Research and Practice in Technology Enhanced Learning
Vol. 3, No. 1 (2008) 1–30

ASSESSING THE ASSESSORS: ADDED VALUE IN WEB-BASED MULTI-CYCLE PEER ASSESSMENT IN HIGHER EDUCATION

YAEL KALI

Department of Education in Technology and Science
Technion – Israel Institute of Technology
Haifa, 32000, Israel
yaelk@technion.ac.il

MIKY RONEN

Department of Instructional Systems Technologies
Holon Institute of Technology, Holon, 52108, Israel
ronen@hit.ac.il

Peer assessment has been shown as a productive strategy for supporting higher education students to learn from each other, yet it also evokes technical, pedagogical, and emotional challenges. This study seeks to benefit from the affordances of peer assessment while reducing typical tensions. Assessing-the-Assessors is an innovative approach to design and conduct web-supported multi-cycle peer assessment activities. The main notion in this approach is that peer assessment is conducted as an exercise to enhance learning; students are rewarded as assessors rather than assessees, while the formal grades for artifacts are provided by the instructor. This approach was implemented in 25 higher-education courses in various contexts with 848 students. Findings indicate that courses with multiple cycles of assessment were highly effective in promoting learning in three aspects: learners, instructors and social climate. Learners refined their artifacts, improved their ability to provide assessment, and increased their confidence in their ability to serve as assessors. Instructors were able to easily distinguish between student bias and their own bias, and to handle it appropriately. Social climate was productive; students' attitudes were positive and their commitment was high. We recommend instructors to utilize the Assessing-the-Assessors approach for increasing the benefits of peer assessment in higher education.

Keywords: Peer assessment; web-based instruction; higher education.

1. Introduction

The past decades of research, since Vygotsky introduced his socio-constructivist views about how people learn (Vygotsky, 1978), have continued to show that learning from peers is a critical factor in the process of human learning (Cobb, 1994; Cole, 1996; Gutierrez & Rogoff, 2003; Lave & Wenger, 1991). More recently, as technology has permeated to all levels of educational systems, researchers have

shown that technology can greatly support collaborative peer learning in various manners (Bransford, Brown, & Cocking, 1999; Kali & Linn, 2007; Linn, Davis, & Bell, 2004; Salomon & Perkins, 1996; Scardamalia & Bereiter, 1996; Stahl, 2006; Stahl, Koshmann, & Suthers, 2006).

Nonetheless, instruction in higher education whether supported by technology or not has usually stayed traditional, and only seldom applies socio-constructivist instructional approaches (Herrington, Reeves, & Oliver, 2005; Levin-Peled, Kali, & Dori, 2007; McCray, DeHaan, & Schuck, 2003). Thus, one of the greatest potential resources for learning available at hand in any class, even when budgets are kept low, i.e. the students themselves, are usually neglected. Peer assessment is one of the strategies which has been shown to successfully utilize this valuable resource. When students assess each other's work, they become more aware of criteria for what accounts as "good quality" in a knowledge domain. As a result, they reflect on and become more attentive to the way they apply these criteria to their own work. In this way, they not only better understand the contents in a knowledge domain, but also develop their metacognitive skills, which can support their learning in other knowledge domains (Pellegrino, Chudowsky, & Glaser, 2001; Stefani, 1994; White & Frederiksen, 2000; Zariski, 1996). However, despite the benefits of peer assessment to student learning, as we describe below, there are technical, pedagogical, emotional and ethical challenges and tensions involved in the implementation of peer assessment (Kali, 2006; Kali & Ronen, 2005; Zariski, 1996). This research describes a model of employing peer assessment in a way that takes advantage of its potential benefits to support learning on one hand, while reducing tensions on the other.

Many researchers have studied peer assessment activities in various higher education contexts (Boud, Cohen, & Sampson, 2001; Brown & Glasner, 1999; Davies, 2004; Falchikov, 2003; Freeman, 1995; Kali & Ronen, 2005; Liu & Yuan, 2002; McConnell, 2002; Miller, 2003; Smith, Cooper, & Lancaster, 2002). The focus of the research in these studies varies from the exploration of student learning gains and attitudes toward peer assessment to examining the appropriateness of using scores from peer assessment for official student grades. In the past decade, researchers have begun synthesizing the body of knowledge gained in this field. Two prominent syntheses are those reported by Topping (1998) and by Falchikov & Goldfinch (2000). The first is a comprehensive review and characterization of 109 studies from which a typology for peer assessment activities in higher education was developed. The second is a meta-analysis of 48 studies, which quantitatively examines the validity of peer assessment scores. In the current research, we build on these syntheses as well as on findings from other research projects to introduce a new approach for peer assessment, which, as we describe below, can increase the benefits for both learners and instructors and reduce tensions such as those reported in the literature.

1.1. *Peer assessment dimensions*

Topping (1998) defines peer assessment as "an arrangement in which individuals consider the amount, level, value, worth, quality or success of the products or

outcomes of learning of peers of similar status" (Topping, 1998, p. 250). In his typology, he refers to 17 dimensions in peer assessment, which emerged from the research studies he reviewed. The dimensions are: (1) Curriculum area (the contents for which the peer assessment activity was designed for); (2) Objectives of the activity (whether a pedagogical objective, or economical one such as saving staff time by relying on peer assessment for students' final scores); (3) Focus (summative or formative orientation of assessment); (4) The Product assessed (tests, oral presentations etc.); (5) Relation to "official" staff assessment (whether scores were substitutional or supplementary); (6) Official weight (the degree to which scores contribute to assessee final official grade; (7) Directionality of peer assessment (unidirectional assessor-assessee, reciprocal, or mutual); (8) Privacy (whether assessment was done anonymously, partially anonymously, or public); (9) Personal Contact (varied from oral assessment of public presentations to assessment of written products from distance); (10) Synchronicity (whether peer assessment was done mutually between peers, or on products from previous years); (11) Ability (same, or cross ability); (12) Constellation of Assessors (individual, pairs or groups); (13) Constellation of Assessees (individual, pairs or groups); (14) Place (in or out of class); (15) Time (class time, free time or informally); (16) Requirement (whether compulsory or voluntary for assessors and assessees); (17) Reward (course credit or other incentives or reinforcement for participation). This typology is used in the current research to characterize the peer assessment activities involved in the study.

1.2. *Advantages of peer assessment*

Many examples of empirical studies indicate that peer assessment as an instructional approach has a great impact on student learning, and that learning is fostered by having students serve both as assessors and as assesses (Dominick, Reilly, & J., 1997; Zariski, 1996). Specifically, researchers illustrate how peer assessment assists students to create higher quality artifacts as a consequence of better understanding of assessment criteria, which they use when they play the role of assessors (Falchikov, 2003; Smith *et al.*, 2002), and of greater reflection of the way they apply these criteria on their own artifacts (Stefani, 1994; White & Frederiksen, 2000). Researchers also emphasize the benefit of peer assessment, which is provided by the opportunity to learn from artifacts created by their peers (Ronen & Langley, 2004). Another advantage stems from the input students receive from other learners, which in some cases complements and even exceeds the instructor's assessment (e.g., Zariski, 1996).

Topping's review (1998), supports this trajectory, and indicates that there are many benefits to peer assessment for both the assessors and assessees. Advantages for assessors include consolidation, reinforcement and deepening of their own understanding as a result of engaging in activities such as reviewing, summarizing, clarifying, giving feedback, diagnosing misconceived knowledge, identifying missing knowledge and considering deviations from the ideal. Advantages for those being assessed include better understanding of what is considered high quality work in

a specific context and attention to important aspects of the assessed assignment according to agreed upon criteria. The review also indicates systemic benefits, which include students' higher appreciation and trust in assessment provided to them by their instructors as a result of experiencing both roles of assessors and assessees.

It is clear from Topping's review (1998) and from the literature reviewed above that students can greatly profit from peer assessment when activities are designed appropriately. The meta-analysis by Falchikov and Goldfinch (2000) further supports this trajectory by examining the validity of peer assessment via quantitative analysis of the data in their review. Their analysis indicates that the mean correlation over all the studies was $r = 0.69$ (on average), showing definite evidence of agreement between peer and instructor scores. Their findings can provide general guidelines as to what "appropriate design" of peer assessment activities might be. One of their most prominent findings regarding design of activities is that students' familiarity with and ownership of criteria tends to enhance peer assessment validity.

1.3. *Technology benefits*

One of the difficulties in implementing peer assessment is the tremendous workload it requires from instructors. Coordinating a peer assessment activity can involve hundreds of student-provided assessments. Examination of these inputs, providing immediate feedback to assessors, conducting quantitative or qualitative analysis, and representation of these outcomes to students are some of the aspects which instructors are required to cope with. Technology can play a crucial role in the facilitation of all these aspects, offering the ability to effectively implement and coordinate peer assessment activities. Indeed, various computerized systems were recently designed to support peer assessment (Davies, 2004; Liu & Yuan, 2002; Mann, 1999; McConnell, 2002). These systems enable students to publish their artifacts online, submit their assessment to their peers, and view the processed quantitative analyses, as well as the pool of written assessments from their peers. Many of these systems provide instructors with tools to create, organize and run their peer assessment activities.

1.4. *Challenges of conducting effective peer assessment*

One of the main challenges in implementing peer assessment is tensions created by student reluctance to assess their peers. This is especially true if the assessment affects final scores (Zariski, 1996), or if the assessed artifact is socially, culturally or personally sensitive (Kali & Ronen, 2005). Additionally, in many cases, students feel that assessing their peers is beyond the scope of their capabilities and responsibilities (Zariski, 1996). Several studies have sought to justify the use of scores compiled from peer assessment for final course grades. A common strategy for such justification is based on examining the correlations between peer scoring, and scores provided by instructors — high correlations in this strategy, justify the use of peer

assessment scores for formal grading of students. However, Flachikov & Goldfinch (2000) emphasize that although on average such correlations tend to be rather high ($r = 0.69$), there are large variations which need to be considered. Some researchers (e.g. Liu &Yuan, 2002) have dealt with this issue by conducting further statistical analysis procedures of scores compiled from peer assessment that reduce variations and bring these scores closer to the instructor's scores.

Recently, researchers have begun paying more attention to an aspect in peer assessment, which traditionally received very little thought, and refers to the 17th dimension from Topping's (1998) typology, i.e. reward. Rewarding students according to the quality of assessments they provide to their peers has been suggested by Davies (2004). A case study he conducted with computer science students showed that the quality of assessments was correlated with scores compiled from essay-writing in this subject area. He concluded that awarding a 'mark for marking' is an appropriate means of assessment, which can even replace other means of assessing students. This approach has the potential to reduce the tensions described above by shifting the focus of using scores compiled from peer assessment, from the assessees to the assessors.

2. The Current Study

In this study we present an innovative approach, which we call Assessing-the-Assessors (AtA) to design and conduct web-supported multi-cycle peer assessment activities. We use the term peer assessment to refer to activities in which students assess their peers' artifacts. One of the main notions of the AtA approach is that the peer assessment activity is conducted as an exercise to enhance learning and promote skills. Thus, scores compiled from this exercise are not used to determine grades for assessees. Instead, they are used to help assessors improve their ability to provide assessments of high quality in a specific context, and thus deepen their understanding of the contents involved. In terms of Topping's (1998) typology, the objectives are purely pedagogical and do not include any economic aspect in which the peer assessment saves staff time by replacing portions of the instructor's scores.

Via the AtA approach, we seek to take advantage of the known benefits of peer assessment to enhance learning and increase these benefits by:

- Minimizing tensions that are usually involved in peer assessment and thus increase learning benefits.
- Increasing students' confidence in their own capability to assess their peers.
- Providing teachers with tools to better understand their students' learning processes and enable them to reflect on their own assessment of student work.

The goal of the research is to explore the value of the AtA approach according to these aspects.

2.1. *The AtA model*

To apply the AtA approach, we developed a model with the following stages (Figure 1).

(1) Students are prepared for the activity. The preparation involves presenting, discussing or creating with students the criteria for assessing the artifacts. In some cases, students are provided with example artifacts created by students from former course enactments (Ronen & Langley, 2004) or with other scaffolds such as templates, resources etc. Students are informed in advance that the grades assigned to them by their peers will not affect their formal grade on the course but that a significant portion of their formal grade for the course will be derived from the quality of assessment they provide to their peers. Ways of providing constructive feedback are discussed and norms are created for appropriate use of language (Kali & Ronen, 2005; Kali, 2006).
(2) Students create the artifacts. Artifacts can vary from short assignments that can be prepared in one week to large projects created throughout the semester. When several cycles of assessments are enacted (Figure 1), instructors may enable students to revise their artifacts based on the published

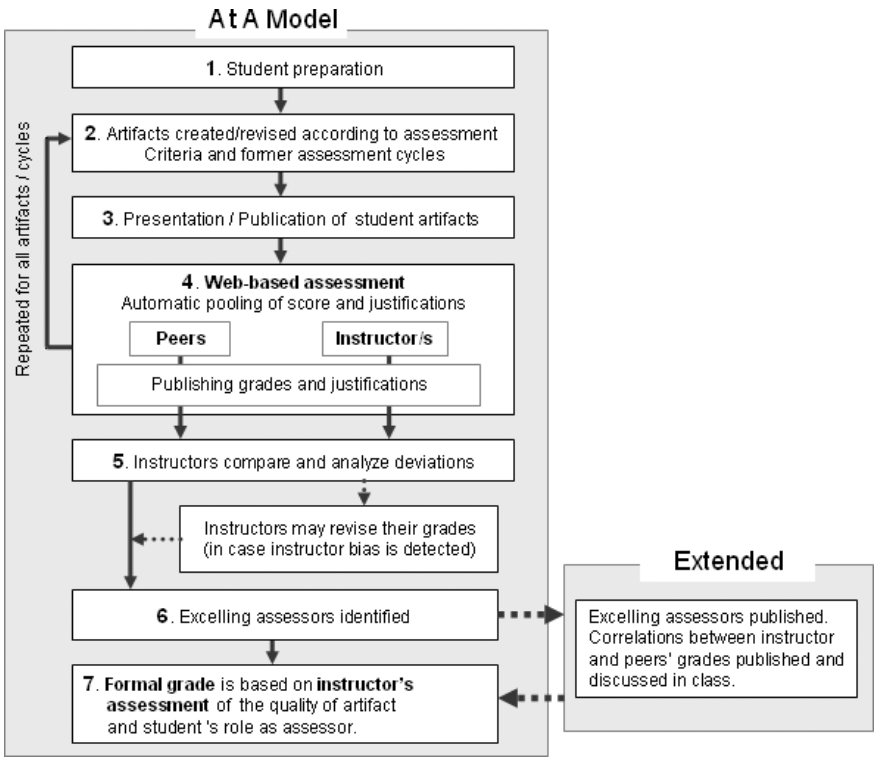


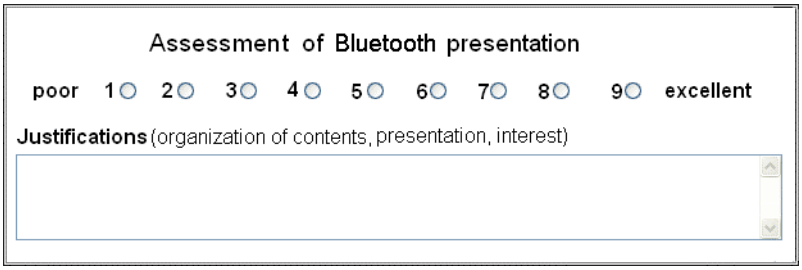
Figure 1. The AtA model.

assessments: (a) prior to the assessment of their artifact based on reviewing their peers' work, and (b) following the assessments of their artifacts based on assessment received from peers.

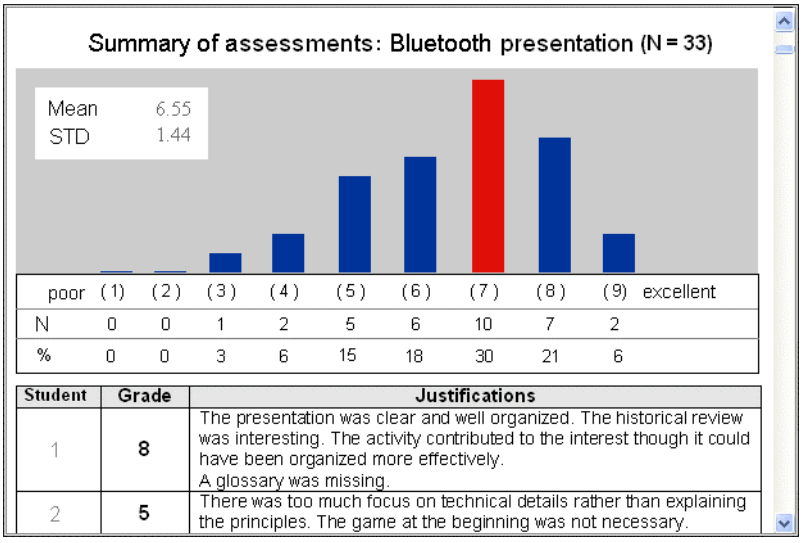
(3) Students either present their artifacts during course sessions or publish them online.

(4) Web-based assessment: students and instructors use a web-based tool to provide their assessments for peers. The assessment includes a score and justifications based on the earlier defined criteria (Figure 2a). Assessments are automatically pooled, synthesized and published online; at the end of each stage students can see the summary of the assessments: the descriptive statistical analysis of the grades and the list of justifications presented anonymously (Figure 2b).

Stages 2–4 are repeated in several cycles according to the number of artifacts and their form of presentation. Providing students an opportunity to assess their



(a)



(b)

Figure 2. (a) Pooling of assessments (b) Presentation of the summary of peer assessments.

peers and view the summary of the assessments in several cycles is an important part of the AtA model.

(5) After all assessments are collected, instructors identify deviations between peer and instructor scores. Students' scores and justification, as indicated from our outcomes below, serve as excellent indicators for instructors about students' understanding of the criteria and contents. Additionally, they can serve as reference for instructors to inspect their own assessment; using the justifications, instructors may choose to revise their own scores, in case they realize they might have been biased in any way.

(6) "Excelling assessors" are identified based on:

   (a) Degree of participation in the peer assessment activity (the extent to which students fulfilled the requirement to provide scores and justifications to peers).
   (b) Degree of correlation with instructor's score (applicable only in activities which require student to assess at least five peer artifacts).
   (c) Quality of justifications (the extent to which justifications were aligned with criteria and depth of ideas in claims).
   (d) Respecting classroom pre-defined norms (use of appropriate language, providing constructive critique, etc.)

   The extended model (applied when time constraints permit) includes publishing the list of excelling assessors and discussing with students the statistical analysis, and especially the meaning of graphs showing correlations between instructor and peer grades. These discussions, as shown below, have great impact in promoting student confidence in their own ability to provide assessment to others and in their instructors' capability to provide sound assessment to all students.

(7) The formal grade for the course includes, in addition to assessment of other course activities, a grade for the artifact (provided by instructor) and a grade for each student as assessor.

2.2. *Research questions*

To study the value of the AtA model, we explored the following questions:

(1) What were the relationships between student and instructor grading in activities that employed the AtA model?
(2) How can discrepancies between student and instructor scores be explained and used to improve instruction?
(3) What is the effect of multiple cycles of implementation of the AtA model?
(4) What is the relationship between the quality of the assessment provided by students to their peers and independent measures of learning?
(5) What is the added value of the extended AtA model?

3. Methods

The sections below describe how data was collected and analyzed to explore the impact of the AtA model on learning and instruction. The effect of the model as a whole was explored via large scale implementation, while the specific impact of the extended model was explored with a smaller portion of that sample.

3.1. *Sample and implementation*

The AtA model was implemented during the academic years 2004–2006 in 25 academic courses in four universities and three colleges in Israel (Table 1). The activities were designed and performed by ten instructors who were guided by the authors. A total of 848 students participated in the study from which 159 students participated in classes in which the extended model was implemented (C14, C19, C20, C24, C25).

The AtA activities in this sample can be characterized according to the 17 dimensions in the typology formulated by Topping (1998) as follows:

(1) Curriculum area: as indicated in Table 1, the courses in the sample represent various subject matter domains including education, health professions, science and information systems.
(2) Objectives of the activity: the AtA approach refrains from using scores compiled from peer assessment to replace any portion of the instructor's scores for the artifacts and is not intended to save staff time. The objective is purely pedagogical.
(3) Focus: in most of the courses several cycles of assessment were enacted over several weeks. This enabled students to revise their artifacts and thus, the assessment to be formative rather than summative.
(4) The product assessed: The artifacts in the peer assessment activities included in the sample were of two types: a product of a development project or artifacts in which students reviewed a certain product and presented a summary of their review. The products are described in Section 3.2.
(5) Relation to "official" staff assessment: in all AtA activities, as applied from the rationale described above, scores were neither subsitutional nor supplementary to staff assessment. Students knew that scores provided to them by peers would not affect their official grade at all. Official scores were provided only by staff.
(6) Official weight: zero percent of the official grade.
(7) Directionality of peer assessment: assessments in all courses was mutual, meaning that each student assessed all, or most of the peers' artifacts.
(8) Privacy: in all courses students knew whose product they were assessing and provided their assessment anonymously.
(9) Personal contact: all the courses were based on public presentations of products, while the assessment was provided in written format via a Web-based tool — the CeLS platform (Ronen, Kohen-Vacs, & Raz-Fogel, 2006) briefly

Table 1. Courses, assessed artifacts, and correlations between peers and instructor grades.

	Name of Course	Degree	Year	Instit.	No. of Students	Artifact			Assess. Per Artifact	Assess. Per Student (Cycles)	Correlation Between Peers and Instructor Grades $r(p)$
						Form of Presentation	Created by	No. of Artifacts			
C1	Computers in education	MA	2004	HU	44	Web	group	15	7	3(1)	0.81(0.0001)
C2	Online instruction	BSc/MSc	2005	TEC	26	Web	group	8	11	5(4)	0.90(0.001)
C3	Designing Ed Technologies	MSc/PhD	2004	TEC	11	class	group	5	7	4(4)	0.87(0.03)
C4	Educational philosophy	BSc	2004	TEC	47	class	group	12	40	11(5)	0.73(0.003)
C5	Educational philosophy	BSc	2004	TEC	27	class	group	9	24	8(4)	0.63(0.03)
C6	Educational philosophy	BSc	2005	TEC	35	class	group	9	27	8(4)	0.70(0.02)
C7	Educational philosophy	BSc	2005	TEC	35	class	group	9	28	8(4)	0.86(0.001)
C8	Electronics Teaching	BSc	2005	TEC	13	class	group	13	12	12(6)	0.83(0.0002)
C9	Information systems	BA	2004	ACC	60	class	group	15	32	13(3)	0.8(0.0002)
C10	Information systems	BA	2004	ACC	46	Web	group	12	27	11(3)	0.74(0.002)
C11	History teaching	MA	2004	TAU	24	Web	group	13	3	3(1)	n.s.
C12	Assistive technology	BA OT	2004	TAU	64	class	individual	64	48	30(7)	0.89(0.004)

Table 1. (*Continued*)

	Name of Course	Degree	Year	Instit.	No. of Students	Artifact			Assess. Per Artifact	Assess. Per Student (Cycles)	Correlation Between Peers and Instructor Grades $r(p)$
						Form of Presentation	Created by	No. of Artifacts			
C13	Assistive technology	BA OT	2005	TAU	58	class	individual	58	45	30(7)	0.70(0.0001)
C14	Assistive technology	BA OT	2006	TAU	59	class	individual	59	45	30(7)	0.86(0.0008)
C15	Instruction in Medical professions	MA OT	2004	TAU	11	class	individual	7	7	5(3)	0.95(0.0006)
C16	Instruction in Medical professions	MA OT	2005	TAU	10	class	individual	6	8	5(3)	0.93(0.003)
C17	Multimedia applications in OT	BA OT	2005	TAU	60	class	group	15	32	14(7)	0.65(0.004)
C18	Science	BED	2004	ORC	16	class	group	6	10	5(5)	n.s.
C19	Web based teaching	MA	2005	HAU	13	Web	individual	13	12	12(4)	0.51(0.04)
C20	Interactive systems	BA	2005	HIT	26	class	group	10	10	9(5)	0.92(0.0001)
C21	New media technologies	BA	2005	HIT	26	class	group	6	16	5(2)	0.75(0.04)
C22	Data bases	BA	2005	HIT	15	class	group	6	12	6(3)	n.s.
C23	Programming	BA	2005	HIT	15	class	group	6	12	6(3)	0.82(0.02)
C24	Interactive systems	BA	2006	HIT	50	class	group	15	40	12(6)	0.49(0.04)
C25	New media technologies	BA	2006	HIT	57	class	group	16	40	12(6)	0.75(0.0003)

Institutions: Hebrew University (HU), Technion (TEC), Ahva College (ACC), Tel-Aviv University (TAU), Haifa University (HAU), Oranim College (OR), Holon Institute of Technology (HIT).

described in Section 3.3. The online assessment form had a similar format in all courses including a space to provide global grade (on a scale of 1–7 or 1–9) and several justifications representing the major evaluation criteria (Figure 2a).

(10) Synchronicity: in all the courses the assessment was conducted on products created at the same run of the course, i.e. assessors and assessees were from the same class.
(11) Ability: students in all activities were at similar stages in their studies.
(12) Constellation of Assessors: individuals in all courses.
(13) Constellation of Assessees: in 19 courses assessees worked in groups, in 6 courses assessees worked as individuals.
(14) Place: following the presentations (which were done in class) or publication of artifacts online, all assessments were provided by students at home or any other computer which they could work on.
(15) Time: all assessments were provided by assessors at their free time, using notes they took during presentations or the online published artifacts.
(16) Requirement: peer assessment was a compulsory activity of all courses.
(17) Reward: this dimension points to the core of the AtA approach; rewarding students for their participation and the quality of their assessment.

3.2. *The nature of the tasks*

Following is a short description of the various products assessed in the courses (Table 1).

(a) Products of a development project:

- In C4–C7, which dealt with philosophy of education, students produced and presented a model they developed of an "ideal school". The ideal school was a project carried on throughout the semester in which students applied knowledge gained via various resources in the course, such as literature reading and analyzing philosophical perspectives of one "interesting school", to develop their own conceptual model of an "ideal school". The product was a fifteen-minute presentation of "a day in the life of a student" in the school, presented as a PowerPoint presentation, a website, or even a short play (Levin-Peled *et al.*, 2007).
- In C8, C15 and C16, which dealt with electronics teaching and instruction in medical professions, students designed a lesson plan on a specific topic and presented it in class.
- In C2, which dealt with design of educational technologies, students designed a demo of an educational technology. The demo was a product of a semester long step-by-step process in which students learned how to design educational technologies. They first picked a topic and potential audience for their technology, then they created a needs and content analysis document, they brainstormed ideas for activities, developed a flow of these activities, and designed software features

for each activity. The final stage was a twenty-minute presentation of the demo in class (Ronen-Fuhrmann, Kali, & Hoadley, 2008).

- In C3, which dealt with online instruction, students created and ran a two-week long online mini-course. The mini-courses were developed by groups of 3 to 4 students using a learning management system used by the institution. The students who assessed each mini-course were the learners (peer students) who had a very deep acquaintanceship with the product (Levin-Peled *et al.*, 2007).
- In C1,C19 and C11, which dealt with web-based teaching and history teaching, students developed WebQuest activities for school students. Peers examined the activities that were presented in class and were uploaded to the course website.
- In C18, which dealt with science for pre-service teachers, students designed scientific experiments for answering questions from everyday life, with answers involving a knowledge in science. Each team had to phrase a question (e.g.: how to determine the quality of diapers?), design and perform an appropriate experiment, and explain and critique the results. The experiments were performed in class in front of the peer students who served as audience.
- In C22, C23, which dealt with databases and programming, students developed a computer program as the final project in the course. In C22 each team designed and developed a database and an interface that allows search and display of database items. In C23 each team developed a computer game in VisualBasic. The projects were presented in class during a dedicated session.

(b) Review and presentation of a topic according to a given format:

- In C12–C14, students reviewed and presented an assistive technology for a specific purpose. Each team was assigned with a general topic of assistive technology, such as technologies that assist in mobility, home affairs, leisure, and work. Each team member had to identify, explore, document and evaluate one product related to the topic. The team members presented the products they have evaluated during one class session. The activity was repeated during 8 weeks for all topics.
- In C17, students reviewed a multimedia application for occupational therapy. The format of this activity was similar to activities in C12–C14.
- In C9, C10, C20, C21, C24 and C25 which dealt with new media technologies, students reviewed a type of technology (e.g. Bluetooth), its history, uses and its potential and limitations. The reviews were presented in class and were uploaded to the course website.

3.3. *The web-based environment*

CeLS (Collaborative e-learning Structures) is a web-based system designed to model, enact, share and reuse online activities and incorporate them in the existing instructional settings for any subject and level (elementary school to higher education). The unique feature in the CeLS's design is the ability to control the flow of data in order to reuse learners' inputs and products from previous stages, and to

relate them to different Social Settings (Figure 3). A script designed in CeLS may include any number of stages. Each stage comprises any combination of building blocks of four types:

- **Presentation objects** are passive elements used to display information and instructions provided by the teacher or to present learners' *products* from previous stages according to the Social Settings defined. A product can be an organized collection of items contributed by different participants (identified or anonymous) or a single item that results from the combined action of a subgroup.
- **Input objects** are interfaces that allow the participant to submit *new data* to the system as text, links, media and files or as grading and voting on various scales, questionnaires and rubrics.
- **Interaction objects** are interfaces that allow participants to *interact with products* submitted by other learners in previous stages by commenting, grading, ranking, categorizing or editing via text or graphic manipulations.
- **Communication objects** are interfaces that allow participants to freely communicate with each other according to the Social Settings, using a synchronous forum.

Each object has properties that can be adjusted by the teacher. Some properties are generic, for instance, if the completion of an object is mandatory or not, and others are particular to the object, for instance, maximum or minimum text length or the vocabulary used for Text Input object. The functionality of an object is determined by its Social Settings so that different participants may encounter different information and perform actions on different data during the same activity stage. The Social Settings are an *integral part of the script definition* and are kept intact when the activity is duplicated for reuse with different students. Subgroups do not have to be 'populated' in order to create or to edit a script. Assigning students to groups and roles can be done just before the actual enactment of the script automatically by the teacher or by students' self registration.

The flexibility of this modular approach, relating actions to social structures enables teachers to design and enact a large variety of complex, multi-stage online activities, including creating and analyzing a common database, reaching an agreement, peer-product evaluation, contest and jigsaw.

CeLS provides a sample of content-free scripts and a searchable domain of activities that were implemented with students by all system users. Teachers can explore these resources, adopt them for personal use and adapt their structure and content to suit their specific needs or create new scripts using the basic building blocks.

The activities included in this study consisted of the following stages: submission of products (in case of online electronic products), pooling assessment (Figure 2a) and presentation of the results (Figure 2b). Teachers have reused formats applied in other courses and adapted them to their specific contents and needs by changing the general instructions and the evaluation criteria.

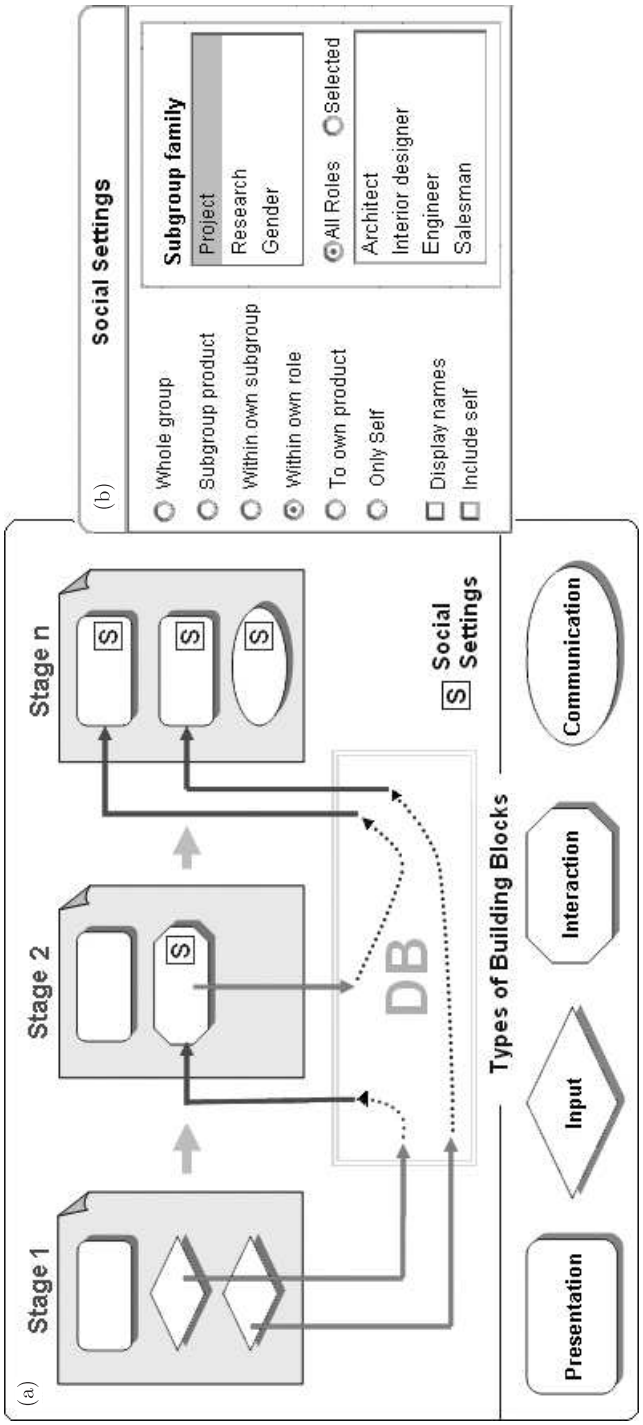


Figure 3. (a) Schematics of an activity in CeLS. (b) The Social Setting attributed to an object.

3.4. *Tools and data sources*

To provide answers for the research questions, we used the following data resources:

- Information gathered in the CeLS environment: the grades and justifications provided by peers and by instructors for the artifacts.
- The websites of each of the courses, from which we gathered information about the contents and the design of each course and the way the AtA activities were embedded in it.
- Phone interviews of about 30 minutes with each of the ten instructors, conducted at the week following the completion of the AtA activity. The interviews included general reflection questions about the enactment of the activity: students' reactions, the quality of artifacts and assessments, possible sources of bias and possible sources of discrepancies between peers' and instructor's assessments.
- A reflective survey, administered at the end of the last session in two classes in which the extended AtA model was implemented (C24, C25, $N = 107$). The survey included the following questions:
  (a) To what extent do you feel that you have learned from the peer assessment activity? (none; a little; some; much; a great amount)
  (b) Have you revised your artifact as a result of assessing peer's artifacts? Explain how (applicable only for students whose artifact was assessed in second or later cycles of assessment) (no; yes)
  (c) To what extent did the discussion in class about the comparison between peer and instructor assessments, and the publication of excelling assessors affect your (explain why for each):
    — Confidence as assessors (less confident, not affected, more confident)
    — Trust in peer grades (less trust, not affected, more trust)
    — Trust in instructors grades (less trust, not affected, more trust)

3.5. *Means of analysis*

To evaluate the impact of the AtA model and to better understand some of the classroom dynamics in its implementation, we analyzed the collected data via the following means:

3.5.1. *Comparison between students' and instructors' grades*

For each of the implementations, correlations between peer and instructor grades were obtained using Pearson correlation. Based on the above literature review, we assumed that the degree of correlation between grades provided by peers and those provided by instructors is a good measure of success for peer assessment activities. When a peer assessment activity is designed and conducted properly (students are well prepared for the activity, students are involved in defining the assessment criteria, anonymity is kept as much as possible, etc.), significant correlations are

expected. Such correlations are considered as a good indicator for student understanding of the assessment criteria, as well as their understanding of the specific contents involved in the activity (Falchikov and Goldfinch, 2000; Topping, 1998).

However, searching for such correlations was not the main focus of our analysis. Rather, the correlation results enabled us to focus on the identification of discrepancies in which assessments provided by peers and instructors were very different. Characterization of these discrepancies, using the information from interviews, enabled us to describe specific types of biases or problems in the implementation.

3.5.2. *The effect of multiple AtA cycles*

To examine the effect of enacting multiple cycles of peer assessment (Figure 1), we examined two courses which included 6 assessment cycles. In courses C24 (Interactive systems) and C25 (New media technologies), taught by two different instructors at the same institution, students were required to explore a technological topic in small groups and prepare a thirty-minute class presentation according to a given format. The class presentations were carried out at the second part of the courses in which two-three presentations were presented in a session during a period of six weeks (cycles). The summary of the peer grades and the peer justifications was published online each week before the next presentations cycle. To examine the effect of the multiple-cycle-implementation format, we divided the data collected in each of these courses into two sections: the first 3 weeks of the AtA activity and the last 3 weeks. Pearson correlations were obtained for each section separately to seek for differences between the two parts. We suspected that the evidence of improvement would be indicated by higher correlations in the second part of the AtA activity.

3.5.3. *Assessor index and independent measures of learning*

To identify excelling assessors, we defined a quantitative measure which was composed of two elements: (a) degree of participation in the peer assessment activity, and (b) correlation with the instructor's grades. The rationale for integrating these two factors (degree of participation and correlation with instructor) into the Assessor Index was that we wanted to encourage students to provide assessments to as many of the artifacts they were assigned to assess and to try to provide their best assessments for each artifact. Excelling assessors were defined as students who have submitted the required number of assessments (as required in each course) and a significant correlation was found between their grades and the instructors' grades for the same artifacts. Non-excelling assessors were defined as students who have not submitted the required number of assessments or whose correlation with the instructor grades was not significant. In this manner, the Assessor Index reflects student's commitment to the task, mastery of the subject domain, understanding the evaluation criteria and ability to apply these criteria. We refer to this Index as representing the assessment performances. We postulated, based on Davies's (2004)

findings, which showed that peer assessment can serve as a good measure of understanding the assessed contents that a positive correlation might be found between the Assessor Index and other independent measures used to assess students in the same course. To check this postulation, the Assessor Index was correlated to grades from independent measures such as scores provided by instructor for other course assignments, in eight of the courses (C2, C4, C5, C10, C12, C14, C20, C22). The correlations were conducted using Spearman's Rho.

3.5.4. *The impact of the extended AtA model*

The added value of the extended AtA model was analyzed using data from the reflective survey. Descriptive statistical analysis was applied to evaluate the extent to which students felt that the AtA activity contributed to their learning, encouraged them to revise their artifacts, and affected their confidence as assessors, their trust in peer grades and their trust in the instructor's grades.

Additionally, since the reflection survey took place after students were exposed to grades provided to them by the instructors for their artifacts and for their role as assessors (excelling or not), it was interesting to check whether students' opinions expressed in the reflection survey were related to these grades. Therefore, Spearman's Rho correlations were used to examine the relation between the survey outcomes and: (a) instructor's grades for the artifact, and (b) the performance of students as assessors (Assessor Index).

4. Results

The analysis of the data, with the means of analysis described above revealed the following results. The insights from each piece of evidence is discussed briefly in each of the sections below and synthesized in the summary section of this paper.

4.1. *Comparison between students' and instructors' grades*

Most of the AtA activities in the sample (88%) revealed high and significant correlations between peer and instructor grades (Table 1), supporting the findings from the meta-analysis by Falchikov & Goldfinch (2000). The high correlations indicate that student understanding of both the evaluation criteria and the contents represented in the artifacts, was good.

As explained above, our analysis focused on explaining cases of discrepancies between scores provided by students and instructors for specific artifacts. We also focused on cases in which no significant correlation was found between students' and instructor's scores. A detailed examination of the qualitative data enabled us to explain and characterize these cases.

4.1.1. *Discrepancies in specific artifacts*

The analysis indicated that cases showing discrepancies between students' and instructor's scores for specific artifacts represent some kind of bias. Four types of biases, which were identified are described and exemplified.

Content related bias: In some AtA activities, we found that when the contents of artifacts included socially or culturally sensitive aspects, students tended to under or over grade the artifact. For instance, course C6 represents one enactment of an educational philosophy course. A main theme in the course is an "ideal school" project, in which groups of 3–4 students construct a conceptual model of a school that meets their evolving educational perceptions. Toward the end of the semester, each group gives a short presentation of one day in their ideal school, which is evaluated via the AtA model. In order to understand the social dynamics in the class, it is important to note that the course was a compulsory course and that the student population of such courses at the university in which it was implemented is typically heterogeneous. The course included about one third of Jewish students who were born in Israel, one third of Jewish students who are relatively new immigrants from the former USSR and one third of Israeli Arab students (Moslem and Christian). This multicultural characteristic of the course is usually a source of enrichment for the course, and enables students to be exposed to a variety of educational perceptions in a productive atmosphere. However one incident, which occurred during the peer-evaluation of a certain group's presentation shows how contents of an artifact can create bias. The main rationale for the "ideal school" presented by that group was to bridge between religious and non-religious students in a certain cultural group. At the end of the presentation, a discussion was held between students as to whether such a school could be applied to bridging between other religious and non-religious groups. The presenters claimed that the problems that they dealt with in their school were unique. This answer triggered a discussion in which cultural tensions were aroused. The peer assessment activity of this presentation reflected the tensions: a few peer scores were biased, showing a large discrepancy with the instructor's score. The justifications for these scores were inappropriate and even offending (Kali & Ronen, 2005).

Personally related bias: In some AtA activities, in which the identity of the assessees was known (such as when assessees presented their artifact in class) bias related to their identity was found. This type of bias was related to a certain belief or common image held by most of the students about the assessees. An example of such bias was evident in course C11, in which the assessed artifacts were WebQuests developed for history teaching. One of the teams in this class consisted of two experienced history teachers who were highly respected by their peers. Peer grades for the product of this group were much higher than the grade provided by the instructors (artifact a2 in Figure 4). The interview with the instructor revealed that the artifact presented by these teachers did not meet the criteria of an excellent

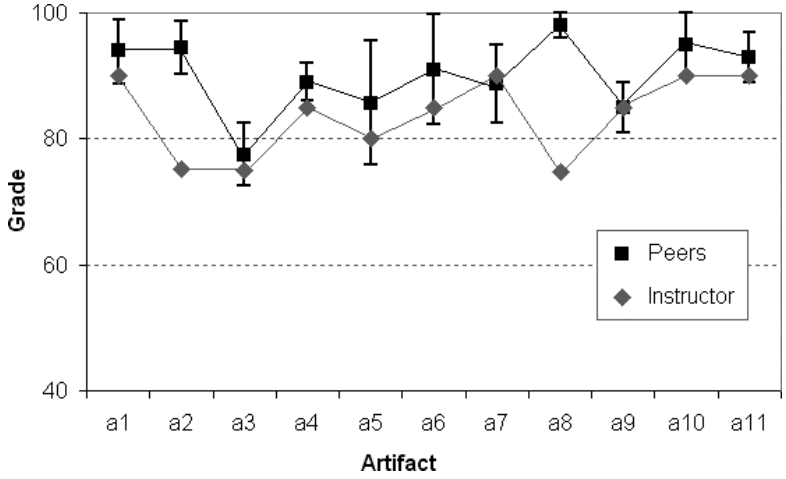


Figure 4. The summary of assessments for course C11 (artifacts are organized according to the order they were presented in class).

WebQuest. The instructor believed that students tended to give higher scores to this group due to their pre-conceptions about this group's teaching expertise.

Interaction between a personal characteristic and content: This third type of bias was found in cases in which there was a special relation between a personal characteristic of the assessees and the content of the artifact they presented. An example of such bias is manifested in course C11 in which students were required to present a WebQuest they had developed. One team, who consisted of two Arab students, a small minority in that class, submitted a product dealing with peace between Palestinians and Israelis. Even though the product itself did not meet the evaluation criteria as indicated by the interview with the instructor and by the grade she assigned them, it was highly graded by the peers, creating a discrepancy between student and instructor grades (artifact a8 in Figure 4). The analysis of the justifications provided by students show that their high grading did not match the criteria, which dealt with the quality of the artifact. Rather, in their justifications students expressed their compassion toward the presenters, and toward the ideas they presented. It is important to note that we are very much in favor of the expression of non-objective views in AtA activities. However, these should not interfere with the agreed-upon objective (as much as possible) evaluation criteria. Indeed, previous work (Kali, 2006; Kali & Ronen, 2005) has shown how to encourage students to express their emotions and non-objective views in a productive manner without causing biased assessment.

Instructors' bias: We have previously stated that we regard instructors' grade as a reference, representing the expert's assessment as recommended by many researchers in the peer evaluation field (e.g. Falchikov & Goldfinch, 2000). Nevertheless, the instructor is not free of bias which may result from the same reasons

described above for causing student bias. In this respect, the immediate availability of the data provided by the web-based environment can help the instructor reflect on his or her own assessment. When identifying a discrepancy between class grading and one's own grades for a specific artifact, the instructor should look for possible reasons for the difference. An honest instructor may sometimes realize that peer grades are more accurate than her own.

An example of such an incident occurred in course C16 which dealt with instruction in medical professions. In this course, MA students assessed lesson-plans presented by their peers. The detailed examination of the grades assigned to each artifact revealed a relatively large discrepancy between instructor and peer grades for a specific artifact (a3 in Figure 5) in which the instructor's grade was significantly higher than the grades provided by students. Artifact a3 was a lesson on "The importance of blood donation" designed for high school students. The interview with the instructor revealed that only after examining the grades and justifications provided by the students, she realized that she may have been too impressed by the many well-presented scientific details. She discovered that she did not pay enough attention to the major evaluation criterion set for this task: "Is the lesson adapted to the goals and the target population?" Justifications provided by peers, such as "the lesson missed the main point of motivating teenagers to donate blood" and the relatively low grades accompanied with these justifications, convinced the instructor to revise her assessment. She admitted that in this case, the peer assessment was more accurate than her own and corrected her grade accordingly before publishing the results (Figure 5).

4.1.2. *Cases showing no significant correlation*

As mentioned above, only in three courses of our sample (12%) there was no significant correlation between scores provided by peers and those provided by the

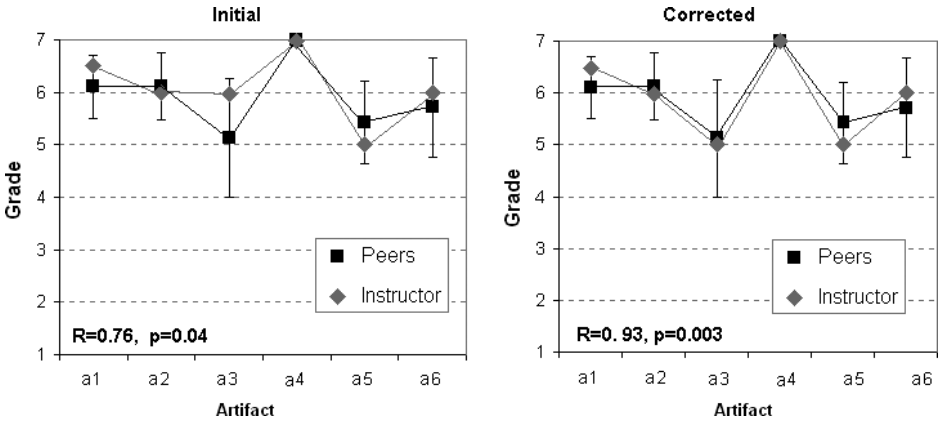


Figure 5. Instructor and peer grades for course C16 (initial on left, corrected on right).

instructor. Our findings indicate that such cases reflect a problem in instruction causing misunderstanding of the contents involved or of the evaluation criteria. This situation was evident in courses C18 and C22 and was validated by the interviews with the instructors. Course C18 was a course for pre-service teachers about science and scientific inquiry. The artifact presented and assessed in the AtA activity was a research question, and the design of a scientific experiment that addresses this question. Each group chose a topic of inquiry of their interest (e.g. one group designed a study to compare the absorption efficacy of different types of diapers). As can be seen in Table 1, no significant correlation was found between peers and instructor grades.

The analysis of justifications showed that students had difficulties applying the criteria. Instead of assessing the degree to which artifacts used appropriately a scientific method, students tended to justify their assessments on the interest they found in the topic explored in the artifacts. A similar case was evident in course C22 dealing with database development. The artifact in this case was a computer program developed by students. Despite the evaluation criteria which focused on the complexity and implementation of database aspects of the program, student assessment and justifications seemed to have been influenced by the graphical design of the interface and the topic it dealt with. Since in both cases such justifications were common, we refer to the inconsistencies between the student and instructor's scores as reflecting a problem in instruction rather than referring to such cases as content related bias for individual artifacts.

4.2. *The effect of multiple AtA cycles*

Like many of the other courses, courses C24 and C25, which included multiple AtA cycles revealed the typical pattern of significant high correlation between the students' and the instructor grades. However, a more informative picture was revealed, when the first and the second part of these longitudinal AtA activities were compiled separately. As can be seen in Table 2 and Figure 6, the correlation became significant only in the second part of the activity, meaning that throughout the AtA multi-cycle activity, students improved their ability to provide assessments to their peers.

4.3. *Assessor index and independent measures of learning*

As described above, we postulated that a possible significant correlation might be found between the Assessor Index and other independent measures of learning in

Table 2. Correlation between peers and instructor grades in courses C24 and C25.

	Course C24	Course C25
As a whole	$r = 0.49$, $p = 0.04$ [15 artifacts]*	$r = 0.75$, $p = 0.0003$ [16 artifacts]**
First part (3 weeks)	$r = 0.12$, $p = 0.4$ [7 artifacts]	$r = 0.53$, $p = 0.09$ [8 artifacts]
Last part (3 weeks)	$r = 0.67$, $p = 0.03$ [8 artifacts]*	$r = 0.91$, $p = 0.001$ [8 artifacts]**

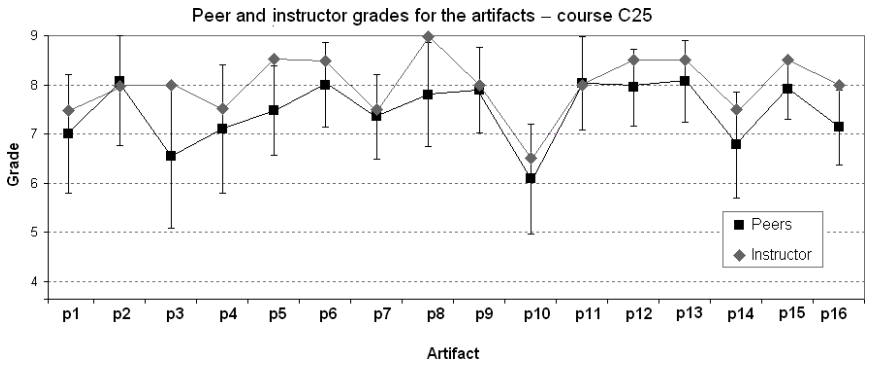


Figure 6. Peer and instructor grades for course C25.

the course. The rationale for this postulation was that the Assessor Index reflects students' understanding of the assessed contents as well as their investment in the course, while other grades in the course might represent the same thing. Our incentive to explore this postulation was that it would help us evaluate the meaning of the Assessor Index and its possible role as a means of assessing student understanding of the course contents. To check this assumption, we compared the Assessor Index with other independent measures of learning. For a sample of eight courses, instructors provided us with scores they gave students for other assignments in the course. The analysis revealed a significant correlation between the Assessor Index and those independent measures (Table 3), showing that excelling assessors were usually the students assigned with the higher grades in other tasks and vice versa. These findings confirm that the Assessor Index that we have used to represent students' assessment performances can also serve as a valid measure for assessing students' understanding of course contents supporting the findings reported by Davies (2004).

Table 3. Correlation between Assessor Index and independent grades in sample courses.

Course (Table 1)	The Independent Grade Correlated with Assessor Index	Correlation of Assessor Index with the Grade	
		r	*p*
C12	Instructor's grade for the personal artifact	0.24	0.02*
C14		0.60	0.02*
C10	Instructor's grade for a personal activity (quiz)	0.31	0.03*
		0.47	0.001**
C20	Instructor's grade for the group artifact	0.41	0.024*
C22		0.37	0.002**
C2	Instructor's grade for all other course activities	0.54	0.002**
C4		0.46	0.001**
C5		0.44	0.012*

4.4. *The impact of the extended AtA model*

The summary of the replies provided by students in course C24 and C25 to the first question of the reflection survey confirmed findings from the literature, which showed that students felt that peer assessment activities contributed to their learning (e.g. Falchikov, 2003; Smith *et al.*, 2002). Our findings indicated that 60% of the students felt that the activity contributed "much" or "to a great extent" to their learning. Typical explanations were: "I had to concentrate in others' presentations, and focus on all details to be able to provide the assessments. This was a good way to learn from others"; "Learning from others' mistakes was the biggest lesson for me"; "I felt that critiquing the other presentations helped me articulate what I need to focus on in our presentation".

Student replies to the second question of the survey indicated that conducting multiple cycles of peer assessment is a productive way to encourage students to revise and improve their artifacts. 65% of the students (excluding those whose artifacts were assessed in the first round of the activity) reported that they revised their own artifacts as a result of assessing their peers' artifacts.

Replies to the third question of the reflection survey, in which students were asked to reflect about how the AtA (extended model) affected their confidence as assessors, trust in instructor's grades, and trust in peer grades, are presented in Figure 7.

4.4.1. *Effect on student confidence as assessor*

Most students (56%) reported that following the activity, they were more confident of their ability to evaluate their peers' work. Typical explanations were: "I saw that my grades were statistically correlated with the teacher's grades"; "The justifications written by other students helped me understand how to evaluate the artifacts

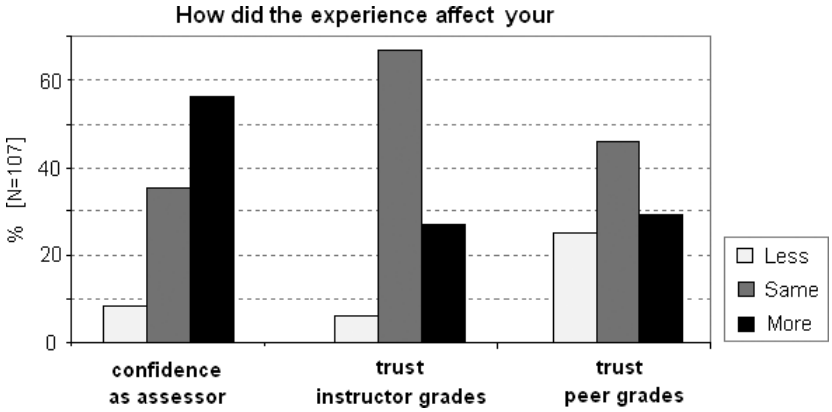


Figure 7. Students reflections to the application of the extended AtA model.

and I am also more aware now of the ability of my peers"; "I realize what the most important aspects of the evaluation are and learned how to avoid personal bias."

Only 8% of the students reported that following the AtA (extended) activity they are less confident of their ability to evaluate their peers' work. Typical explanations were: "Since I was not among those who were the excelling assessors, I think that maybe I do not know how to evaluate"; "In some of the cases, my grades did not match the teacher's grade".

4.4.2. *Effect on trust in teacher's grades*

For most students (67%), the trust in the teacher's grades was not affected by the AtA activity. An example of a typical explanation was: "I had trust before in the instructor's ability to grade me — he is the authority". 27% of the students stated that following the AtA activity they have more trust in the instructor's grades. Typical explanations were: "I saw that the instructor's grades and the justifications were quite similar to mine"; "I have mixed feelings: we were disappointed of the low grade for our artifact but I trust the evaluation and I'll try to learn from it". The 6% who reported a decrease in their trust in the instructor's grades were three students who argued with the instructor about a specific remark made by one of the peers about their presentation.

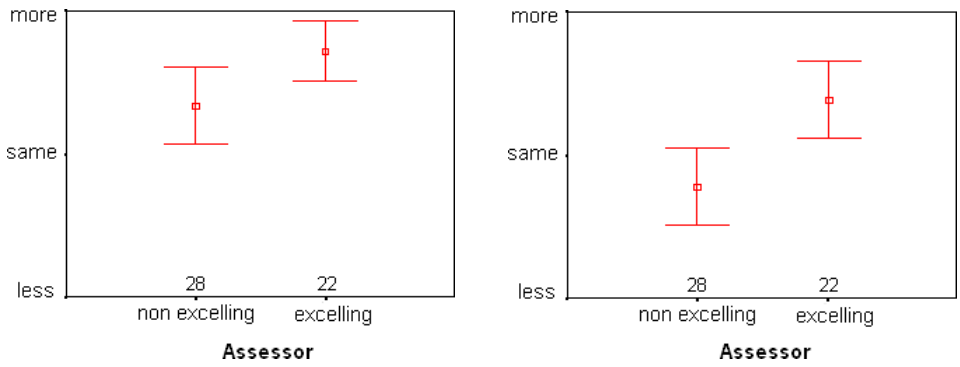
4.4.3. *Effect on trust in grades provided by peers*

The opinions on the effect of the activity on students' trust in grades provided to them by peers were divided as follows: 29% reported that following the AtA activity their trust in peer grades increased, 25% reported decrease, and 50% stated that their initial level of trust in peer grades has not changed. Typical explanations for increase were: "The immediate exposure of peer grades and evaluations enabled us to see the peers' opinions"; "At first I found it difficult to evaluate others, my skills improved in the next cycles and I believe that the same happened for many of us. The final results of correlation with the instructor's grades confirm it". Typical explanations for decrease in trust in peer grades were: "Some of the assessments were biased — students assigned high scores to their friends' artifacts"; "There was a discrepancy between the instructor and the peer grades for our product."

4.4.4. *Relation between confidence, trust and personal grades*

As described above, we wanted to check whether the change in students' attitudes was related to the grades provided to them by the instructor prior to the survey.

Relation to instructor's grades for the artifact: No significant correlation was found between the grade for the personal artifact and students' level of trust in the instructor's grades or in peer grades. This outcome indicates that the students accepted the grades as fair evaluations of their artifacts.



(a) Confidence in own ability as evaluator (b) Trust teacher's grades

Figure 8. Relation between students' change in confidence as assessors: (a) With regards to trusting teachers' grades (b) With regards to their identification as 'excelling assessors' (C24).

Relation to the performance of a student as assessor (Assessor Index): A different picture was revealed for the relation with the Assessor Index. Significant correlations were found between the Assessor Index and students' opinions regarding two aspects (Figure 8): Students who were defined as excelling assessors reported an increase of their confidence in their own ability to assess their peers' work more than those who were defined as non-excelling ($r = 0.3$, $p = 0.02$). Additionally, excelling assessors reported that the experience increased their trust in the instructor's grades ($r = 0.4$, $p = 0.002$).

These findings might reflect the result of the positive feedback students received when they were defined as "excelling assessors", which contributed to their self confidence. On the contrary, students who were not defined as "excelling" tended to express their mistrust in the instructor's grades. This last finding should be taken very seriously. It seems that the feedback students receive from instructors regarding their performance as assessors has a crucial role on their trust in the instructor's grading. To prevent a regression in students' trust, it is important that the AtA extended model would be conducted early enough within a course for students to improve their assessment performances and become excelling assessors. In this manner, the feedback they receive from the instructor regarding the assessment performances would serve as formative rather than summative assessment.

5. Summary and Discussion

In the beginning of this article, we highlighted the gap between research findings about the potential of socio-constructivist instructional approaches to support learning and the current state of instruction in higher education. The above findings indicate that the AtA multi-cycle peer assessment approach is a highly effective way to promote learning and instruction which takes advantage of this potential. We

summarize the affordances of this approach in three aspects: learners, instructors and social climate.

5.1. *Learners*

Many of the affordances reported in the literature for peer assessment were reproduced in this research: students were attentive to their peers' ideas and learned from the artifacts of their peers. By knowing the criteria in advance or in some cases being involved in creating the criteria, students were aware of what is expected from them, reflected on their own process in creating the artifact and produced artifacts that corresponded to the criteria. In addition to these benefits, our findings indicate an added value, unique to the AtA multi-cycle peer assessment approach: First, we found that most students improved their artifacts following their experience in providing assessment to others' artifacts as a result of better understanding of the assessment criteria and of adopting good ideas from their peers' artifacts into their own work. Second, in courses with multiple cycles of assessment, students improved their ability to provide assessment of high quality to their peers. Third, since students were presented with the correlation analysis and since in most cases there was a good correlation between peers and instructor scores, the confidence of most students in their ability to provide assessment increased. This was especially true for those who were identified as excelling assessors. Students' confidence in their instructors' scoring, which was high from the beginning, did not change much. However, the findings also indicate that to avoid non-excelling assessor from decreasing their trust in the instructor's grading, this approach should be conducted cautiously, enabling students to improve their assessment performances.

5.2. *Instructors*

Since the activity was performed in cycles, each with several assessed artifacts, which were analyzed immediately, instructors were able to easily identify discrepancies in the pattern of correlation between their own grading and the students' grading of their peers. The justifications, pooled together using the CeLS environment, enabled instructors to understand the sources of these discrepancies and distinguish between student bias (content related, personally related, or the interaction between them) for specific artifacts, and their own bias. When instructors found that their own assessment was biased, they were able to fix their assessment before it was published. In cases in which no significant correlation was found between peer and instructors' scores, instructors were able to pinpoint the source of the misunderstanding of the contents or assessment criteria. The immediate feedback for instructors enabled them to adapt their instruction and remedy many of these issues.

5.3. *Social climate*

Another important added value of the AtA approach for both learners and instructors is the fact that it reduces tensions, usually reported in peer assessment

activities, in which students are reluctant to grade and be graded by their peers. Such tensions were absent in the current study. Students realized that their grading has no effect on their peers' official scores and that they are being assessed as assessors. Therefore, their attitudes were positive and commitment was high.

5.4. *Concluding remark*

In the light of the findings of this research, we recommend instructors in higher education to utilize the AtA multi-cycle peer assessment approach and model to increase the benefits of peer assessment and enable their students to participate in a most productive activity. We highly recommend using the many tools that have already been developed and are continuing to evolve for web-based peer evaluation.

Finally, we would like to note that the CeLS environment, which was used in the current study continues to develop. The statistical analysis, such as the correlations between peer and instructor grades and the Assessor Index, which were compiled with other means in this study are planned to become automated features of the CeLS environment.

References

Boud, D., Cohen, R., & Sampson, J. (Eds.) (2001). *Peer learning in higher education: Learning from & with each other*. Routledge.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Research Council.

Brown, S., & Glasner, A. (Eds.) (1999). *Assessment matters in higher education: Choosing and using diverse approaches*. Philadelphia, UK: Open University Press.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23*(7), 13–20.

Cole, M. (1996). *Culture in mind*. Cambridge, MA: Harward University Press.

Davies, P. (2004). Don't write, just mark: The validity of assessing student ability via their computerized peer-marking of an essay rather than their creation of an essay. *Research in Learning Technology, 12*(3), 261–277.

Dominick, P. G., Reilly, R. R., & J., M. (1997). The effects of peer feedback on team member behavior. *Group and Organization Management, 22*, 508–520.

Falchikov, N. (2003). Involving student in assessment. *Psychology Learning and Teaching, 3*(2), 102–108.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287–322.

Freeman, M. (1995). Peer assessment by groups of group work. *Assessment & Evaluation in Higher Education, 20*(3), 289–300.

Gutierrez, K. D. & Rogoff, B. (2003). Cultural ways of knowing: Individual traits or repertoires of practice. *Educational Researcher, 32*(5), 19–25.

Herrington, J. A., Reeves, T. C., & Oliver, R. (2005). Online learning as information delivery: Digital myopia. *Journal of Interactive Learning Research, 16*(4), 353–367.

Kali, Y. (2006). Collaborative knowledge building using the Design Principles Database. *International Journal of Computer Support for Collaborative Learning, 1*(2), 187–201.

Kali, Y., & Linn, M. C. (2007). Technology-enhanced support strategies for inquiry learning. In J. M. Spector, M. D. Merrill, J. J. G. V. Merriënboer & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed.) (pp. 445–490). Mahwah, NJ: Erlbaum.

Kali, Y., & Ronen, M. (2005). Design principles for online peer-evaluation: Fostering objectivity. In T. Koschmann, D. Suthers & T. W. Chan (Eds.), in *Proceedings of CSCL 2005: Computer support for collaborative learning: The Next 10 Years!* (pp. 247–251). Mahwah, NJ: Lawrence Erlbaum Associates.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. In R. Pea & J. S. Brown (Eds.), *Learning in doing: Social, cognitive, and computational perspectives* (pp. 29–129). Cambridge, MA: Cambridge University Press.

Levin-Peled, R., Kali, Y., & Dori, Y. J. (2007). Promoting collaborative learning in higher education: Design principles for hybrid courses. In C. A. Chinn, G. Erkens & S. Puntambekar (Eds.), in *Proceedings of the computer supported collaborative learning (CSCL) 2007: Mice, minds and society* (pp. 418–427). International Society of the Learning Sciences, Inc.

Linn, M. C., Davis, E. A., & Bell, P. (Eds.) (2004). *Internet environments for science education.* Mahwah, NJ: Lawrence Erlbaum Associates.

Liu, E. Z., & Yuan, S. J. (2002). Alternatives to instructor assessment: A case study of comparing self and peer assessment with instructor assessment under a networked innovative assessment procedures. *International Journal of Instructional Media, 29*(4), 395–404.

Mann, B. (1999). Web course management in higher education. *Australian Educational Computing, 14*(1), 15–20.

McConnell, D. (2002). The experience of collaborative assessment in e-learning. *Studies in Continuing Education, 24*(1), 73–92.

McCray, R. A., DeHaan, R. L., & Schuck, J. A. (2003). *Improving undergraduate instruction in science, technology, engineering, and mathematics: Report of a workshop*: National Reseach Council, Washington, DC : National Academies Press.

Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment & Evaluation in Higher Education, 28*(4), 383–394.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Ronen, M., Kohen-Vacs, D., & Raz-Fogel, N. (2006). Adopt & adapt: Structuring, sharing and reusing asynchronous collaborative pedagogy. In S. Barab, K. Hay & D. Hickey (Eds.), *Making a difference: 7th international conference of the learning sciences.* Mahwah, NJ: Lawrence Erlbaum Associatesh.

Ronen, M., & Langley, D. (2004). Scaffolding complex tasks by open online submission: Emerging patterns and profiles. *Journal of Asynchronous Learning Networks, 8*(4), 39–61.

Ronen-Fuhrmann, T., Kali, Y., & Hoadley, C. M. (2008). Helping education students understand learning through designing. *Educational Technology*, *48*(2), 26–33.

Salomon, G., & Perkins, D. N. (1996). Learning in wonderland: What computers really offer to education. In S. Kerr (Ed.), *Technology and the future of education* (pp. 111–130). Chicago: University of Chicago Press.

Scardamalia, M., & Bereiter, C. (1996). Computer support for knowledge-building communities. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm* (pp. 249–268). Mahwah, NJ: Lawrence Erlbaum Associates.

Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: A case for student and staff development. *Innovations in Education and Teaching International, 39*(1), 71–81.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Stahl, G., Koshmann, T., & Suthers, D. (2006). Computer-supported collaborative learning. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–425). Cambridge: Cambridge University Press.

Stefani, L. A. J. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19*(1), 69–75.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

White, B. Y., & Frederiksen, J. R. (2000). Technological tools and instructional approaches for making scientific inquiry accessible to all. In M. J. Jacobson & R. B. Kozma (Eds.), *Innovations in science and mathematics education* (pp. 321–359). Mahwah, NJ: Lawrence Erlbaum Associates.

Zariski, A. (1996). Student peer assessment in tertiary education: Promise, perils and practice. In J. Abbott & L. Willcoxson (Eds.), in *Proceedings of the 5th annual teaching learning forum: Teaching and learning within and across disciplines* (pp. 189–200). Perth: Murdoch.