Education Tech Research Dev (2010) 58:311–324
DOI 10.1007/s11423-008-9099-0

DEVELOPMENT ARTICLE

Formative assessment in an online learning environment to support flexible on-the-job learning in complex professional domains

Tamara van Gog · Dominique M. A. Sluijsmans · Desirée Joosten-ten Brinke · Frans J. Prins

Published online: 25 July 2008

Abstract This article describes a blueprint for an online learning environment that is based on prominent instructional design and assessment theories for supporting learning in complex domains. The core of this environment consists of formative assessment tasks (i.e., assessment *for* learning) that center on professional situations. For each professional situation, three levels of situational complexity are defined, and within each of these three levels, tasks are offered that differ in the degree of support offered to the learner. This environment can support (beginning) professionals in complex domains in gaining insight into the available repertoire of behavior in professional situations, as well as into the quality and effectiveness of that behavior (assessment criteria), while simultaneously helping them to develop insight into the standards that their own behavior should (eventually) match.

Keywords Online learning · Formative assessment · Instructional design

Introduction

Continuing to learn on the job is increasingly required in many professions (cf. life-long learning; Boud 1990), and to do so requires the ability to effectively self-regulate ones own learning process (i.e., acquire self-regulated learning competence; Van den Boom et al. 2004; Zimmerman 2002). An important pre-requisite for self-regulation, in turn, is the ability to accurately diagnose ones own needs for performance improvement, that is, to self-assess. The question that led to the development of the online learning environment described here, was: How can we support (beginning) professionals, who are working in highly demanding jobs in various places, and who have different levels of competency

T. van Gog (✉) · D. M. A. Sluijsmans · D. Joosten-ten Brinke · F. J. Prins
Ruud de Moor Center for the Professionalization of Teachers, and Educational Technology Expertise Center, Open University of the Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands
e-mail: tamara.vangog@ou.nl

Present Address:
F. J. Prins
Department of Pedagogical and Educational Sciences, Utrecht University, Utrecht, The Netherlands

development and different learning needs, in continuing their professional development and acquiring (self-)assessment skill?

This online learning environment is based on prominent instructional design (i.e., the 4C/ID model; Van Merriënboer 1997) and assessment theories (see e.g., Segers et al. 2003) for supporting learning in complex domains. We will first discuss complexity in terms of domain and task characteristics as well as in terms of the cognitive demands these characteristics pose on the learners. Then, we will discuss the design of the online learning environment in terms of components, ways of use, and the underlying theoretical rationale for why these components and ways of use provide optimal support for professional learning and learning to (self-)assess in complex domains.

Learning in complex domains

Complexity can be defined in three different ways. One can interpret it as being primarily subjective, that is, on psychological dimensions such as perceived task significance and identity, as primarily objective, that is, as a function of objective task characteristics, or as an interaction of subjective and objective characteristics (Campbell 1988). The objective interpretation is based on the degree of structuredness of a task or the possibility of multiple solution paths (Byström and Järvelin 1995; Campbell 1988). When the process of task performance in a domain can be described in detail a priori (i.e., is well-structured), it is considered less complex, in contrast, when it is less structured or ill-structured it is highly complex. In unstructured or ill-structured domains, such as teaching, medicine, or engineering, it is often the case that multiple paths can lead to an acceptable solution for tasks, or even that multiple solutions are possible. A teacher, for example, needs to use his or her conceptual knowledge structures of classroom management, of learning content, and of pedagogical and instructional strategies simultaneously. Moreover, this needs to be done in constantly varying situations, in which there is always another party involved (mostly students, but sometimes also parents or colleagues), that is reacting on the teacher's performance and thereby influencing the outcome.

In terms of Van Merriënboer's (1997) Four Component Instructional Design-model, (4C/ID-model; see also De Croock et al. 2002; Van Merriënboer et al. 1992, 2002), complex domains are characterized by a large number of non-recurrent skills, that is, skills that have to be applied differently and flexibly from situation to situation, whereas less complex domains are characterized by a large number of recurrent skills, which are performed similarly from situation to situation and can therefore be automated. This ability to flexibly apply parts of previously learned skills in solving new problems or problems new situations (i.e., transfer; Mayer and Wittrock 1996), requires that learners not only learn how to complete a task, but are also able to recognize the situational conditions under which (parts of) task completion procedures are applicable (i.e., when and why). Because the application of non-recurrent skills varies from situation to situation, these can, in contrast to recurrent skills (e.g., typing) not be automated. Since conscious processing requires more cognitive capacity than automated processing, learning in complex domains with a large number of non-recurrent skills imposes a high cognitive load on the learner (Sweller 1988; Sweller et al. 1998; Van Merriënboer and Sweller 2005).

An important element of the 4C/ID model is reliance on authentic, meaningful, and whole complex tasks in learning trajectories. Given the high cognitive load imposed by such tasks, they should be offered in such a way that learners are not cognitively overloaded by their complexity. That is, learners should be given the opportunity to practice simplified but increasingly complex versions of authentic whole tasks (Van Merriënboer 1997). This is accomplished by grouping tasks in task classes of different levels of complexity and scheduling those from lower to higher complexity, and by presenting multiple tasks within each task class that are of the same level of complexity, but differ in amount of instructional support offered. The first task starts with a high level of support, such as a process-oriented worked-out example showing the learner the solution procedure to be used, and the rationale behind the selection of the solution steps (Van Gog et al. 2006, 2008). The next task offers a little less support, for example a worked example that no longer explicates the rationale, but only shows solution steps. The next task provides even less support, for example, a completion problem that shows part of the solution procedure in which learners have to complete the missing steps (Paas 1992), or a process worksheet outlining the steps to be taken (but these steps are not worked out; Hummel et al. 2004), and the final task does not contain any support (i.e., conventional problem solving). A related but slightly different way to achieve different levels of support is to first limit the demands placed on a learner (e.g., "in this situation, focus only on …") and gradually loosen these until the full task is performed. This principle of decreasing support not only prevents cognitive overload, but also gives learners insight into how tasks need to be or can be successfully completed.

In sum, it is important to confront learners with authentic, meaningful, whole tasks, but in such a way that they are not cognitively overloaded by the complexity of those tasks. We argue here that this should not be any different when the skill to be learned is to (self-)assess.

Learning through formative assessment

Insight into how tasks need to be or can be successfully completed, is also an important element of formative assessment. Whereas summative assessment is conducted after a learning phase (ranging from a single course to an entire curriculum) and serves accountability or certification purposes ('assessment *of l*earning'), formative assessment is conducted during a learning phase with the goal of promoting learning ('assessment *for* learning'; Birenbaum 1996; Sadler 1989, 1998). Both formative and summative assessment can be directed at the learners' level of acquired *knowledge*, however, the preference has started to shift in the past decades towards alternative assessment forms that allow measurement of acquired knowledge as well as cognitive and/or physical skills through *performance* on relevant whole tasks (see Segers et al. 2003).

Learning is fostered through formative assessment when it succeeds in helping learners identify their weaker and stronger points, and in guiding them to overcome the weaker points during the learning process (Boud 1990; Dierick and Dochy 2001; Topping 2003). This requires learners to develop an understanding of the performance criteria and standards, and helping them do so is a crucial aspect of formative assessment. That is, learners should know what aspects of performance should be assessed (criteria) and what constitutes poor, average, good or excellent performance on those aspects (standards; cf. Arter and Spandel 1992).

It follows logically from the above, that in order to be successful, instruction and assessment should be aligned (Arter 1996; Biggs 1996; Dochy and McDowell 1997), that is, learners should be able to practice the kinds of tasks that they will also be summatively assessed on, because understanding performance criteria will not help much if the tasks to be performed during assessment are not aligned with those during instruction. In case of formative assessment, instructional tasks and formative assessment tasks can be identical. Learners can work on a task, assess their performance on that task, decide which aspects of their performance need improvement, and select a next task to work on that will help them improve these performance aspects. They can also have their performance assessed, diagnosed and have the task selected by someone else (e.g., a teacher, supervisor, or peer; Sluijsmans et al. 1998, 2006).

In complex domains, however, the questions of how to perform a certain task or what constitutes "weak" and "strong" performance, are not answered easily, because these are also partly influenced by the changing situational demands. In other words, in complex domains, defining assessment criteria and standards is difficult, and so is learning to understand and apply them. To provide learners with an environment in which they can practice both their domain-specific and assessment skills while task complexity and instructional support are taken into account, an online learning environment blueprint was developed. This blueprint is based on principles of the 4C/ID model, and an environment developed based on this blueprint provides learners with formative assessment tasks with different levels of task demands (support) in combination with different levels of situational complexity (task classes). Through working on these tasks, learners can simultaneously improve their domain knowledge and skill, acquire knowledge of assessment criteria and standards for the domain, as well as (self-)assessment skill, by assessing the quality of their own or a peers' performance.

Components of the online learning environment

Below, the components of the environment are discussed. A schematic overview of how these components are interrelated is given in Fig. 1. For each professional situation (left 'column'), three levels of complexity are defined (second 'column' from the left), and at each level of complexity, three types of tasks are offered that differ in task demands/support (third 'column' from the left). The "Act" task is the task offering the least support (or placing the highest demands), because it has to be executed in the learner's own

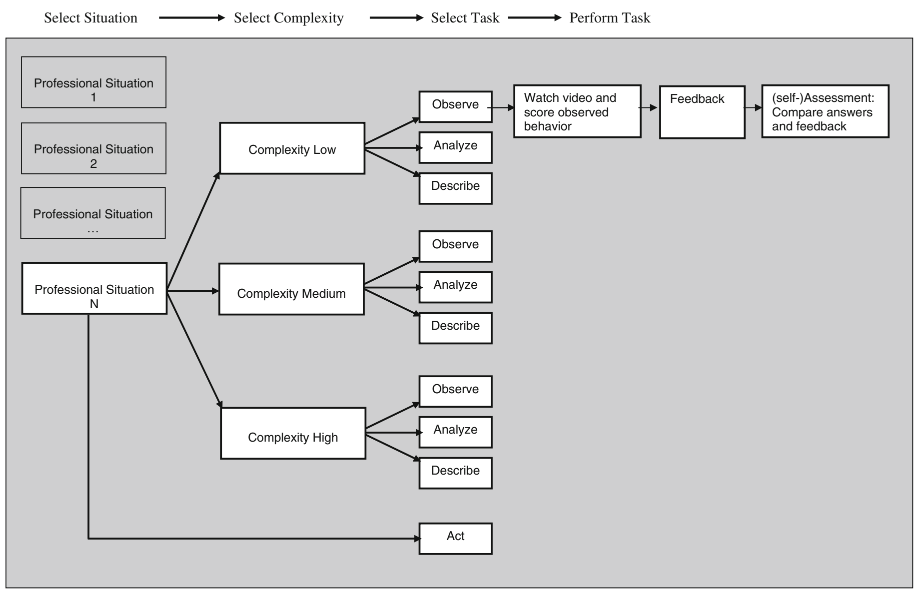

Fig. 1 Blueprint of the online learning environment components

professional real-world setting (performance on the job). This task is not tied to a situational complexity level, because the exact complexity of the real-world setting can be quite unpredictable, but the same assessment criteria and standards are applied and complexity of the situation can be rated and taken into account.

Authentic whole tasks: professional situations as a basis

For many professions, competency maps or profiles have been created. However, when performing in professional situations, competencies are not required in isolation, but rather, they interact. In teaching, for example, being didactically competent (e.g., giving good instruction) and pedagogically competent (e.g., being capable of maintaining order) often interact, and are both necessary for good performance outcomes (e.g., the students learning the particular topic being taught). For tasks to be "whole" tasks with a high degree of authenticity and relevance then, they should centre on performance in professional situations, instead of on isolated competencies.

Levels of situational complexity

With professional situations as a basis, it is important to provide tasks at different levels of situational complexity (cf. Van Merriënboer's 1997, "task classes"), for different reasons. For absolute beginners, scheduling situations from lower to higher complexity, is more optimal in terms of cognitive load and therefore will lead to better learning outcomes when this sequence is followed (see Van Merriënboer and Sweller 2005). Learners are not obliged to follow a fixed path through the online learning environment, but even with flexible paths, there are two more reasons for offering different levels of complexity. First, the complexity of the real-life professional situations that learners operate in is likely to differ from learner to learner, so multiple complexity levels are needed in order to provide a level that reflects an individual learner's situation. Secondly, it is desirable to offer experiences that differ from ones own personal situation. When ones daily professional environment is overall *lower* in situational complexity than other environments, it is important to realize that although the learner may be performing at a very high level, this might not necessarily transfer to more complex environments (which may pose a problem when changing jobs). When the learner's daily professional environment is overall *higher* in situational complexity than other environments, the learner may well be overwhelmed by this complexity. In this case, it is critical for a learner's feelings of mastery and success to realize that this is a highly complex environment, and to be able to experience what his or her performance in less complex situations might be like. Offering tasks at multiple levels of situational complexity will help to give learner a sense of the complexity of situations they are facing everyday, and will place their performance in broader perspective.

Unfortunately, however, determining situational complexity is not an easy task. We attempted to define objective factors that determine complexity, using a procedure described later on in the section "Acquiring input for the online learning environment".

Levels of task support

For beginners in a domain, it is important to offer a range of tasks that differ in the amount of support offered to learners (high to low), which can also be applied as differences in demands placed on learners (low to high; cf. Van Merriënboer's, 1997, principle of "decreasing supportive information within task classes"). Four types of tasks are delivered in the online learning environment, that place increasing demands on the learners (or give decreasing support), of which the first three are offered at each level of complexity, while, as indicated before, the fourth is not tied to a complexity level (see also Fig. 1):

- Observation tasks to assess ones knowledge about the available repertoire of performance
- Analysis tasks to assess ones insight into the conditions for effectiveness of performance
- Description tasks to assess ones own self-reported performance
- Hands-on practice tasks to assess ones own actual performance (not tied to a situational complexity level, but in the assessment criteria and standards complexity of this situation can be taken into account).

While there is no increase in complexity of the tasks within one level of situational complexity, the tasks do differ with regard to the amount of support provided to learners, by proceeding from assessment of knowledge of adequate performance (i.e., high performance restriction) to independent performance (i.e., no performance restriction). More information on how the tasks themselves were designed and had to be performed is provided in section "Using the online learning environment".

Acquiring input for the online learning environment

Although techniques such as Cognitive Task Analysis (CTA; see Clark et al. 2008; Cooke 1994) may be very helpful in defining lists of key professional situations, levels of complexity of situations, and task content, instructional designers will also have to collaborate closely with professionals to test their analysis and to fill in required details. A means to elicit the required information from a group of professionals is the 'structured expert session' (Goes et al. 2005). In this method, expert professionals with a substantial level of experience are asked to participate. They are chosen in such a way that a group of 5–10 individuals encompasses different levels of age, years of experience, and professional setting. Prior to each session, the experts complete a "homework assignment", that asks them to describe—from their own experience—three specific situations of different levels of complexity for a given professional situation. They are also required to state the factors that determine the level of complexity in their opinion. This assignment serves as input for a group session. This session takes ca. 2 h, in which it should first be attempted to come to agreement on the complexity factors, then on the appropriate repertoire of performance, and finally on the associated assessment criteria and standards. In combination with more objective techniques (e.g., CTA), these expert sessions provide a wealth of input data for designing the content of the online learning environment.

Using the online learning environment

In this section, we will shortly describe a possible user scenario for the online learning environment, which will provide some more insight into the tasks selection options and into how the tasks have to be performed.

Task selection

Upon entering the online learning environment, a learner first selects the professional situation s/he wants to work on from a list of available situations (Fig. 2). Then, s/he

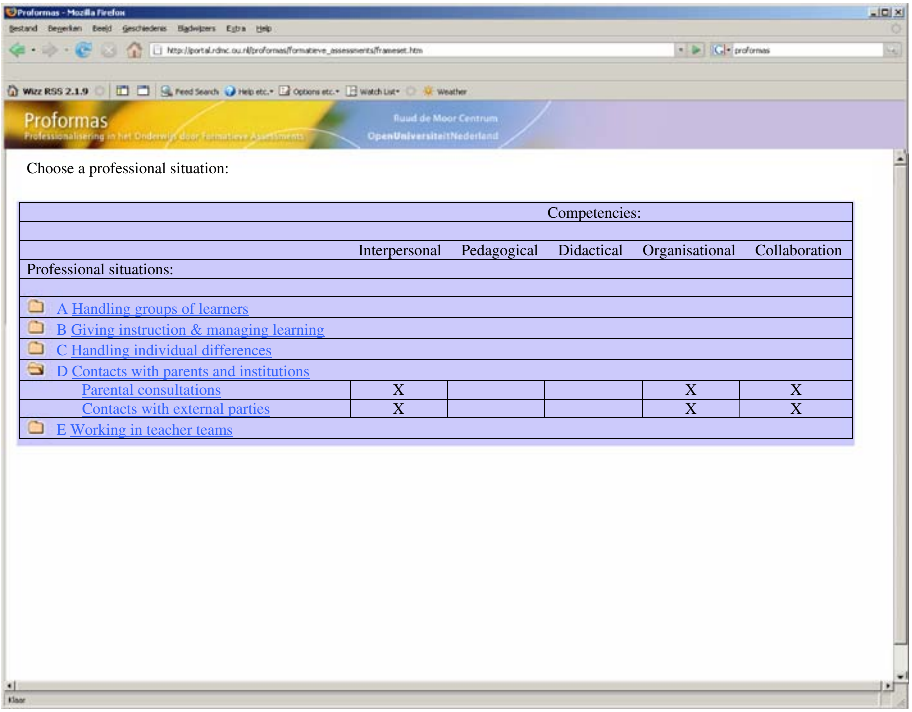

Fig. 2 Screenshot: Selecting a professional situation (NB: Original text on page is in Dutch)

selects the complexity level (Low, Medium, High) at which s/he wants to perform the tasks for this situation, or the Act task to be performed in practice (Fig. 3). After (when applicable) a complexity level is selected, the learner selects the type of task s/he wants to work on (Observe, Analyze, or Describe; Fig. 4).

Task performance

The Observe and Analyze tasks are video-based, thereby providing instances of realistic models in professional situations (Olivero et al. 2004). A video is shown of the professional situation, with a relevant model (e.g., for student teachers, this would be a teacher in the classroom, or during parental/colleague consultation). The learner can watch this video full-screen a first time (this is optional), and then (or immediately) proceed to a split-screen interface in which the video is shown again on one side, and text boxes on the other side, in which the learner can type his or her observations and/or analysis of the performance displayed by the model in relation to a number of performance criteria (Fig. 5). The video can be controlled, so it can be paused or replayed if necessary. As mentioned above, the difference between Observe and Analyze tasks is that Observe tasks require only observation of the model's performance (i.e., actions taken), whereas Analyze tasks also require an analysis of the effectiveness of those actions. The Describe task can be video or text-based. Merely a beginning of a situation is shown or verbally sketched in one half of the split-screen, based on which the learner is required to give a description of how s/he would handle this situation. This description can be typed in the textboxes in the other half.

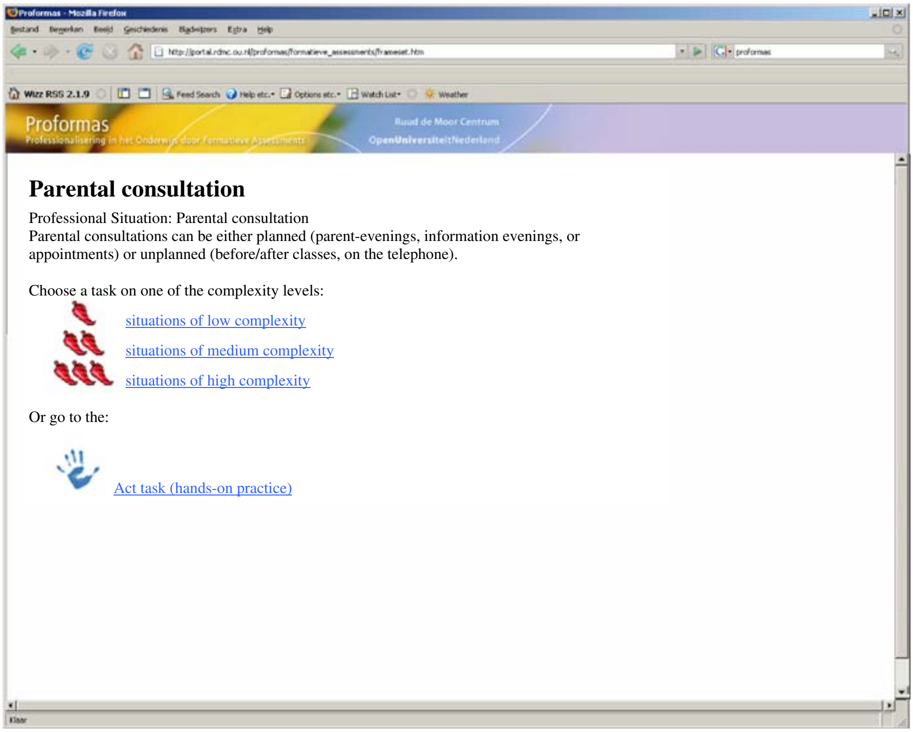

Fig. 3 Screenshot: Selecting a complexity level (NB: Original text on page is in Dutch)

When the learner submits his or her observations/analysis/self-description, feedback is provided in the form of an expert's observations/analysis/self-report, respectively. It is explicitly indicated that this expert solution is an example of a very good solution, but that other ways to reach that solution or other solutions may have been appropriate as well.

The Act task consists of a task assignment that the learner has to perform in a real-world setting. Because it is often not possible to chose or control situational complexity in these settings, the assignment is accompanied by a rating sheet for situational complexity (to be filled out by the learner afterwards or by a peer or teacher/mentor during task performance), so that the learner can get some indications regarding situational complexity. The environment provides the option to upload digital video recordings of the learner performing the Act task.

Assessment: self, peer or teacher

After receiving the feedback on the Observe, Analyze, or Describe tasks, the learner can either assess his or her own performance by comparing the expert's answer to his or her own, or can ask a peer or teacher/mentor to do so. This is again done by typing in textboxes while the answers and feedback are still available on screen. In this case, only task performance is assessed. However, it is also possible to first provide a self-assessment and then ask a peer or mentor to judge the quality of this self-assessment, in which case the peer- or teacher assessment is focusing on the learner's ability to (self-)assess.

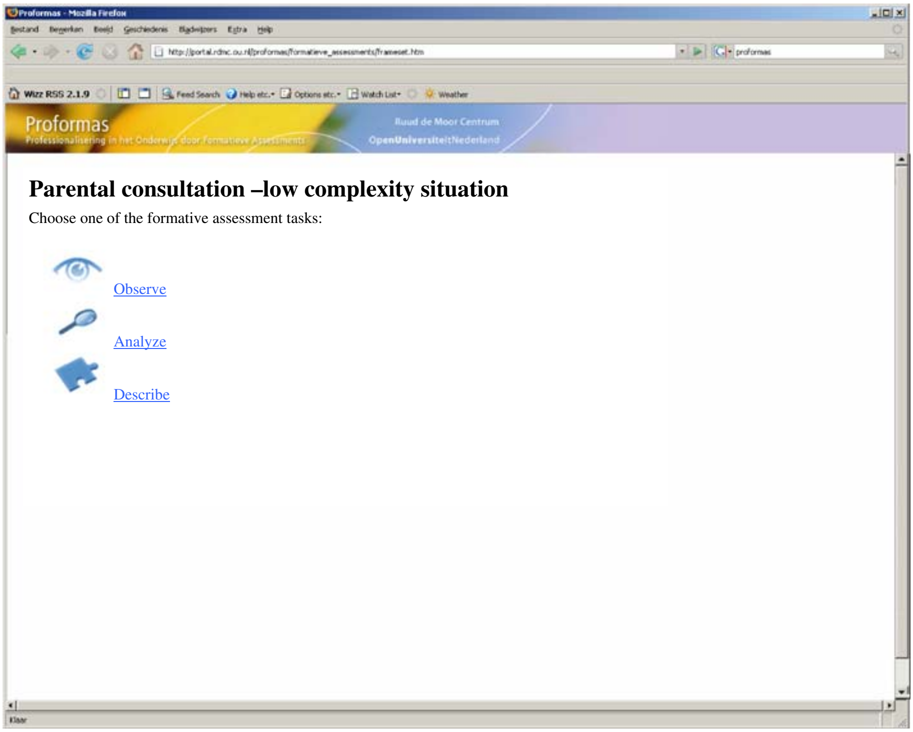

Fig. 4 Screenshot: Selecting a formative assessment task (NB: Original text on page is in Dutch)

For the Act task, performance can be rated by the learner afterwards (which may or may not be supported by a video record of the real-world performance), by a peer or teacher/mentor during task performance, or both, using a list of performance criteria provided along with the task and the complexity rating form. When both the learner and someone else assess performance (i.e., co-assessment), this is again an opportunity to evaluate not only the learner's performance, but also his or her ability to (self-)assess.

On the one hand, asking peers or teachers/supervisors to assess, may diminish the benefits of time and place independency an online learning environment can provide, because the learner has to wait for the peer or teacher/mentor. On the other hand, it offers the opportunity to circumvent a weakness inherent to online learning environments, especially in complex domains: the provision of *personalized* feedback. Moreover, based on the assessment outcomes, the selection of a suitable next situation, complexity level, and/or task can be discussed with the supervisor, which may have added benefits because being able to assess which aspects of performance should be worked on and being able to select a task that will address precisely those aspects, are two related but different skills.

Tracking system

When tasks have been completed, the learner can save his or her answers, feedback, and (self-)assessment in a text file in a personal folder in the online learning environment, or e-mail this file (in order to be able to save it in another location). A leaner can also check a task as completed and provide a rating for that task, and record of this is kept in a personal

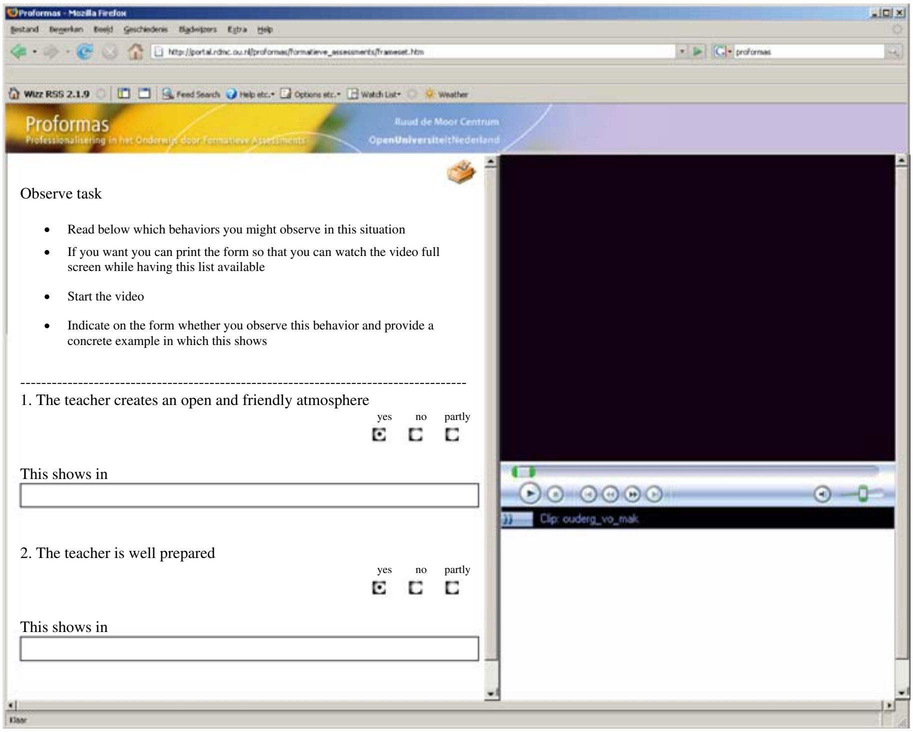

Fig. 5 Screenshot: Working on an Observe task (NB: Original text on page is in Dutch)

task overview, so that the learners can see which tasks they completed and what the outcome was. When the most important competencies tasks call on would be described in metadata, this overview could also document competency development. These folder and personal task overview functionalities can assist learners in selecting future tasks to work on.

User evaluation

The development principles underlying this online learning environment are applicable in a variety of complex professional domains, but we originally developed it in the context of teacher education. Beginning teachers have little time available for further 'formal professional learning' because of their teaching load, and the schools they teach at vary widely in terms of location, student population, and the types of education offered. The same goes for student teachers on internships, they also get their practical experience at a variety of schools. Moreover, in the Netherlands, beginning teachers may differ immensely in their amount of relevant experience, as some of them no longer follow 'traditional' teacher education trajectories. Some beginning teachers come from another professional background (business, industry, social services) and switch their career-path to teaching. They start teaching almost from day one, while simultaneously receiving education from Teacher Training Colleges for primary or secondary education (this was initiated by the government in order to reduce the imminent shortage of primary and secondary school teachers in the Netherlands). For these reasons, online education solutions were sought to support the professional development of the beginning teachers, one of which was the online learning environment for the acquisition of formative (self-)assessment skill described here. We called this environment "Proformas" (**PRO**fessional development of teachers through **FORM**ative **AS**sessments; Van Ingen et al. 2007).

This environment was tested in a small user evaluation study (Van Ingen et al. 2007). Thirty-eight participants (60% women) were given access to the environment, and were asked to use the environment for some time and to evaluate it based on a questionnaire. They could do so in a single session or in multiple sessions, at home or at school. Because of this, we unfortunately have no reliable data as to the amount of time they spent on the environment before filling out the questionnaire, but we do know that participants actually completed several tasks in the environment (mostly Observe and Analyze, but also Describe, and in some cases even the Act task). Half of the participants were teachers, half of them teacher educators. Results regarding the evaluation of the environment as a whole and the components of the environment are shown in Table 1 (results regarding layout/usability are not presented here given the scope of this article, but these were used to alter some technical aspects of the environment).

The online learning environment was evaluated positively. Participants indicated that it focused on relevant situations, was useful for learning about formative assessment and effective teacher behavior, and could contribute to further professional development of beginning teachers. The division in complexity levels was positively evaluated, and participants recognized the difference between Observe and Analyze tasks (which might seem very similar at first glance). They felt the amount of questions to be answered with the tasks (see Fig. 4 for an impression of the questions) was too high, but really appreciated the feedback option. Interestingly, only half of the participants used the flexibility offered by the system, the other half chose a fixed order. Note though, that this does not have to be problematic, because this order offers a logical learning path (simple to complex situations and high to low

Table 1 User evaluation of the environment

	Negative (%)	Positive (%)
Environment as a whole		
I understand how to use the environment to learn about formative assessment	5	95
The content of the situations is familiar	0	100
The environment gives insight into effective teacher behavior	3	97
The environment can contribute to professional learning of beginning teachers	0	100
I would recommend colleagues to use this environment	11	89
Components of the environment		
The division of situations (and as a result, tasks) in 3 complexity levels (low, medium, high) has added value	8	92
The indicated complexity levels were in line with how complex I perceived the situation to be	9	91
The difference between Observe and Analyze tasks was clear to me	21	79
I worked in the environment based on the presented order of situations and tasks	51	49
Too many questions need to be answered with each task	72	28
The feedback provided with the tasks was useful	25	75
I used the tracking system	76	24

support in tasks at each complexity level). Most of the participants did not use the tracking system, but indicated they probably would do so when using the environment for extended periods (i.e., when it really has a function for tracking personal development).

Discussion

To build ones own broad and flexible repertoire of knowledge and skills, it is imperative for (beginning) professionals to develop insight into the complexity of the domain, the performance repertoire available, and the quality standards for performance. Moreover, to continue professional development, it is necessary to be able to identify ones strengths and points for improvement. The online learning environment we described here, addresses all of the above by relying heavily on formative assessment of performance on authentic whole tasks tied to professional situations. By working on these tasks, learners can simultaneously develop domain knowledge and skills, knowledge of assessment criteria and standards employed in the domain, and (self-)assessment skill, that is, the ability to assess the quality of (their own) performance. This way, this online learning environment is believed to support flexible learning in complex domains.

Our online learning environment also has great benefits in terms of accessibility and flexibility. Regarding accessibility, it is imperative for (beginning) professionals who are often spread over different locations and are often quite overburdened with trying to cope with the new job demands, to have anytime anywhere access to the learning environment. Regarding flexibility, the learners may follow a logical linear path. For example they could chose to complete all tasks at the Low complexity level of a certain situation, then all tasks at the Medium complexity level and then all tasks at the High complexity level, before proceeding to the Act task. Or they could first work on the tasks of the Low complexity levels of all situations, before proceeding to the Medium level, et cetera. However, learners are not *forced* to follow any path. This ensures the usability of the environment for learners of different levels of expertise and/or different learning goals.

In sum, by providing formative assessment tasks with different levels of task demands in combination with different levels of situational complexity, this online learning environment aims to:

- allow learners to improve their performance
- allow learners to acquire self-assessment skill
- support learners to seek personalized feedback from peers or teachers on their task performance or assessment skill
- be flexible, by offering logically built-up sequences of tasks, but allowing full choice
- be accessible, by allowing learners to work on the tasks whenever they wish, wherever they have Internet access.

Another benefit of the online learning environment we described here, is that it can be widely implemented. The principles underlying this environment are applicable in many if not all complex domains, so based on this "blueprint", online learning environments can be designed for many different professional domains.

We presented some evaluation data based on our implementation of the environment in the domain of teacher education. Future research should aim to uncover the effects on learning of this online learning environment in different complex domains. These effects can be studied both in terms of learning outcomes, that is, task performance and assessment ability, and in terms of how learners interact with the environment.

Acknowledgments The authors gratefully acknowledge the contributions of the Project Team Assessment RdMC in the realization of this environment. In particular, they would like to thank Marieke Dresen and Jo Knarren, for their contributions to the discussions that led to the development of the online learning environment as it is described here, and Sonja van Ingen, for her contributions to the small-scale user evaluation described here.

References

Arter, J. (1996). Using assessment as a tool for learning. In R. Blum & J. Arter (Eds.), *Student performance assessment in an era of restructuring* (pp. 1–6). Alexandria, VA: Association for Supervision and Curriculum Development.

Arter, J. A., & Spandel, V. (1992). An NCME instructional module on: Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice, 11*, 36–45. doi:10.1111/j.1745-3992.1992.tb00230.x.

Biggs, J. (1996). Assessing learning quality: Reconciling institutional, staff and educational demands. *Assessment & Evaluation in Higher Education, 21*, 5–15. doi:10.1080/0260293960210101.

Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3–29). Boston: Kluwer Academic Publishers.

Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education, 15*, 101–111. doi:10.1080/03075079012331377621.

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management, 31*, 191–213. doi:10.1016/0306-4573(94)00041-Z.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of Management Review, 13*, 40–52. doi:10.2307/258353.

Clark, R. E., Feldon, D. F., Van Merriënboer, J. J. G., Yates, K. A., & Early, S. (2008). Cognitive task analysis. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd rev. ed., pp. 577–593). New York: Erlbaum/Routledge.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies, 41*, 801–849. doi:10.1006/ijhc.1994.1083.

De Croock, M. B. M., Paas, F., Schlanbusch, H., & Van Merriënboer, J. J. G. (2002). ADAPT^{IT}: Tools for training design and evaluation. *Educational Technology Research and Development, 50*(4), 47–58. doi:10.1007/BF02504984.

Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation, 27*, 307–329. doi:10.1016/S0191-491X(01)00032-3.

Dochy, F., & McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation, 23*, 279–298. doi:10.1016/S0191-491X(97)86211-6.

Goes, M., Dresen, M., & Van der Klink, M. (2005). Zonder leraren geen meesterlijke ontwikkeling: Het uitwerken van kenmerkende beroepssituaties [Professional development of teachers: Elaborating characteristic professional situations]. Working paper Ruud de Moor Centre, Heerlen, The Netherlands.

Hummel, H. G. K., Paas, F., & Koper, E. J. R. (2004). Cueing for transfer in multimedia programmes: Process worksheets versus worked-out examples. *Journal of Computer Assisted Learning, 20*, 387–397. doi:10.1111/j.1365-2729.2004.00098.x.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47–62). New York: Macmillan.

Olivero, F., John, P., & Sutherland, R. (2004). Seeing is believing: Using videopapers to transform teachers' professional knowledge and practice. *Cambridge Journal of Education, 2*, 179–191. doi:10.1080/03057640410001700552.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology, 84*, 429–434. doi:10.1037/0022-0663.84.4.429.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144. doi:10.1007/BF00117714.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles. Policy & Practice, 5*, 77–84. doi:10.1080/0969595980050104.

Segers, M., Dochy, F., & Cascallar, E. (2003). *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Sluijsmans, D. M. A., Dochy, F., & Moerkerke, G. (1998). Creating a learning environment by using self-peer- and co-assessment. *Learning Environments Research, 1*, 293–319. doi:10.1023/A:1009932704458.

Sluijsmans, D. M. A., Prins, F. J., & Martens, R. (2006). A framework for integrated performance assessment in e-learning. *Learning Environments Research, 9*, 45–66. doi:10.1007/s10984-005-9003-3.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–295. doi:10.1023/A:1022193728205.

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 55–89). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Van den Boom, G., Paas, F., Van Merriënboer, J. J. G., & Van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior, 20*, 551–567. doi:10.1016/j.chb.2003.10.001.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction, 16*, 154–164. doi:10.1016/j.learninstruc.2006.02.003.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction, 18*, 211–222. doi:10.1016/j.learninstruc.2007.03.003.

Van Ingen, S., Joosten-ten Brinke, D., Schildwacht, R., & Knarren, J. (2007). Formatieve assessments voor docenten: Een evaluatierapport [Formative assessments for teachers: An evaluation report]. Working paper, Ruud de Moor Center, Open University of The Netherlands, Heerlen, The Netherlands.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

Van Merriënboer, J. J. G., Clark, R. E., & De Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID model. *Educational Technology Research and Development, 50*(2), 39–64. doi:10.1007/BF02504993.

Van Merriënboer, J. J. G., Jelsma, O., & Paas, F. (1992). Training for reflective expertise: A four-component instructional design model for training complex cognitive skills. *Educational Technology Research and Development, 40*(2), 23–43. doi:10.1007/BF02297047.

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*, 147–177. doi:10.1007/s10648-005-3951-0.

Zimmerman, B. J. (2002). Achieving academic excellence: A self-regulatory perspective. In M. Ferrari (Ed.), *The pursuit of excellence through education* (pp. 85–110). Hillsdale, NJ: Erlbaum.

Tamara van Gog is associate professor at the Educational Technology Expertise Center of the Open University of the Netherlands (OUNL). Her research focuses on instructional design, cognitive load and cognitive processes, expertise (development), and (self-)assessment. The work presented here was conducted at the Ruud de Moor Center for the Professionalization of Teachers at the OUNL.

Dominique M. A. Sluijsmans is assistant professor at the Educational Technology Expertise Center of the Open University of the Netherlands and lector (associate professor) at the Professional University of Arnhem-Nijmegen. Her research interests encompass sustainable assessment and lifelong learning, with a focus on the design of sustainable assessment and the involvement of learners through self- and peer assessment.

Desirée Joosten-ten Brinke studied Applied Educational Science at the University of Twente. Since 1990, she works at the Educational Technology Expertise Center of the Open University of the Netherlands. In October 2004, she started her PhD project on Assessment of Prior Learning.

Frans J. Prins is now assistant professor at the Faculty of Social and Behavioural Sciences, Department of Pedagogical and Educational Sciences, Utrecht University, The Netherlands. His research interests include metacognition and self-regulated learning, feedback, peer assessment, and competence-based education.