The Visual Assessment Method for Maintenance training (VAMM)

Nieke Janssen & Eddy Boot
TNO Defence, Security and Safety
Soesterberg, The Netherlands
nieke.janssen@tno.nl, eddy.boot@tno.nl

Toine Broers
Royal Netherlands Air Force
Volkel, The Netherlands
ajm.broers@mindef.nl

ABSTRACT

Observing maintenance competencies in training settings is difficult. Many critical attitudes such as safety-awareness or team skills are covert and occur irregularly. As a result, assessing such competences for learning and testing often is subjective and not transparent. TNO and the Royal Netherlands Air Force (RNLAF) together have created a method for learning and assessing the development of such critical attitudes and skills, called the Visual Assessment Method for Maintenance training (VAMM). In this method, video-cameras are set up in the workshop, monitoring a team of learners doing realistic F16 maintenance tasks. One of the team members is responsible for evaluation of task performance. A commercial computer-based tagging tool (Darthfish) is used to mark incoming video events, and quickly classify them according to predefined evaluation criteria such as 'bookwork discipline' or 'team co-operation'. A selection of marked video-episodes is used in an After Action Review (AAR), showing examples of good and bad displays of attitudes and skills. Both task performers and evaluator can learn from this reflection process. Instructors may apply the method in exam settings, providing an objective and transparent manner of assessing realistic tasks. The VAMM method is currently being tested by means of a pilot project at a RNLAF school with two teams of four learners. The first results show that learners are capable to assess mutual task performance by means of video analyses; (a) a sufficient number of events was tagged for proper AAR and examinations and (b) the tagged events covered the most important attitudes and skills relevant for maintenance training to be learned. It can be concluded that the VAMM method has great potential in the maintenance training domain. Currently, an experimental design study is set up to test the impact of the method more rigorously.

ABOUT THE AUTHORS

Nieke Janssen is a scientific researcher at TNO Human Factors, and is involved in research and development projects concerning the application of Information and Communication Technology (ICT) to improve learning processes. She holds a M.Sc. in cognitive science and specializes in integrating and applying instructional theories in the design of multimedia learning environments. In her research she focuses on learner centered approaches to learning such as self-organised, peer-to-peer, and ubiquitous learning.

Eddy Boot is a researcher at TNO Human Factors, and is involved in R&D projects concerning the application of Information and Communication Technology (ICT) to improve learning processes. He holds a Ph.D. in instructional technology and specializes in complex learning and competency-based learning by means of advanced learning technology. Much of his research is related with the integration of work and learning and ubiquitous learning.

Toine Broers is an experienced instructor at the Royal Netherlands Air Force (RNLAF). He is about to finish his bachelor in education and specializes on the application of competency-based learning methods in F16 workshop learning. From his role of instructor he is involved in many instructional design and development projects for complex technical training. Currently he is responsible for the development of a new training program for learning on the job.

The Visual Assessment Method for Maintenance training (VAMM)

Nieke Janssen, Eddy Boot
TNO Defence, Security and Safety
Soesterberg, The Netherlands
Nieke.janssen@tno.nl, eddy.boot@tno.nl

Toine Broers
Royal Netherlands Air Force
Volkel, The Netherlands
ajm.broers@mindef.nl

INTRODUCTION

A trend in practical technical training programs for complex system maintenance is to immerse the learners as much as possible in authentic work environments (cf. the principles of 'situation based learning' or 'situated cognition' (Brown, Collins, & Duguid, 1989; Lave, J., & Wenger, E. (1991); Vygotsky, 1978). Learning in the real workplace is often too dangerous or inefficient, and therefore learning takes often place in workshops; practice rooms at schools that mimic the real workplace. Acquisition of competencies is done by offering realistic learning tasks that integrate sets of (meta-cognitive) skills, knowledge and attitudes.

Assessment of these competences is often limited to practical tests, mostly at the end of the training. An individual instructor is responsible for observing and assessing learner's task performance and progress in acquiring competences, mostly in a one-to-one setting. However, complex maintenance tasks are often team tasks, involving complex cognitive skills (Van Merriënboer & Kirschner, 2007). Besides team skills, team-related attitudes and cognition play a significant role in effective team work. Measuring these types of team competences is difficult because of covert behaviors regarding to team members' state of mind (e.g. situation awareness, team monitoring) (Cannon-Bowers & Sales, 1998; Smith-Jentsch et al (p81)). Currently, assessment of these competences in realistic task contexts is often subjective, not transparent, and leaves little space for self-reflection of the learner or second opinions. Besides this, it conflicts with new approaches to learning that aim at increasing responsibility of the learner for their own learning processes. Embedding assessment activities, like peer-assessment, in the learning process, would help instructors and learners to keep track of learning progress and increase self-regulation in learning. Moreover, it would increase objectivity in After Action Review (AAR).

The Royal Netherlands Air Force (RNLAF) is currently facing assessment problems in their training program for F16 maintenance engineers. In the current training program, there is much focus on acquiring theory about the F16's complex systems and little attention to the integration of theoretical knowledge and practical skills, particularly with regard to the social and physical context in which the task will be applied. Together with TNO Human Factors, the Air Force searched for a solution to embed assessment of generic maintenance competences in the process of learning and testing. The results of a case study on this topic are described in this paper.

BACKGROUND

The F16 maintenance engineers of the RNLAF are trained to work in small teams, performing (a) preventive maintenance and, when necessary, (b) fault isolation and/or (c) perform corrective actions. The emphasis lies on using the correct Technical Orders (TO's) with the official (up-to-date) procedures. Therefore, training should mainly be focused on acquiring generic maintenance competences such as working accurately with these TOs, in small teams, and upholding the highest level of safety awareness. Amongst others, this implies that the new maintenance engineers must acquire a professional attitude; learning to trust team members as well as be critical towards each other.

The Four Component Instructional Design (4C/ID) model (van Merriënboer & Kirschner, 2007) provides a model for designing competency-based learning programs. It emphasizes a whole-task approach, in which meaningful tasks are performed in authentic (simulated) task environments. The 4C/ID model provides detailed, prescriptive guidelines for implementing technical training. The four components of the 4C/ID model are (a) learning tasks, (b) part-task practice, (c) supportive information, and (d) Just-in-Time (JIT) information. TNO and the RNLAF used the 4C/ID model to develop a method for learning and assessing the development of maintenance skills, which has been called the Visual Assessment Method for Maintenance training (VAMM).

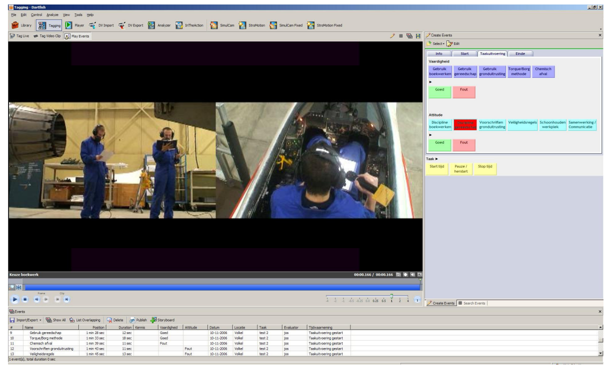
VAMM is an evaluation method that combines a visual evaluation approach and guided peer-to-peer

assessment. VAMM uses video techniques to record task performance. The video episodes are displayed on a PC monitor for evaluation and assessment by a peer-learner.
VAMM focuses on this new generation of learners who have a visual way of thinking and working. Images and audio are becoming more and more part of the communication culture of these learners who communicate via web channels such as web logs, pod- and vodcasting, and YouTube (Goldschmeding et al, 2006). *Streaming video* in the learning process can support these learners in reflecting on their own learning processes. The process of evaluation is supported and structured by a computer-based tagging tool with pre-structured evaluation criteria. A selection of marked video-episodes are brought in for evaluation and reflection during After Action Review (AAR). As theory is immediately connected to images, it supports an optimal transfer to practice. The method can be applied both for teams working together on one location as well as distributed teams, allowing real-time or delayed monitoring, and supporting distance learning approaches.

The core of the method is assessment of whole task performance (according to recommendations of the 4C/ID model). In this case, the learning task is a complex F16 maintenance activity that integrates all routine and non-routine elements of the real task. It involves working accurately according TOs, in small teams, and in work circumstances such as noise, technical limitations, and safety hazards. VAMM applies a model of shared ownership of the learning task. This model refers to the natural tendency of teams with shared task experiences to commit to tasks and claim responsibility for learning processes (Broers, A.J.M., 2007). Each member has been assigned to a role with specific responsibilities mirroring responsibilities in the real work context; either as dock-chief, task performer, or evaluator. First, the dock-chief manages the team, is responsible for work delegation amongst team members, and is first point of contact for other team members during task execution. Second, the task performer is responsible for the practical performance of the task. And third, the evaluator is responsible for peer-to-peer assessment of task performance and provides cognitive feedback (supportive information) during the After Action review (AAR).

To support the process of peer-to-peer assessment, VAMM uses a pre-structured evaluation panel with domain specific evaluation criteria for technical maintenance tasks.

Figure 1. The tagging panel



The panel is used by the evaluator to assess peer-learners. As the buttons on the panel provide a complete overview of all criteria that are relevant for strong task performance, the panel should support teams in building a frame of reference that will help them to assess their performance collectively. According to Cannon-Bowers, Salas, and Converse (1993), *shared mental models* help explain how teams are able to cope with difficult and changing task conditions. When team members share accurate mental models of the teamwork processes that influence their performance, they should be better able to (a) uncover performance trends and diagnose deficiencies, (b) focus their practice appropriately on specific goals, and (c) generalize the lessons they learn to new situations (Cannon-Bowers & Sales, 1998; Smith-Jentsch et al (pp 273)).

In our case study, VAMM is used to focus on the acquisition of attitude in the setting of learning complex cognitive skills. The goal of this case study is to examine if the VAMM concept works; (a) 'are the learners capable of assessing mutual task performance by means of a computer-based tagging panel?'; and (b) 'do they manage to mark sufficient and the most relevant episodes for assessment of core maintenance competences?'.

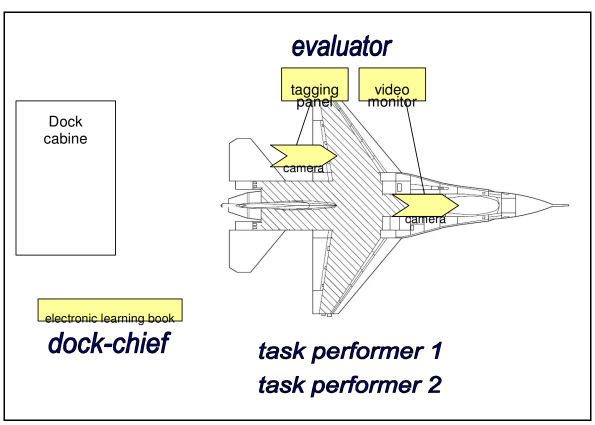
METHOD

Eight learners and three instructors participated in the case study. The learners were 16-18 years old trainees from civil schools for aircraft maintenance who join a training period at the RNLAF. These schools are responsible for the initial aircraft maintenance training program which is the first part of the training of maintenance engineers at the RNLAF. The main objective of this part of the training is the acquisition of a strong attitude towards a safe and conscientious way of working. Although all learners had the same background and prerequisite knowledge, four out of eight learners had slightly more practical experience in aircraft maintenance. The instructors, all experienced maintainers themselves, all had a long history in teaching F16 maintenance.

The learners were split up in two small teams with equal experience levels amongst team members, team B having slightly more practical experience than team B. This team classification has been chosen to support a positive atmosphere within the teams and prevent negative effects of competition. Each learner within the team was designated to a specific role; two task performers, one evaluator, and one dock-chief.
Both teams performed two tasks. Task 1 was to perform a 'Safe for maintenance procedure' and task 2 was to perform an 'operational check-out on the Arresting Hook of the F16'. The learning tasks were designed according to 4C/ID-based whole-task approach. This means that the task circumstances represented the social and physical task context of F16 maintenance, including all relevant aspects of the task as it occurs in practice.

The learning environment was comprised of a real F16 aircraft, TOs, and maintenance tools and equipment. The learning tasks were described in an electronic learning book which provided access to background information for all steps involved (e.g., procedure descriptions or video-episodes of expert task performance). Two video-cameras were set up to monitor each team of learners. One camera registered activities executed in the cockpit and the other camera was set up at the back of the airplane to register activities at the arresting hook.

Figure 2. Experimental set-up



We used a commercial computer-based tagging tool (Darthfish; see www.darthfish.com) to mark and classify incoming video episodes according to predefined evaluation criteria for assessing core maintenance competences. The classification was done by selecting a limited set of options from a tagging panel.

In order to design a tagging panel that matches the needs and task context of the target users, the panel was designed in close cooperation with Subject Matter Experts (SMEs) and instructors of the Air Force. To support acquisition of integrated maintenance competences on basis of concrete observable criteria, the evaluation criteria for aircraft maintenance have been listed and classified according to knowledge, skills and attitudes involved. The practical criteria (skills and attitudes) are the main elements of the evaluation model, as they are concrete and observable during task performance. The theoretical criteria (knowledge) have been added to support assessment of required knowledge as part of the integrated skill. As availability of knowledge is hard to observe during task performance, assessment of knowledge is based on learner's responses on specific task-related questions during task preparation. The fifteen resulting criteria are implemented in the Dartfish software. Table 1 below provides an overview of the criteria.

Table 1. The evaluation criteria

Knowledge	Aircraft systems TO's Tools Ground equipment
Skills	TO selection Tools selection TO use Tools use Ground equipment use Torque methods and safety wiring
Attitudes	TO's discipline Tool discipline Safety rules Orderly way of working Communication and coo-operation

The tasks had to be executed as in operational circumstances. The learners in the role of task performer were responsible for task execution. Each session started with a task preparation in which the learners with the role of dock-chief and evaluator prepared themselves to the task procedure by means of the technical manual plus electronic learning book. The dock-chiefs had the final responsibility for the task. They were asked to make an action plan and bring this over to the team. Additionally, they were told to check whether the minimal required knowledge for safe task performance was available by questioning the learners. The learners with the role of evaluator were responsible for assessment of correct and incorrect task execution for the specific categories involved. They received a thirty minute instruction on how to use the tagging panel. In order to trigger them to mark as much episodes as possible (showing examples of good and bad displays of attitudes and skills), they were asked to bring in a selection of marked video-episodes during AAR.

Data were collected by logging, observation techniques and interviewing. The logging data were quantified on total number of tags that were made, showing correct or incorrect task performance in categories involved. To determine the core evaluation criteria for the training program, the instructors were asked to seek consensus and select the five most important criteria. The quality of tags was measured by mapping the tagging results to these main evaluation criteria. During the study, two experiment leaders were available to operate the camera's and observe task performance. Afterwards all learners filled in a form with four statements, expressing on a 4-point Likert scale how they perceived working according to the new method. There were two types of forms; (A) a form for dock-chiefs and task performers and (B) an evaluator form. Main topics in both forms were related to the function of VAMM (does it support assessment of maintenance competences) and pleasure of use. Additionally, form B contained statements to rate the usability of the tagging panel. Finally, the learners were asked to write an evaluation report to self-reflect on positive and negative (learning) experiences.

RESULTS

Analyses of the logging data shows that a large number of video-episodes was tagged for AAR (see table 2).

Table 2. Number of tags

	Team A	team B	All
Correct	52 (94,5%)	93 (94,9%)	145 (94,8%)
Incorrect	3 (5,5%)	5 (5,1%)	8 (5,2%)
Total	**55 (100%)**	**98 (100%)**	**153 (100%)**

Team A, the team with slightly less practical experience, scored significantly less tags then team B did. Table 2 shows that the ratio between correct and incorrect tags is equally deviated for both teams.

The three instructors selected five (out of the fifteen) evaluation criteria to be the main evaluation criteria for the initial maintenance training program. The results are displayed in table 3.

Table 3. Main evaluation criteria for initial training

TO use (skill)
Communication and co-operation (attitude)
Tool discipline (attitude)
Safety rules (attitude)
TO discipline (attitude)

Logging data of the tagged video-episodes provide insight in the number of times specific events have been tagged. Table 4 provides a list of the top 5 most tagged events.

Table 4. Top 5 most tagged events

Evaluation criterion	*number of tags*
1. TO use (skill)	39
2. Communication/co-operation (attitude)	33
3. Tool use (skill)	23
4. Safety rules (attitude)	17
5. TO discipline (attitude)	12

The mapping of results from table 3 and 4 resulted in a close match between the most important evaluation criteria and the criteria that have been tagged most. Four out of five criteria are covered by both lists. With regard to the fifth criterion the instructors rated *attitude* with regard to tool discipline to be one of the most important evaluation criteria. The learners on the other hand, reflected mostly on the *skill* of using tools. Eighty-one percent of the tags covered a criterion out of the top 5. Only two criteria were not used at all. These were criteria were 'tool selection skills' and 'torque and safety wiring skills' (the latter was not applicable to the experimental task).

Table 2 shows that ninety-five percent of the video-episodes that were tagged by the peer-evaluators were examples of correct task performance. Although the teams did perform well and hardly any faults were made, instructors and team members were able to trace additional deficiencies during AAR.

Observation of AAR showed that the confrontation with images of own task behavior supported the start of a dialog and detailed discussions about task performance. In some situations the participants were able to uncover performance trends, like careless behavior (which often are hard to assess because of being covert). An example during AAR was the evaluation of a part of the procedure where the team had to bring the Arresting Hook down. The learner in the role of evaluator tagged the video-episode because of communication problems. One of the learners sat in the cock-pit while the others stood at the back of the plane by the Arresting Hook. In these circumstances, the person in the cockpit is not able to see his team-mates at the back of the plane and must make sure that no people are hurt when the hook comes down. Therefore, one of the safety rules is that the person in the pilot seat holds his hands visible outside of the cock-pit so that other team members can be sure that he does not touch any controls at crucial moments. While viewing the video-episode during AAR it was noticed that this safety condition had not been taken correctly. In the discussion that followed the video-episodes were used to show that the learner did intend to perform the procedure, but his timing was incorrect. Another example of unsafe behavior that was uncovered during AAR was the tagged video-episode of a learner who uses a pencil to point at a specific control panel in the cock-pit. To prevent for Foreign Object Damage (FOD) it is strictly forbidden to use any other objects in the cock-pit than the required maintenance tools. From the observations during AAR, the confrontation of learners with images of this kind of covert but risk full behavior appeared to make a large impression.

During task performance, observations were made of learner's enthusiasm and motivation in team performance. Learners were strongly committed to the task and claimed an increasing responsibility for good task performance. These observations were confirmed by the learners' responses in the evaluation reports. The reports describe the pleasure of use that the learners experienced working with VAMM. They emphasized enjoying performance of realistic learning tasks, involving many aspects of their future function. In addition, the results on the statements (table 4 and 5) show that learners were very positive with regard to the function of VAMM (supporting peer-assessment) and no large usability problems were found.

Table 4. Subjective opinions task performers and dock-chiefs (n=6)

	m	*sd*	*range*
1. Accepted feedback from peers	3.50	0.58	1
2. Video-tags made feedback understandable	3.50	0.58	1
3. Agreed on results of peer-assessment	3.25	0.96	1
4. Enjoyed working with VAMM	3.25	0.50	0

*1= totally disagree, 4 = totally agree

Table 5. Subjective opinions evaluator (n=2)

	m	*sd*	*range*
5. Video-tags supported assessment	3.50	0.71	1
6. Enjoyed working with VAMM	3.00	0.00	0
7. Tagging panel was easy to use	3.00	0.00	0
8. Understood evaluation criteria	3.00	0.00	0

*1= totally disagree, 4 = totally agree

CONCLUSION

VAMM is designed to evaluate acquisition of attitudes in the context of learning complex cognitive tasks in the initial maintenance training program. In this context, assessment of these competences is very crucial because seemingly innocent deficiencies in task performance may result in serious safety hazards.

VAMM appeared to be a excellent method for evaluating the performance of particular aspects of complex cognitive maintenance skills. First, application of the method in this papers' case study proved that learners in the role of evaluator were able to tag a sufficient number of events for proper AAR. Second, the events that were tagged appeared to be the most important attitudes and skills to be acquired during the training.
It is now demonstrated that the method uncovers covert task behavior. Moreover, it identifies performance deficiencies that are usually hard to measure, because they are an integral part of a complex cognitive team skill.

Often, when learning complex cognitive maintenance skills, maintenance engineers tend to focus on (technical) details and lose the broader overview. The results of the case study show that VAMM supports the target learners to focus on critical skills and attitudes and pay less attention to the evaluation of detailed technical skills such as 'tool selection'. VAMM seems to support the learners in taking a different perspective. First, the method forces them to act in different roles and take different perspectives. Second, the availability of a shared frame of reference appears to support them in focusing on a wider range of goals.

VAMM only provides evaluation support to the learner with the role of evaluator. Consequently, much of the learning of other team members takes place during AAR. In future research it would be interesting to search for other ways to make evaluation support available for all team members, already during task performance. An example would be the integration of reflection activities in learning tasks, supporting teams to improve or recover from faults.

The new way of assessment that VAMM applies appears to stimulate learners to become active learners and take responsibility for own learning processes. When acting in teams the learners tend to take increasing responsibility for tasks. The study results confirm that learners are supported by the application of streaming video as part of the communication process. During AAR, the objective character of the visuals stimulates the (rather inexperienced) learners in starting the dialog and supports interpretation of comments of the instructor and peer learners.

A limitation of VAMM is the focus on positive feedback. During the case study learners tended to provide mainly positive feedback. One explanation would be that the evaluators did not recognize the deficiencies (as they were no expert task performers). Another explanation would be that they hesitated to be critical towards team members and were prone to decision biases. A third explanation would be that the evaluators did not have enough experience in using the evaluation panel. Currently, an experimental design study is set up to test the impact of the method more rigorously. The impact of the third explanation will be tested in this study by adding an extra practice trail. In this trail all team members will have to score a large number of example video-episodes, increasing inter-evaluator reliability and decreasing the threshold to tag negative examples.

VAMM can be used to improve training effectiveness as well as training efficiency. In the current study, VAMM has been tested within the initial training program for maintenance engineers. An additional study will be required to test the impact of the method in other parts of the maintenance training program.
The instructor can apply VAMM to provide an objective and transparent manner of assessing realistic tasks and to stimulate learners to become active learners. However, the method may also be applicable in examination settings, for supporting distributed teams, allowing retrospective monitoring, and supporting learning on distance. Instructors, who often experience high work loads, are then able to safe time for other purposes such as coaching and assessment.

Given these results, it can be concluded that VAMM has great potential in the maintenance training domain.

ACKNOWLEDGEMENTS

The study described in this paper has been supported by the following people who were part of the project team: Jos van der Arend (project leader), Toon van der Steen (instructor at the RNLAF), Martin Bouwer (instructor at the RNLAF), Willem Hylkema (policy maker at the RNLAF), and Nicolet Theunissen (group leader of research area learning innovations).

REFERENCES

Broers, A.J.M., (2007). *Workshop learning at the RNLAF; from civil learner to RNLAF CAT-A maintenance engineer [Klu Opleidingsdock 2008; van ROC leerling tot Klu CAT-A monteur].* Royal Netherlands Air Force, Technical Report. Volkel, The Netherlands.

Brown, J. S., Collins, A., & Duguid, S. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Cannon-Bowers, J.A., Salas, E., & Converse, S.A. (1993). Shared mental models in expert team decision making. In N.J. Castellan, Jr. (Ed.), *Current issues in individual and group decision making* (pp. 221-246). Hillsdale, NJ: Erlbaum.

Goldschmeding, J. T.; Leijen-de-Vos, M.; Nijlunsing, W.; Oomen, J. (2006). Streaming video: huge impact (*Webstroom: genzeloze impact).* SURF Onderwijsdagen 14 en 15 november 2006.

Lave, J., & Wenger, E. (1991). Situated learning: legitimate peripheral participation. Cambridge: Cambridge University Press.

Smith-Jentsch, K.A.; Johnston, J.H.; Payne, S.C. (1998). Measuring team-related expertise in complex environments, In Cannon-Bowers, J.A. (Ed); Salas, E. (Ed), *Making decisions under stress: Implications for individual and team training*. (pp. 61-87). Washington, DC, US: American Psychological Association.

Smith-Jentsch, K.A.; Zeisig, R.L.; Acton, B.; McPherson, J.A. (1998). Team dimensional training: A strategy for guided team self-correction, In Cannon-Bowers, J.A. (Ed); Salas, E. (Ed), *Making decisions under stress: Implications for individual and team training*. (pp. 271-279). Washington, DC, US: American Psychological Association.

Van Merriënboer, J. J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning*. Mahwah, NJ: Lawrence Erlbaum.

Vygotsky, L. S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.