

From Research Instruments to Classroom Assessments: A Call for Tools to Assist Teacher Assessment of Collaborative Learning

Organizers

Jan-Willem Strijbos, Department of Psychology, Ludwig-Maximilians-University Munich, Leopoldstrasse 13, D-80802, Munich, Germany, jan-willem.strijbos@psy.lmu.de

Frank Fischer, Department of Psychology, Ludwig-Maximilians-University Munich, Leopoldstrasse 13, D-80802, Munich, Germany, frank.fischer@psy.lmu.de

Panelists

Ulrike Cress, Knowledge Construction Lab, Knowledge Media Research Center, Schleichstrasse 6, D-72076, Tuebingen, Germany, u.cress@iwm-kmrc.de

Chee-Kit Looi, National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616, Singapore, cheekit.looi@nie.edu.sg

Sadhana Puntambekar, Department of Educational Psychology, University of Wisconsin-Madison, 1025 West Johnson Street, WI 53706-1706, Madison, USA, puntambekar@education.wisc.edu

Peter Reimann, Coco Research Center, Faculty of Education and Social Work, University of Sydney, Camperdown NSW 2050, Sydney, Australia, peter.reimann@sydney.edu.au

Carolyn Rosé, Language Technologies Institute and HCI Institute, Carnegie Mellon University, 5000 Forbes Avenue, PA 15213-3891, Pittsburgh, USA, cprose@cs.cmu.edu

Jim Slotta, Ontario Institute for Studies in Education, University of Toronto, 252 Bloor St. West, Ontario M5S-1V6, Toronto, Canada, jslotta@oise.utoronto.ca

Abstract: When asked about their experiences with collaborative learning, students typically mention (a) unequal participation by students—up to free-riding, and (b) dissatisfaction with the assessment of collaborative learning. In fact, as inequality of participation increases – especially when there is a free-rider in a group – the call for diversified assessment intensifies. The topic of classroom assessment has remained implicit in wide areas of research on (CS)CL. More specifically, the issue as to how we can support teachers and students in both monitoring and assessment of collaborative learning processes and products has hardly been addressed systematically. This panel brings together researchers from the (CS)CL community to discuss and explore how research instruments can be transformed for classroom assessment purposes – including the role of technology – and what we as (CS)CL community could offer to teachers and students alike.

Assessment and Its Purpose

Assessment is the process whereby information on a students' performance is collected and interpreted. However, what are considered relevant outcomes is governed by their (a) operationalization and (b) subsequent measurement. The assessment of (computer-supported) collaborative learning (CS)CL) is directly shaped by what is measured – yet it also contains a statement on the quality of CL in relation to pre-specified criteria.

Assessment criteria, in turn, are shaped by the purpose of assessment. Broadly two purposes are distinguished: summative and formative. 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, and often applies a single performance score. Formative assessment (also referred to as 'assessment for learning') is contextualized, an integral part of the 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 the use of assessment information is an issue of interpretation. A further important distinction is the scale at which the assessment is conducted. In so-called 'large-scale assessment', the performances of hundreds up to thousands of students are assessed on carefully designed tasks, for example to determine and compared the effectiveness of school systems in different countries (PISA, TIMSS, PIRLS, etc.). In contrast, 'small-scale assessment' concerns the performance of a small number of students – typically within a single classroom (or several classrooms in a single school) – to determine the students' performance-level as well as indicators for self-group- and teacher-monitoring and need for learning support.

Irrespective of the purpose of an assessment (be it summative, formative, large-scale or small-scale) the operationalization of relevant CL outcomes is crucial to the assessment of CL. More specifically, it is the question what should be assessed, why it should be assessed, by whom, how, and when it should be assessed.

Assessment of Collaborative Learning in the Classroom

In a recent review on assessment of (CS)CL, Strijbos (2011) showed that revisiting the approaches developed in the 1970s and 1980s is highly insightful for understanding the current CL assessment practices. Slavin (1996) achieved individual accountability through reward interdependence – or more precisely – summative assessment using tests or quizzes. In fact, common to all approaches developed in that era (e.g., Jigsaw, STAD, Complex Instruction, Learning Together, Group Investigation) is that achievement is the principal outcome variable, and typically measured in terms of ‘scores’ on (standardized) individual tests or quizzes.

Although assessment of CL gained attention in the past decade in face-to-face and online CL contexts, it is currently still (a) mostly summative, (b) designed and conducted by the teacher, (c) consists 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 (d) nearly exclusively focused on cognitive outcomes. Especially group grades and/or a mix of group and individual tasks can be problematic. Despite their appeal for efficiency, Kagan (1995) argues that group grades should never be used because: (a) they violate individual accountability and invite free-riding or that the most able group member conducts most (or all) of the work, (b) 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’s competence, (c) low- and medium-ability students generally profit more from group grades than their high-ability counterparts, and (d) unsatisfactory experiences with group grades often result in a reluctance for CL among students.

A group task combined (or supplemented) with one or more individual tasks becomes problematic when a 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 and free-riding, assuming that performance on individual tasks reflects that individuals’ contribution to the group task, and validly compensates for a possible lack of individual effort and/or quality of contributions. Moreover, the approaches to such weighting vary and no clear guidelines exist. In fact, the percentage of the final grade contributed by the group task and individual tasks can each range from 10 to 90%. If the group grade is only 10% of the final grade then CL is devalued (Boud, Cohen, & Sampson, 1999), whereas if the group grade makes up 90% of the final grade free-riding is invited.

At present, assessment of CL is typically conducted after the collaboration and disconnected from the collaborative setting (i.e., a lack of ‘constructive alignment’ exists). Instead, processes and outcomes associated with CL should feature in assessment of CL. Moreover, assessment of CL contexts is predominantly focused on cognitive outcomes (achievement), and not so much on social and motivational outcomes. On the basis of the literature review on assessment of CL, Strijbos (2011) formulated three needs for assessment of CL: assessment of the individual and group-level, assessment of transformation over time – before, during and after CL, and assessment of multiple concurrent processes and outcomes (cognitive, social and motivational).

From Research Instruments to Classroom Assessment

Over the past decades the (CS)CL research community has developed and applied a wide variety of research instruments and analysis methods to uncover and understand collaborative learning processes and products (see the special issue edited by Strijbos & Fischer, 2007), ranging from Social Network Analysis (SNA), content analysis, conversation analysis, interviews, questionnaires, rating scales, log-files, scripts, roles, etc. Whereas these instruments have been pivotal for the advancement of the (CS)CL research community and our understanding of (CS)CL processes and products, the instruments largely remained tools for researchers. As (CS)CL research matures a natural transition would be to extend and/or adapt these tools for use by teachers and students for classroom assessment and monitoring purposes. Some examples of assessment methods that have been used in the context of (CS)CL are (web-based) peer assessment (PA) and portfolio assessment.

In assessment research, PA of group work has been investigated since the 1990s. Most of these studies involve the application of PA to convert a group grade to individual grades (Lejk & Wyvill, 2001). However, apart from the product (or achievement) group members’ contribution to the collaborative process can be assessed (Prins, Sluijsmans, Kirschner, & Strijbos, 2005), as well as social aspects like interpersonal relations (Phielix, Prins, & Kirschner, 2010). The rapid development in the past decade of computer supported and web-based PA systems signifies that PA is an attractive approach to assess CL. An example of student portfolio assessment of their own CL processes and products – combined with log-file data – is the research by Lee, Chan and van Aalst (2006). They applied the built-in Analytic ToolKit (ATK: component of the Knowledge Forum© environment) in combination with student portfolios to assess the knowledge building discourse within a community of learners. Finally, Meier, Spada and Rummel (2007) developed a rating format for an overall assessment of the quality of CL and they are to date the only approach to explicitly include a motivational component as part of assessing the quality of CL.

Although the literature on (CS)CL, PA and portfolio assessment shows many promising directions for the assessment of CL, the major omissions are that (a) there is no generic set of agreed-upon CL indicators that can be used for assessment of CL, (b) the availability of teacher and student tools for monitoring and assessment of CL processes and products is limited (Gress, Fior, Hadwin, & Winne, 2010), (c) if available, the information

collected by these tools (e.g., most systems collect some type of log-file data) is commonly not applied for teacher and student assessment of CL, and (d) actual teacher practices of CL assessment are sporadically investigated. In relation to assessment of CL there are some practitioner-oriented initiatives to apply complex coding schemes developed for (CS)CL research to assessment of CL (e.g., Crisp, 2007), which clearly signifies a need for transformation of research instruments into easy accessible and manageable assessment tools for teachers and students. Technology can potentially support the teacher with the assessment of collaborative learning – and specifically strides that have been made in the (CS)CL community with, for example, automated analysis and visualization – could assist teachers with this demanding task.

Panel Format and Interactivity

The panel will focus specifically on small-scale classroom assessment. To achieve a balanced representation of issues and the (CS)CL community, panelists were invited from different geographical regions, with a variety in research topics and perspectives on assessment of CL. Some of the themes to be covered are:

- How to address the practice of group grades in classrooms?
- What aspects of collaborative learning can (and ought to be) assessed?
- Does a mix of graded individual and group assignments make sense?
- How can technology assist teachers in the process of grading collaborative learning?
- What are the benefits/ drawbacks of automated assessment in a classroom context?
- How can we use artifacts/products by groups in classroom contexts to assess a group's learning?

The focus of this panel on small-scale classroom assessment and support for teacher and student monitoring and assessment, covers the topic of CL assessment in tandem with the invited panel organized by Gijsbert Erkens, Chee-Kit Looi, and Sadhana Puntambekar. Their panel focuses on large-scale assessment within the context of international PISA studies and is entitled “Can we measure collaborative skills through human-agent collaboration” (Confirmed panelists are: Chee-Kit Looi, Pierre Dillenbourg, Patrick Griffin, Jim Pellegrino and a representative of the OECD).

Interactivity

The panelists will be grouped in pairs, who will then individually prepare and briefly present their (contrasting) perspective(s) to (one or more) of the themes. The audience will be involved both prior and during the panel discussion via the web-based system “Understood It”, where we aim to collect opinions in advance and during the conference. Audience responses will be monitored and projected on a separate presentation screen during the panel discussion. The input of the audience will be used to prioritize themes for the panel discussion.

References

- Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education*, 24, 413-426.
- Crisp, G. (2007). *The e-assessment handbook*. New York: Continuum.
- Gress, C. L. Z., Fior, M., Hadwin, A. F., & Winne, P. H. (2010). Measurement and assessment in computer-supported collaborative learning. *Computers in Human Behavior*, 26, 806-814.
- Kagan, S. (1995). Group grades miss the mark. *Educational Leadership*, 52(8), 68-71.
- Lee, E. Y. C., Chan, C. K. K., & van Aalst, J. (2006). Students assessing their own collaborative knowledge building. *International Journal of Computer-Supported Collaborative Learning*, 1, 57-87.
- Lejk, M., & Wyvill, M. (2001). The effect of inclusion of self-assessment with peer-assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment and Evaluation in Higher Education*, 26, 551-561.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration process. *International Journal of Computer-Supported Collaborative Learning*, 2, 63-86.
- Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL environment: Effects of peer feedback and reflection. *Computers in Human Behavior*, 26, 151-161.
- Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. W. (2005). Formative peer assessment in a CSCL environment. *Assessment and Evaluation in Higher Education*, 30, 417-444.
- Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, 21, 43-69.
- Strijbos, J.W. (2011). Assessment of (computer-supported) collaborative learning. *IEEE Transactions on Learning Technologies*, 4(1), 59-73.
- Strijbos, J. W., & Fischer, F. (2007). Methodological challenges for collaborative learning research. *Learning and Instruction*, 17(4), 389-394.