

Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously

M.J. van Zundert , K.D. Könings , D.M.A. Sluismans & J.J.G. van Merriënboer

To cite this article: M.J. van Zundert , K.D. Könings , D.M.A. Sluismans & J.J.G. van Merriënboer (2012) Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously, *Educational Studies*, 38:5, 541-557, DOI: [10.1080/03055698.2012.654920](https://doi.org/10.1080/03055698.2012.654920)

To link to this article: <https://doi.org/10.1080/03055698.2012.654920>



Published online: 31 Jan 2012.



Submit your article to this journal [↗](#)



Article views: 240



Citing articles: 5 [View citing articles](#) [↗](#)

Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously

M.J. van Zundert, K.D. Könings*, D.M.A. Sluijsmans and J.J.G. van Merriënboer

Department of Educational Development & Research and Graduate School of Health Professions Education, Maastricht University, Maastricht, Netherlands

(Received 27 December 2011; final version received 3 January 2012)

Instruction in peer assessment of complex task performance may cause high cognitive load, impairing learning. A stepwise instructional strategy aimed at reducing cognitive load was investigated by comparing it with a combined instructional strategy in an experiment with 128 secondary school students (mean age 14.0 years; 45.2% male) with the between-subjects factor instruction (stepwise, combined). In the stepwise condition, study tasks in Phase 1 were domain-specific and study tasks in Phase 2 had both domain-specific and peer assessment components. In the combined condition, these two components were present in all tasks in both phases. Final performance (i.e. speed and accuracy in domain-specific skills and peer assessment skills) showed no significant differences, but performance improved more from Phase 1 to Phase 2 in the stepwise condition than in the combined condition. The results suggest that, with complex study tasks, it might be beneficial to teach domain-specific skills before peer assessment skills.

Keywords: instruction; peer assessment; domain-specific; task complexity; cognitive load

In modern society, people are expected to perform increasingly complex professional tasks in complex workplaces (Van Merriënboer and Kirschner 2007). Since education has to prepare students for tomorrow's work environments, there have been shifts in education to accommodate changing societal demands. Where traditional education emphasises domain-specific knowledge and skills, modern education also addresses so-called twenty-first-century skills (Scardamalia 2001): more flexible, higher order skills, including skills for problem solving, critical thinking and lifelong learning.

Peer assessment in education, that is, learners evaluating the work of other equal-status learners (Topping 2009), is in line with developments in society, where peer assessment is increasingly regarded as an important professional skill for professionals and employees. In fact, the Dutch government announced in 2011 that all Dutch school teachers would be required to assess the teaching skills of their colleagues (The Netherlands Ministry of Education, Culture, and Science 2011). Similarly, employees in the US Army engage in after action review: a discussion with superiors and peers to assess aspects of their performance in combat training exercises (Fletcher 2009). And in academia, peer review of research manuscripts has

*Corresponding author. Email: kd.konings@maastrichtuniversity.nl

long been a common practice. Because of its match with educational goals, peer assessment is often applied in schools, with students evaluating the quality of the school work of their fellow students.

Despite the connection between peer assessment and developments in education, little is known about instructional methods for peer assessment skills (Van Zundert, Sluijsmans, and Van Merriënboer 2010). These methods are important, because students cannot be expected to properly perform a complex activity like peer assessment unless they have received guidance or training in that area (Sluijsmans 2002; Sluijsmans and Prins 2006). Indeed, Tsivitanidou, Zacharia, and Hovardas (2010) found that, without support, students' peer assessments were invalid. This underlines the importance of sound instructional strategies for peer assessment, particularly in educational contexts where increasingly complex tasks such as projects, problems, and professional tasks drive the learning process. Van Zundert, Sluijsmans, Könings, and Van Merriënboer (2012) showed that high complexity of domain-specific tasks had a stronger detrimental effect on the performance of peer assessment than on domain-specific task performance. This finding highlights the importance of taking into consideration the complexity of tasks when developing instructional strategies for peer assessment.

The current study reports on an experiment with a stepwise instructional method aimed at counteracting the adverse effect of task complexity on peer assessment instruction. The proposed method consists of a first phase in which only domain-specific skills are taught, followed by a second phase in which domain-specific skills are taught in combination with peer assessment skills. It was hypothesised that this stepwise method would yield better task performances (i.e. higher accuracy and speed) than a method in which domain-specific skills and peer assessment skills are combined in both phases. In the next two sections, two complementary areas, together providing the rationale for this stepwise method, will be discussed: the hierarchical structure of cognitive skills and the capacity of human working memory.

Hierarchical structure of cognitive skills

There is a conditional relationship between domain-specific and peer assessment skills (Van Zundert et al. 2012). For example, it would be very difficult, if not impossible, for students to assess a classmate's history essay without them having any knowledge or skills in the history domain. Thus, peer assessment skills are of a higher order than domain-specific skills, in other words, they are higher in the learning hierarchy. The classic description of learning hierarchies by Gagné, published in 1968, has had a profound impact on instructional theory (Smith and Ragan 2000). According to Gagné's description, different types of cognitive skills are positioned at different levels in the learning hierarchy depending on the extent to which their acquisition is conditional on the prior mastery of other skills. The ability to recognise stimuli, for example, is conditional on the ability to generate a response, and problem solving presupposes the ability to apply rules (Gagné 1985). In a similar vein, Anderson and Krathwohl (2001) revised Bloom's taxonomy of educational objectives (Bloom and Krathwohl 1956), incorporating the recent notion that knowledge is one of the dimensions of cognitive processes, ranging from memorising and understanding to more complex ones such as applying, analysing, evaluating and – at the highest level – creating knowledge. Evaluating or assessing is high in the hierarchy, since it is contingent on the presence of many other cognitive skills. These

ideas support the notion that peer assessment skill (i.e. evaluating) is higher in the hierarchy than the domain-specific skills that are the subject of peer assessment.

Working memory capacity: cognitive load theory

The instructional method for peer assessment, tested in this study, relies not only on the structural explanation of learning hierarchies but also on cognitive load theory. The latter is based on the notion of the limited cognitive capacity of working memory vs. the virtually unlimited capacity of long-term memory. In other words, storage capacity in memory is huge, but the resources available for the processing and storing of new information are limited (Sweller 1988, 2010; Sweller, Van Merriënboer, and Paas 1998). This explains why learning complex tasks may require multiple practice sessions. Complex tasks are often ecologically valid, have more than one possible solution and consist of many interacting elements that have to be processed simultaneously (Paas, Renkl, and Sweller 2003; Paas and Van Gog 2006; Van Merriënboer, Kester, and Paas 2006). Given the limited processing capacity of working memory, high element interactivity makes complex tasks difficult to understand and learn, in other words, it creates high cognitive load. When task complexity is due to high element interactivity and therefore intrinsic to the nature of the task, it is referred to as intrinsic cognitive load. Cognitive load due to (suboptimal) instructional procedures is known as extraneous cognitive load, which does not contribute to learning. Germane cognitive load refers to the working memory capacity actually allocated to intrinsic task aspects, and hence is the type of cognitive load that is most relevant to learning (Ibid.; Van Merriënboer and Sweller 2010).

Another important factor in cognitive load theory is the learner's level of expertise. The actual amount of cognitive load induced by the number and interactivity of elements depends on the prior knowledge of individual learners (e.g. Van Merriënboer, Kester, and Paas 2006; Van Merriënboer and Sluijsmans 2009). It is notably high for novice learners who have to process all elements of a complex task as separate units because they have not yet formed mental models, consisting of chunks of related elements (i.e. cognitive *schemata*). With growing expertise, elements are increasingly combined to form chunks, which function as single elements in working memory. Consequently, with increasing expertise, the number of elements in complex tasks decreases and cognitive load is reduced.

A problem of peer assessment instruction

Learning hierarchies and cognitive load theory provides the following notions, underpinning the peer assessment method that was tested in our study: (1) peer assessment skills are superposed on domain-specific skills and (2) complex tasks use up many cognitive resources. When a complex domain-specific task is combined with a peer assessment task, cognitive load increases and cognitive overload may occur. In other words, the available cognitive resources do not suffice to perform the combination of a complex domain-specific task plus a peer assessment task. Since peer assessment skills are superposed on domain-specific skills, the domain-specific task will be addressed first, using up most of the available cognitive resources, leaving few cognitive resources for the peer assessment task. Consequently, when task complexity increases, the peer assessment skill will be the first to suffer (Van Zundert et al. 2012).

Figure 1 illustrates this mechanism. Imagine a novice student, performing a simple task (see left part of Figure 1). Because the student is a novice, he or she has not constructed cognitive schemata yet. Thus, the task elements are perceived as separate units, which makes cognitive overload likely. The student firstly calls on his or her domain-specific skills that are prerequisite for performing the superposed peer assessment skills. The task being simple, the student's cognitive load will be relatively low, leaving sufficient working memory capacity for the peer assessment task. However, the situation is different with complex tasks (see the right part of Figure 1). The novice student again firstly uses cognitive resources for the domain-specific skills, but, the task being a complex one, the cognitive load will be high and most of the student's working memory capacity will be used up, leaving little or no surplus capacity for the peer assessment task. So, if task complexity is high, cognitive overload might occur and peer assessment suffers.

This appears to create a paradox. Peer assessment skills cannot be taught *detached* from the conditional domain-specific skills, but – if tasks are complex – cognitive overload makes it impossible for novice learners to learn the two skills *simultaneously*.

A potential solution to the problem of peer assessment instruction

In order to resolve the paradox, we propose a stepwise, two-phase instructional method. In Phase 1, only domain-specific task components are presented to enable students to form cognitive schemata of the domain-specific information. Then, in Phase 2, peer assessment task components are added to the domain-specific ones. It is expected that the domain-specific schemata acquired in Phase 1 will reduce the demand on cognitive resources from the domain-specific component, leaving more resources available for the peer assessment task, thereby preventing cognitive overload with its detrimental effects on peer assessment.

This stepwise method is derived from Pollock, Chandler, and Sweller (2002), who used a similar method with complex educational material in electrical engineering. They started by presenting undergraduate students with material containing

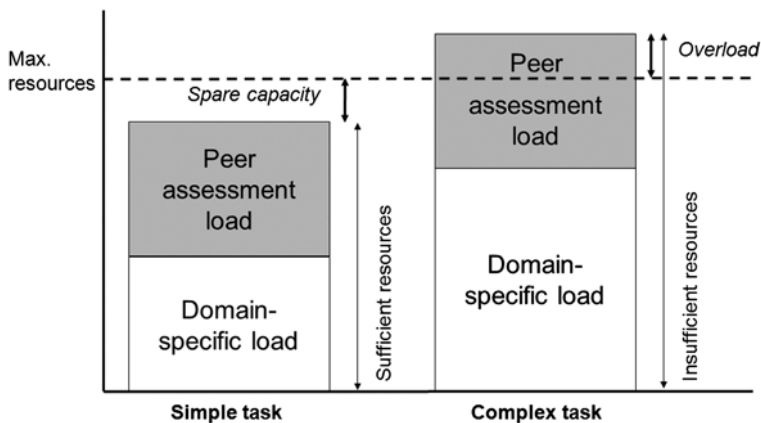


Figure 1. Effects of task complexity on cognitive load for the combination of domain-specific skills and superposed peer assessment skills.

only a limited amount of information elements, adding the remaining elements required for full understanding at a later stage. Their results indicated that the stepwise method yielded better task performances of novices than when all information elements were presented during all instructional stages. However, the study by Pollock et al. involved only first-order skills. The present study builds on this approach by investigating whether this instructional method may also be beneficial in situations that require integration of domain-specific skills with higher order skills, such as peer assessment (cf. Van Merriënboer and Sluijsmans 2009).

The aim of the present study was to investigate the effects of a stepwise method for novices learning complex tasks, comprising domain-specific and peer assessment skills. In Phase 1, students were presented with only domain-specific instruction and in Phase 2 with a combination of domain-specific and peer assessment instruction. The effects of this method were compared with those of a method presenting identical instructional tasks but combining domain-specific and peer assessment instruction in both phases. It was hypothesised that the stepwise method would yield superior results in terms of performance improvement from Phase 1 to Phase 2 and higher final performance on domain-specific tasks and peer assessment tasks.

Method

Participants

The participants were 128 secondary school students who were novices in the domain of the test tasks (i.e. steps in doing research). The mean age of the students was 14.00 years (SD=2.17) and 45.2% were male. The students received passwords for the electronic learning environment and were randomly allocated to either the stepwise ($N=61$) or the combined instruction condition ($N=67$). All participating students received a gift voucher of 10€ for their participation.

Materials

Within the electronic learning environment, specifically designed for this study, all participating students proceeded through two instructional phases, each consisting of three study tasks followed by two test tasks. The electronic learning environment consisted of domain-specific study tasks, combined (domain-specific and peer assessment) study tasks and test tasks, which will be described in more detail below. Figure 2 shows the sequence of the tasks in each condition.

Domain-specific study tasks

For Phase 1 of the *stepwise instruction condition*, three complex study tasks were developed, exclusively aimed at the acquisition of domain-specific skills. The objective of the tasks was to teach students how to identify the six main steps of research: (1) observation, (2) problem statement, (3) hypothesis, (4) experimental stage, (5) results and (6) conclusions (Table 1).

Each of the three study tasks described a scientific experiment in areas including operant conditioning, appetite and the effects of taking vitamin C. The descriptions contained, in a random sequence, events representing the six main steps of scientific research, but it was not indicated which event matched which step.

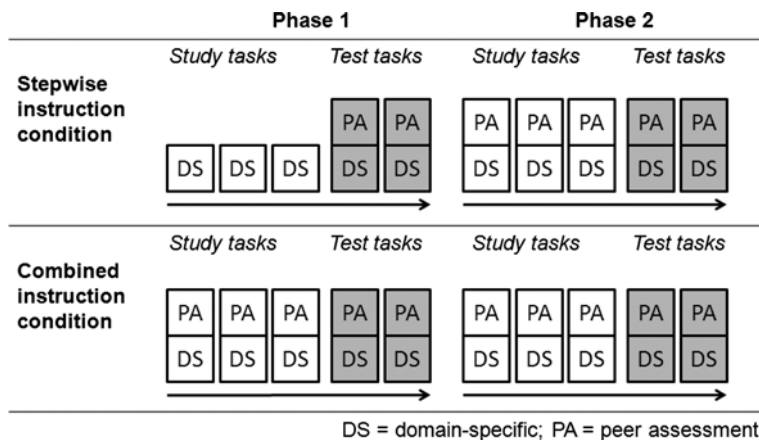


Figure 2. Sequence of the tasks for each condition.

The students were taught to identify the six research steps by means of worked examples (Sweller, Van Merriënboer, and Paas 1998), presenting the identification, with argumentation, of each research step. The argumentation consisted of an explanation of each research step with a link to the corresponding step in the description of the experiment. The instructional format of the worked examples is presented in Appendix A.

Combined study tasks

For Phase 1 of the *combined instruction condition*, three study tasks were developed, of which the domain-specific component was identical to that of the tasks in the *stepwise instruction condition*. For Phase 2 of both instruction conditions, three combined study tasks were developed, aimed at the simultaneous acquisition of domain-specific and peer assessment skills. Thus, the students in both conditions performed the same six domain-specific tasks and received the same instructions for these tasks. In addition to the domain-specific component, the combined study tasks contained peer descriptions of scientific experiments with fictitious peer solutions, that is, a peer’s identification of a research step with supporting argumentation. The fictitious peer solutions were designed by the researchers and were identical for all

Table 1. Explanations of the research steps.

Research step	Explanation
Observation	A certain phenomenon that is considered for further research is perceived
Problem statement	The researcher experiences the observation as a problem and formulates a problem statement
Hypothesis	An attempt is made to provide a logical explanation for the problem
Experimental stage	It is tested whether the hypothesis is correct or incorrect. In order to do this, the researcher conducts an experiment and collects data
Results	The collected data are displayed synoptically, for example, in graphs, diagrams or tables
Conclusion	The results are compared with the hypothesis

students. The students were taught how to assess the domain-specific task solutions of the peers by means of worked examples, showing a fictitious teacher’s judgement of the accuracy of the peer’s solution along with an argumentation for this judgement. The argumentation consisted of an explanation of the judgement for each step, and a link of this judgement to the peer’s task solution. The peer assessment instructional format of the worked examples is presented in Appendix B.

Figure 3 depicts a fragment of an integrated study task as presented in both phases of the combined instruction condition and in the second phase of the stepwise instruction condition. In the upper text box, a full description of an experiment (i.e. all six research steps) is presented. Below this box, one research step is shown. The left-hand box, labelled “teacher”, contains domain-specific instruction about the research step. The right-hand box, labelled “classmate”, presents the task solution of a fictitious peer, that is, the research step as identified by the peer. The lower box, labelled “assess whether your classmate recognised the research step correctly”, provides peer assessment instruction, explaining how to assess the peer’s solution.

In this box, a full experiment description is displayed (i.e., the six research steps in random order)

Here, one of the six research steps is presented

This box shows the solution of a fictitious peer

This is the domain-specific instruction

Here, the peer assessment instruction is given

Figure 3. Example fragment of an integrated study task.

Combined test tasks

Four test tasks were developed to measure students' domain-specific and peer assessment skills. All test tasks consisted of a domain-specific and a peer assessment component.

Domain-specific test tasks. The domain-specific test tasks were developed to measure students' domain-specific skill, that is, their *own* knowledge and understanding of the six research steps. Like the study tasks, the test tasks consisted of a description of a scientific experiment in which the six steps of scientific research were presented, unidentified, in a random sequence. The topics of the test tasks differed from those of the study tasks. Students were given open questions, asking them to correctly identify the six research steps.

Peer assessment test tasks. The peer assessment test tasks were developed to measure students' ability to assess a *peer's* knowledge and understanding of the research steps. Students were given solutions of a fictitious peer performing the domain-specific tasks. For each peer solution, the students answered open questions, asking them to independently assess their classmate's domain-specific accuracy.

Film

To provide a break between Phase 1 and Phase 2, the students watched a short film showing a monkey in a funny research setting. The break was inserted to prevent students from getting tired during Phase 2.

Questionnaire

To control for differences in students' motivation towards the tasks, Keller's (1983) Instructional material motivational survey (IMMS) questionnaire was used. This questionnaire consists of 36 items, measuring four constructs: (a) attention items (e.g. There was something interesting at the beginning of each task that got my attention), (b) confidence items (e.g. As I worked on these tasks, I was confident that I could learn the content), (c) relevance items (e.g. I could relate the content of these tasks to things I have seen, done or thought about in my own life) and (d) satisfaction items (e.g. Completing the tasks gave me a satisfying feeling of accomplishment). Cronbach's alpha for attention, confidence, relevance and satisfaction was .86, .91, .71 and .89, respectively. The five-point Likert-type items ranged from 1 (totally disagree) to 5 (totally agree).

Procedure

We used a mixed factorial design with the between-subjects factor "instruction" (i.e. stepwise, combined). All students logged on to a computer and entered the electronic learning environment. Immediately thereafter, they were randomly assigned to one of the two instructional conditions: (1) stepwise or (2) combined. After answering demographic questions about age, gender and school class, the students undertook Phase 1. The students in the *stepwise instruction condition* studied three domain-specific study tasks: tasks with only domain-specific instructions. The students in the *combined instruction condition* studied three combined study tasks:

domain-specific instructions combined with peer assessment instructions. After that, all the students were given two combined test tasks (a domain-specific test and a peer assessment test). The test was followed by the break with a short film and a snack. Phase 2 commenced after the break. The procedure was identical, only now all students received combined study tasks, followed by two combined test tasks. Finally, all students filled in the motivation questionnaire IMMS. The entire procedure took approximately 90 min.

Scoring

Domain-specific accuracy measure

Each domain-specific test task consisted of six research steps. For each correctly identified research step, a score of 1 was awarded. If the research step was not identified correctly, a score of 0 was assigned. Hence, domain-specific task scores varied between 0 and 6. As there were two test tasks in each phase, the results per phase will be reported as the mean scores of the two test tasks.

Peer assessment accuracy measure

Each peer assessment test task consisted of six research steps identified by a fictitious peer. The peer's answers could be: completely correct (i.e. correct step identification and correct argumentation), completely incorrect (i.e. incorrect step identification and incorrect argumentation) or partially correct (i.e. either correct step identification and incorrect argumentation or vice versa). Students assessed which of the three judgements applied to the peer's answers. If the student's judgement corresponded with the norm score (e.g. if a partially correct peer answer was assessed as partially correct by the student), a score of 1 was awarded. If the student's judgement deviated from the norm score (e.g. if a partially correct peer answer was assessed as completely incorrect by the student), a score of 0 was assigned. As students had to assess six answers in each test task, accuracy scores per task varied between 0 and 6. Students received two test tasks in each phase; the results per phase are reported as the mean score of the two test tasks.

Data analyses

The aim of the data analysis was to determine whether stepwise instruction, compared to combined instruction, led to more improvement of performance and to better final test performance. *Performance improvement* was defined as the difference between test performances in the two phases (i.e. performance in Phase 2 – performance in Phase 1). *Final performance* was operationalised as test performance in Phase 2. Performance was defined in terms of accuracy and speed. So, accuracy and time on test tasks were combined to estimate performance. Firstly, multivariate analyses of variance (MANCOVAs) on performance improvement were conducted for domain-specific and peer assessment test accuracy and time on test tasks, with time on study tasks as a covariate. Subsequently, univariate analyses of variance (ANCOVAs) on accuracy and time on test tasks were conducted with time on study tasks as a covariate. These analyses were then repeated for final test performance.

Results

The results for time on study tasks will be presented first, followed by the multivariate results for domain-specific and peer assessment test performances and the univariate results for time on test tasks, domain-specific accuracy and peer assessment accuracy. Finally, the results from the motivation questionnaire IMMS are reported. Partial eta-squared is provided as an estimate of effect size, with .01 corresponding to a small effect, .06 to a medium effect and .14 to a large effect (Tabachnick and Fidell 2007).

Study tasks

The upper part of Table 2 presents the means and standard deviations for *time on study tasks*. There was a significant effect of instruction on time spent on study tasks, $F(1, 126)=36.36$, $MSE=11,854.95$, $\eta_p^2=.22$, $p=.00$. In the combined instruction condition, the average time per study task decreased more rapidly from Phase 1 to Phase 2 than in the stepwise condition. In the combined instruction condition, the mean difference between Phases 1 and 2 was -128.85 ($SD=115.41$) seconds and in the stepwise instruction condition, it was -13.96 ($SD=101.11$) seconds. Time on study tasks in Phase 2 also differed significantly between the two instructional conditions, $F(1, 126)=9.00$, $MSE=4943.66$, $\eta_p^2=.07$, $p=.00$, with students in the stepwise instruction condition needing more time ($M=287.48$; $SD=82.99$) than the students in the combined instruction condition ($M=250.57$; $SD=56.57$). Time on study tasks was included as covariate in the analyses of the test tasks.

Domain-specific and peer assessment test performance

MANCOVA revealed a significant effect of instruction on *domain-specific performance*, $F(2, 124)=4.61$, $\eta_p^2=.07$, $p=.01$, indicating that performance improved more from Phase 1 to Phase 2 in the stepwise than in the combined instruction condition. The effect for final domain-specific performance was in the expected direction, but the difference between the two instructional conditions was not significant, $F(2, 124)=1.15$, $\eta_p^2=.02$, $p=.32$. As the multivariate analyses for final domain-specific performance revealed no significant effects, univariate results for final domain-specific performance are not reported.

MANCOVA also revealed a significant effect of instruction on *peer assessment performance*, $F(2, 124)=4.17$, $\eta_p^2=.06$, $p=.02$, indicating that in the stepwise instruction condition, peer assessment performance improved more from Phase 1 to Phase 2 than in the combined condition. The effect for the final peer assessment performance was in the expected direction, but there was again no significant difference between the two instructional conditions, $F(2, 124)=.59$, $\eta_p^2=.01$, $p=.56$, and therefore univariate results for final peer assessment performance are not reported.

ANCOVA on *time on test tasks* showed a significant effect of instruction, $F(1, 125)=8.00$, $MSE=54978.19$, $\eta_p^2=.06$, $p=.01$. There was a larger decrease in the mean time per test task between Phase 1 and Phase 2 in the stepwise instruction condition (-404.16 s, $SD=250.04$) than in the combined instruction condition (-360.79 s, $SD=247.90$). This is depicted in Figure 4.

Table 2. Means and standard deviations of time on study tasks, time on test tasks and test accuracy.

	Stepwise instruction (N=62)						Combined instruction (N=69)					
	Phase 1		Phase 2		Difference (Phase 2-Phase 1)		Phase 1		Phase 2		Difference (Phase 2-Phase 1)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
<i>Study tasks</i>												
Time on task ^a	301.44	73.98	287.48	82.99	-13.96	101.11	379.43	106.97	250.57	56.57	-128.85	115.41
<i>Test tasks</i>												
Domain-specific accuracy ^b	3.48	1.00	3.75	.89	.27	.90	3.56	.97	3.52	1.00	-.04	.99
Peer assessment accuracy ^b	3.75	.83	3.95	.80	.20	.93	3.85	.79	3.84	.84	-.01	.90
Time on task ^a	969.95	338.79	554.31	197.78	-404.16	250.03	922.12	296.73	548.33	189.20	-360.79	247.90

^aTime in seconds; ^bscale=0-6.

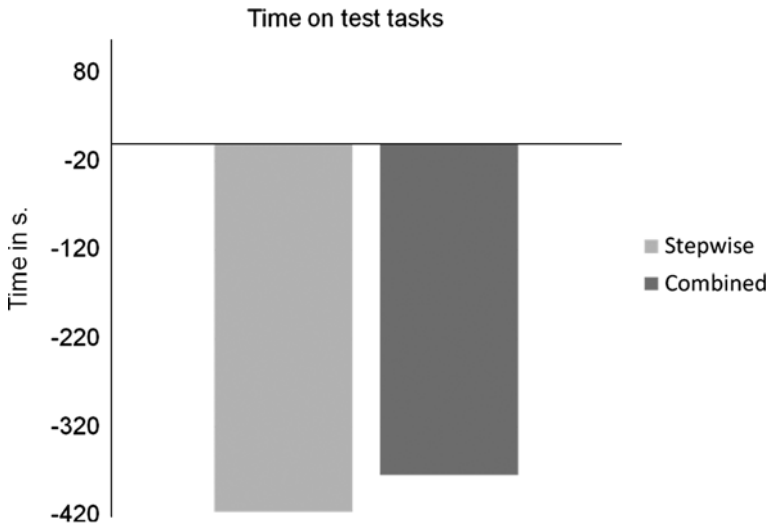


Figure 4. Mean differences in time on test tasks between Phase 1 and Phase 2 for each instructional condition.

Univariate analyses on accuracy improvement showed no significant effects of instruction. ANCOVA showed no significant effect of instruction on either *domain-specific accuracy*, $F(1, 125)=1.12$, $MSE=.90$, $\eta_p^2=.01$, $\theta=.29$, or *peer assessment accuracy*, $F(1, 125)=.33$, $MSE=.83$, $\eta_p^2=.00$, $p=.57$. The mean differences in domain-specific accuracy and peer assessment accuracy between the two phases of instruction are shown in Figure 5.

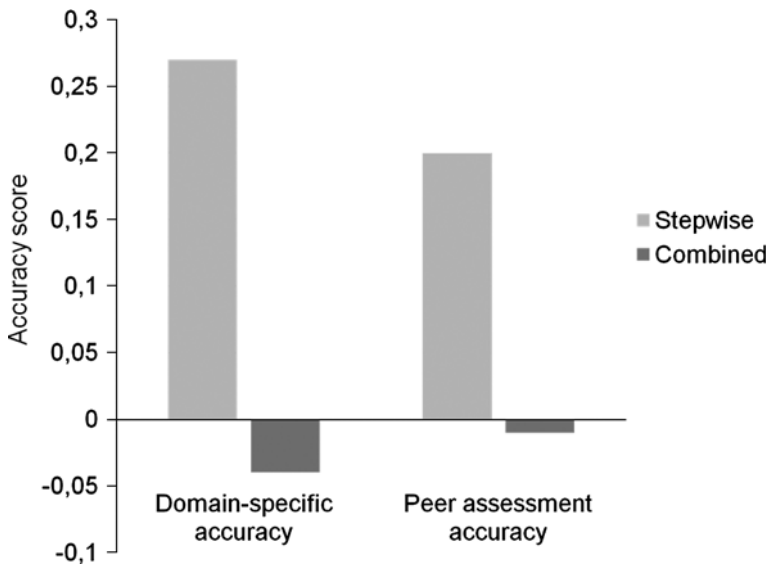


Figure 5. Mean differences in domain-specific accuracy and peer assessment accuracy between Phase 1 and Phase 2 for each instructional condition.

Student motivation

No significant effects were found for any of the four constructs measured by the questionnaire (attention, confidence, relevance and satisfaction). This supports our assumption of equivalence between the experimental groups. As task motivation was comparable between the two groups, there was no reason to control for any differences in motivation. The mean scores of the whole group were 2.68 (SD=.69) on attention, 3.31 (SD=.71) on confidence, 3.36 (SD=.62) on relevance and 2.68 (SD=.80) on satisfaction.

Conclusions and discussion

This study investigated an instructional method for peer assessment of complex tasks for novice learners by comparing test performance during two different two-phase instruction methods: a combined method with domain-specific and peer assessment tasks in both phases and a stepwise method, which had no peer assessment tasks in the first phase but was otherwise identical. We hypothesised that the stepwise instruction method would yield more improvement of performance and a higher final test task performance than the combined instruction method. We did indeed find more performance improvement in accuracy and time with the stepwise instruction method. These findings provide support for preferring the stepwise instruction method over the combined instruction method for teaching novice students peer assessment skills for complex tasks.

Our multivariate results show more performance improvement with stepwise instruction compared to combined instruction, but no significant differences in final performance. The univariate results show no significant differences between the two methods in domain-specific or peer assessment test *accuracies*. Yet, students in the stepwise condition performed at a slightly higher level of accuracy and needed less time for their performances. It is an interesting point for future research to discover how test accuracies can be further improved. In the next paragraph, we propose some explanations for the absence of significant findings for final test performance and make some suggestions for future studies.

One possible cause for the absence of significant differences between the instructional methods in final performance is task complexity. As all students performed reasonably well on the test tasks, the question might be raised of whether the study tasks were sufficiently complex for the participants in this study to observe the predicted effects. If the tasks had been more complex, it is well conceivable that significant differences in test performance would have been found. Another explanation may be the short duration of the instruction. There were only six study tasks, of which three differed between the two instructional conditions. Maybe differences in accuracy would have occurred with longer training periods and more tasks in each condition. Also, the test tasks may not have been sufficiently sensitive to measure accuracy differences resulting from only three distinctive study tasks. Further studies of stepwise instruction for peer assessment should address the potential effects on test accuracies of using longer training periods, more complex tasks and more sensitive measurement instruments.

It is important to consider the generalisability of the findings to authentic settings. In the current experiment, peer assessment was performed anonymously in an electronic learning environment. The assessments were not communicated to the assessees, and the peer solutions that were judged were artificial solutions provided

by the researchers and not by real peers. Hence, social and cognitive factors that may play a role in more authentic peer assessment situations (e.g. Johnson and Winterbottom 2011) were neglected. As social interaction and communication are significant aspects of peer assessment skills, future research should investigate the stepwise instruction strategy in the presence of these factors.

The focus of the current study was on one particular aspect of peer assessment: judging the performance of a peer. Although the domain-specific tasks were relatively complex, the actual peer assessment tasks were relatively simple. In reality, peer assessment constitutes more than just judging peer performance. For example, Sluijsmans (2002) identified *three* important sub-skills of peer assessment: (1) defining assessment criteria, (2) judging the performance of a peer and (3) providing feedback for future learning. It might therefore be worthwhile to examine possible advantages of the stepwise instruction method with regard to the sub-skills “defining assessment criteria” and “providing feedback for future learning”. Adding these sub-skills would also make the peer assessment task more complex, probably increasing cognitive overload and thus making the advantage of the stepwise instruction more visible. It remains a challenge to discover potential effects of instructional methods aimed at reducing cognitive overload – like the stepwise instruction method – in settings with not only complex domain-specific tasks, but also more complex peer assessment tasks.

To conclude, the current study provides some insight into a possible solution to the cognitive overload problem with instruction in peer assessment of complex domain-specific tasks. A stepwise instruction method was found to enhance performance improvement. Apparently, if students are first presented with only domain-specific information and later with both domain-specific and peer assessment information, they need more time to achieve slightly higher accuracies compared to students receiving only combined instruction. Educational practitioners should therefore consider that it may be beneficial to teach students domain-specific skills *before* starting on peer assessment skills.

Acknowledgement

This research was supported by the Netherlands Organisation for Scientific Research (NWO) under project number 411-05-110.

Notes on contributors

Marjo van Zundert was working as a PhD candidate at the Department of Educational Development and Research, Graduate School of Health Professions Education of Maastricht University. Her main research interests are peer assessment and instructional design. After recently having completed her PhD thesis “Conditions of peer assessment for complex learning”, she currently works as an educationalist at Leeuwenborgh Institute for Vocational Education and Training.

Karen Könings works as an assistant professor at the Department of Educational Development and Research, Graduate School of Health Professions Education of Maastricht University. Her PhD thesis “Student perspectives on education: implications for instructional redesign” (2007) received the dissertation award of the Netherlands Educational Research Association. Her main research interests are the participatory design of instruction (to better account for students’ perspectives), the role of student expectations on their later perceptions of education, and effectiveness of blended learning. She is coordinating the project “Learning and Working” in which different sub-projects are aimed to promote learning on

the workplace by including e-learning tools for students in both the medical curriculum and the health sciences curriculum. She is also involved in the Master of Health Professions Education.

Dominique Sluijsmans works as an associate professor at the Faculty of Education at HAN Professional University of Applied Sciences. Her PhD thesis “Student involvement in assessment: The training of peer assessment skills” (2002) received the dissertation award of the Netherlands Educational Research Association. Her main area of expertise is sustainable assessment. She received a research grant, and has given many lectures for a diverse audience in higher and vocational education. Her doctoral address “Involved in assessment” presents the ambition for a holistic perspective on assessment in education, which aims at an active participation of students in assessment practice for professional and sustainable learning. The design of assessment, the communication between learners and teachers about assessment and self-directed learning in assessment are key themes. Furthermore, she is the co coordinator of the EARLI SIG 1 “Assessment and evaluation”.

Jeroen van Merriënboer works as a full professor at the Department of Educational Development and Research, Graduate School of Health Professions Education of Maastricht University, where he is holding a chair in learning and instruction. He is currently the program director of Research in Education (RiE), the research and PhD program of the School of Health Professions Education (SHE). He is also the scientific director of the Interuniversity Centre for Educational Research, a collaboration of 10 Dutch universities offering a joint course program for PhD students in educational sciences. His main area of expertise is learning and instruction, in particular instructional design and the use of new media in innovative learning environments. He has published widely on four-component instructional design, cognitive load theory, and lifelong learning. He holds several academic awards for his research and has been a supervisor for over 25 PhD students.

References

- Anderson, L.W., and D.R. Krathwohl, eds. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman.
- Bloom, B.S., and D.R. Krathwohl. 1956. *Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. Handbook 1: Cognitive domain*. New York, NY: Longmans.
- Fletcher, J.D. 2009. The value of expertise and expert performance. A review of evidence from the military. In *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments*, ed. K.A. Ericsson, 449–69. New York, NY: Cambridge University Press.
- Gagné, R.M. 1968. Learning hierarchies. *Educational Psychologist* 6: 1–9.
- Gagné, R.M. 1985. *The conditions of learning*. New York, NY: Holt/Rinehart & Winston.
- Johnson, N., and M. Winterbottom. 2011. Supporting girls' motivation in science. A study of peer- and self-assessment in a girls-only class. *Educational Studies* 37: 391–403.
- Keller, J.M. 1983. *Development and use of the ARCS model of motivational design*. Enschede: Toegepaste Onderwijskunde, Technische Hogeschool Twente.
- Paas, F., A. Renkl, and J. Sweller. 2003. Cognitive load theory and instructional design: Recent developments. *Educational Psychologist* 38: 1–4.
- Paas, F., and T. Van Gog. 2006. Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction* 16: 87–91.
- Pollock, E., P. Chandler, and J. Sweller. 2002. Assimilating complex information. *Learning and Instruction* 12: 61–86.
- Scardamalia, M. 2001. Big change questions: Will educational institutions, within their present structures, be able to adapt sufficiently to meet the needs of the information age? *Journal of Educational Change* 2: 171–6.
- Sluijsmans, D.M.A. 2002. Student involvement in assessment: The training of peer assessment skills. PhD diss., Open Univ. of The Netherlands.

- Sluijsmans, D.M.A., and F. Prins. 2006. A conceptual framework for integrating peer assessment in teacher education. *Studies in Educational Evaluation* 32: 6–22.
- Smith, P.L., and T.J. Ragan. 2000. The impact of R.M. Gagné's work on instructional theory. In *The legacy of Robert M. Gagné*, ed. R.C. Richey, 147–81. Syracuse, NY: ERIC Clearinghouse on Information & Technology.
- Sweller, J. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12: 257–85.
- Sweller, J. 2010. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review* 22: 123–38.
- Sweller, J., J.J.G. Van Merriënboer, and F. Paas. 1998. Cognitive architecture and instructional design: Recent developments. *Educational Psychology Review* 10: 251–96.
- Tabachnick, B.G., and L.S. Fidell. 2007. *Using multivariate statistics*. Boston, MA: Pearson Education.
- The Netherlands Ministry of Education, Culture, and Science. 2011. *Action plan teacher 2020*. <http://www.rijksoverheid.nl/ministeries/ocw/documenten-en-publicaties/kamerstukken/2011/05/23/actieplan-leraar-2020.html>
- Topping, K.J. 2009. Peer assessment. *Theory Into Practice* 48: 20–7.
- Tsvitanidou, O.E., Z.C. Zacharia, and T. Hovardas. 2010. Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction*, Advance online publication. doi: 10.1016/j.learninstruc.2010.08.002.
- Van Merriënboer, J.J.G., L. Kester, and F. Paas. 2006. Teaching complex rather than simple tasks: Balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology* 20: 343–52.
- Van Merriënboer, J.J.G., and P.A. Kirschner. 2007. *Ten steps to complex learning: A systematic approach to four-component instructional design*. Mahwah, NJ: Lawrence Erlbaum.
- Van Merriënboer, J.J.G., and D.M.A. Sluijsmans. 2009. Toward a synthesis of cognitive load theory, four-component instructional design, and self-directed learning. *Educational Psychology Review* 21: 55–66.
- Van Merriënboer, J.J.G., and J. Sweller. 2010. Cognitive load theory in health professions education: Design principles and strategies. *Medical Education* 44: 85–93.
- Van Zundert, M.J., D.M.A. Sluijsmans, and J.J.G. Van Merriënboer. 2010. Effective peer assessment processes: Research findings and future directions. *Learning and Instruction* 20: 270–9.
- Van Zundert, M.J., D.M.A. Sluijsmans, K.D. Könings, and J.J.G. Van Merriënboer. 2012. The differential effects of task complexity on domain-specific and peer assessment skills. *Educational Psychology* 32: 137–45.

Appendix A

Domain-specific instructional format of the worked examples

General format

This is step ... [*step number*] ... , the ... [*step name*] ... In the ... [*step name*] ..., it is ... [*explanation of step*] ... This is the case here, because ... [*link to description of experiment*] ...

Sample

This is step 6, the conclusion. In the conclusion, the researcher compares the results of the experiment with the hypothesis that was formed. That is indeed the case here, because John finds out that taking vitamin C is beneficial: the results are compared with the hypothesis.

Appendix B

Peer assessment instructional format of the worked examples

General format

If you have to assess whether your classmate's answer is correct, first of all you check whether or not your classmate entered the correct research step. That is ... [*the case/not the case*] ... here: this is namely step ... [*step number*] ..., the ... [*step name*] ..., because ... [*explanation*] ... Then, you check whether or not your classmate added an argumentation (i.e. why does s/he think this is a certain step?). Your classmate ... [*did/did not*] ... do this: s/he writes that ... [*argumentation peer*] ... Finally, you decide whether or not your classmate's argumentation is correct. The argumentation of your classmate is ... [*correct/incorrect*] ... Hence, your classmate's solution is ... [*completely correct/completely incorrect/partially correct*] ...

Sample

If you have to assess whether the solution of your classmate is correct, first of all check whether your classmate has chosen the correct step. That is indeed the case here: this is namely step 6, the conclusion. Subsequently, you check whether he or she provided an argumentation for the chosen step. Your classmate did this too: he or she writes that this is the conclusion because John finds out that taking vitamin C is beneficial. The results are compared with the hypothesis. Finally, you decide whether the argumentation is correct. This argumentation is correct as well. Hence, your classmate's solution is completely correct.