSCL” (2004). His current research interests are the design of (computer-supported) collaborative learning environments, peer assessment, and discourse analysis and methodology for (CSCL) analysis and assessment. He is a member of the EARLI and the ISLS.