Multimodal Learning Experience for Deliberate Practice

Di Mitri, Daniele; Schneider, Jan; Limbu, Bibeg; Mat Sanusi, Khaleel Asyraaf; Klemke, Roland

DOI
10.1007/978-3-031-08076-0_8

Publication date
2022

Document Version
Final published version

Published in
The Multimodal Learning Analytics Handbook

Multimodal Learning Experience for Deliberate Practice

Daniele Di Mitri, Jan Schneider, Bibeg Limbu, Khaleel Asyraaf Mat Sanusi, and Roland Klemke

Abstract While digital education technologies have improved to make educational resources more available, the modes of interaction they implement remain largely unnatural for the learner. Modern sensor-enabled computer systems allow extending human-computer interfaces for multimodal communication. Advances in Artificial Intelligence allow interpreting the data collected from multimodal and multi-sensor devices. These insights can be used to support deliberate practice with personalised feedback and adaptation through Multimodal Learning Experiences (MLX). This chapter elaborates on the approaches, architectures, and methodologies in five different use cases that use multimodal learning analytics applications for deliberate practice.

Keywords Deliberate practice · Psychomotor learning · Sensor devices · Multimodal interfaces · Intelligent tutoring systems

1 Introduction

In times of physical distancing and learning in isolation, digital learning is commonly associated with video-conferencing tools or online learning platforms such as Learning Management Systems (LMS). Defining digital learning to these well-

D. Di Mitri (✉) · J. Schneider
DIPF | Leibniz Institute for Research and Information in Education, Frankfurt, Germany
e-mail: dimitri@difp.de

B. Limbu
Leiden Delft Erasmus-Center for Education & Learning, Technical University Delft, Delft, The Netherlands

K. A. Mat Sanusi
Cologne Game Lab, TH Köln, Cologne, Germany

R. Klemke
Cologne Game Lab, TH Köln, Cologne, Germany

Open University of the Netherlands, Heerlen, The Netherlands

M. Giannakos et al. (eds.), *The Multimodal Learning Analytics Handbook*,
https://doi.org/10.1007/978-3-031-08076-0_8

known digital tools is a limitation, both for what modern digital technologies can offer and what someone can learn. With a desktop computer, one can sit down, watch a video, read some material, but there is no guarantee that it will translate to real-world performance. For example, it is not recommended to jump into the ocean just after reading the manual on how to swim. The discrepancy between learning time and learning gain can be attributed to the lack of actual authentic practice. The authentic practice is performed within the context/physical world where often more than one modality is involved. Learning in the physical world is multimodal, which may explain the limitations of traditional desktop-based digital learning, which only utilises visual and auditory modalities (Gruber et al., 1995). The current landscape of educational technologies is also constrained by the modalities of interaction with computers and smartphones.

The majority of current digital education still relies on traditional desktop experiences, which are still the most common computational interfaces. However, the increased availability of sensors as consumable technologies has paved the way for the potential use of multimodal technologies for learning (Schneider et al., 2015a; Di Mitri et al., 2018). By translating measurements of the physical world into digitally readable formats, sensors enable awareness of the physical environment in computing units which facilitate digital instructions in authentic practice. For example, sensors such as microphones that capture audio signals can be used to naturally train certain voice aspects like the voice volume for public speaking or voice pitch for singing. Furthermore, processing units have gotten smaller and faster, reducing the overall size of computing devices, making wearable technologies more comfortable and unobtrusive to use, and allowing for near to real-time calculations on sensor data input. Works on sensor fusion further facilitate a complex eco-system of sensors that can monitor various attributes of the physical environment and interactions between them. For example, Augmented Reality (AR) utilises multiple sensors such as cameras, accelerometers, and gyroscopes to provide users with affordances for natural multimodal interactions and wearable displays via which digital instructions can be provided during authentic practice (Guest et al., 2017). Such technologies offer unique learning opportunities via multimodal learning experiences (MLX). In this book chapter, we elaborate on the approaches, architectures, and methodologies derived from five different case studies of technologies providing MLX for deliberate practice with the purpose to present a novel perspective and the potential impact of educational technologies with new multimodal interfaces.

We define MLX as any learning activity using more than two modalities within an authentic learning setting. The use of multimodality is founded on three basic principles: (1) sensors, (2) authentic practice, and (3) immersive and ubiquitous technologies.

Sensors Data from a single sensor or modality are usually insufficient to reliably capture and explain the complex interactions between the learner and the environment that results in learning (Di Mitri et al., 2018). For example, by tracking the heartbeat of a learner, it might be difficult to identify if the learner is sleeping or fully

concentrated watching an educational video. An additional eye-tracking sensor, in this case, can be helpful to support the interpretation of the heartbeat. Learning in an authentic setting (e.g. an industry workplace) naturally comprises interaction using multiple modalities. MLX makes use of multiple modalities and hence, associated sensors to support digital learning in an authentic setting.

Authentic Practice Desktop-based interactions have constrained support provided by educational technologies to the cognitive domain, limiting practice to traditional naive drill-based learning. This limitation often results in learners using educational technology failing to transfer learning and demonstrate expert performance in the real world (Clark & Voogel, 1985). Watching video instructions on how to ride a bicycle is not likely to automatically translate to learning to ride a bicycle. To learn how to bike, the student must use a bicycle, and in doing so, they are practicing in an authentic context. Multimodal learning technologies merge learning and authentic practice, which increases the likelihood of better performance in a real-world setting. Furthermore, multimodal technologies also show affordances that support deliberate practice in authentic settings, fostering efficient and effective learning of skills.

Immersive and Ubiquitous Technologies Beyond only supporting learning in an authentic setting, multimodal technologies can push the learning experience further by creating immersive and ubiquitous learning environments. Multimodal technologies such as AR create immersive environments by augmenting digital content to the physical world, enabling immersive learning at any time and place. These technologies make it possible to simulate environments like the audience of a presentation while practicing to speak in public. Moreover, such technologies can present feedback and instruction unobtrusively. For example, in the aircraft maintenance training use case, the necessary instructions can be shown in the direct path of vision of the learner.

Multimodal technologies enable educational technologies to be designed around the learning activities instead of the device. They promote ubiquitous and constructivist learning while facilitating immersive, authentic practice. This chapter will use the term "*MLX systems*" to indicate the various multimodal sensor-based immersive and ubiquitous technologies that support learning. By practicing in an authentic setting, learners can reduce the gap between knowledge acquisition and application. However, it does not account for the attainment of expertise. Consistent progress towards expertise requires *deliberate practice,* a particular type of purposeful and systematic practice that needs to focus attention and repetitive executions to improve a particular skill (Ericsson et al., 1993), which will be further discussed in the following section.

2 Multimodal Learning Theories

2.1 Embodied Learning

Multimodal learning, simply put, is learning with multiple modalities/senses. In a standard classroom, we often use two modalities to learn, namely visual and auditory. In line with this, multimodal learning proposes using multiple senses, often more than two, for effective learning and skill acquisition. For example, Juntunen (2020) and Odena (2012) argue that multimodal learning helps learners retain knowledge for a longer time. Juntunen (2020) also argues that multimodal learning is inseparable from embodied learning and is a way to promote embodied learning. Embodied learning views learning and skill acquisition as grounded in the body and the environment it is operating. Since senses, which are functions of the human body, are the only way humans can perceive the environment, it can be argued that senses play a vital role in learning. Multimodal learning aims to leverage this and design learning environments that use multiple modalities/senses and foster embodied learning.

Embodied learning involves perception via senses, motor activity, and introspection that helps to assimilate better knowledge (Robbins & Aydede, 2009; Clark et al., 2019). For example, children learning language by writing have shown better recall and recognition of the characters and more brain activation (Longcamp et al., 2003). This claim is homogeneous to Clark and Paivio (1991)'s assumption in Dual coding theory, which states that humans process information via visual and auditory means without a cognitive load overhead. For example, Chandler and Tricot (2015) note that body movement, such as gestures, can help offload working memory, which in turn allows working memory resources to be used in creating a deeper understanding. Besides, Clark et al. (2019) and Kiefer et al. (2015) state that embodied learning also leads to better recall rates of training. Embodied learning and, therefore, Multimodal Learning offers various benefits for learning. Additionally, in the following sections, we would like to argue that multimodal learning and embodied learning-based learning environments have the potential to foster deliberate practice.

2.2 Deliberate Practice

Learning occurs via the repeated perception-action cycles, which is central to the ideas of Embodied learning. In addition to sheer repetition, Ericsson et al. (1993) claim that conscious and structured practice is required for sustained improvement. Goldman Schuyler (2010) also states that embodied learning does not occur automatically via sheer repetition and requires awareness. In both cases, the authors claim that a conscious form of practice is required, also known as deliberate practice. This practice, however, requires a high cognitive load. Learning in an embodied

manner distributes the cognitive load among the body and environment, i.e. the senses, as also claimed by the Multimedia theory of cognitive load (Hutchins 1995). Therefore, integrating embodied learning into deliberate practice can help better manage cognitive load. Hence, we view deliberate practice and embodied learning as complementary concepts.

Ericsson et al. (1993) define deliberate practice as a conscious and highly structured activity, the explicit goal of which is to improve performance. However, it is difficult for novice learners to practice deliberately, as it is cognitively demanding to be conscious of their performance (Rikers et al., 2004; Ericsson et al., 2007). Goldman Schuyler (2010) also claims that developing the capacity to act with awareness – to be fully present to what is taking place – is fundamental to embodied learning as well. Nonetheless, embodied learning and multimodal learning claim to distribute the cognitive load among the body/its senses and modalities, increasing the likelihood for the learner to practice deliberately. A key assumption behind the distribution of cognitive load to the environment is that the learner must practice within the context, i.e. learners will learn a skill better by actually practicing it in the authentic context. Neelen and Kirschner (2016) also claim the same by arguing that deliberate practice must be done in an authentic setting. This requirement for an authentic setting can also better explain/foster transfer of learning observed in embodied learning.

Practicing deliberately is a complicated endeavour that entails many more variables in practice and their interrelationship, which must be explored further. Regardless, embodied learning methods are apt to support deliberate practice, especially in novices, as they provide strategies for better cognitive load management. Multimodal learning technologies such as sensors, augmented reality, etc., can facilitate learning environments that foster embodied learning and, eventually, deliberate practice. Hence, multimodal learning technologies show potential for training psychomotor skills. Therefore, in this chapter, we aim to present our experience working with it in the context of psychomotor skills development.

3 Engineering Aspect

MLX systems need to be able to track the learners' performance while deliberately practicing their tasks. Ideally, as unobtrusively as possible, learners can focus on the practice rather than on the interaction with the technology. Sensors are technologies that enable the unobtrusive tracking of learners while engaging in deliberate practice. Sensor data, however, is noisy and generally has poor semantic value; therefore, in many cases, multiple data sources are needed to do a reliable analysis of the learners' performance. Multimodal data introduces multidimensional complexity. Tracing or analysing multiple modes of interaction implies that the efforts for data gathering and inference making must be multiplied. As compared to log data, sensor-based signals are highly dimensional and have low semantic values (Dillenbourg, 2016). Learning analytics, by their definition, aims to ultimately

support learners in the environments where the learning occurs (Greller & Drachsler, 2012). The use of MMLA for deliberate practice can be pictured as a *Multimodal Feedback Loop*, i.e. a data-informed cycle in which the data gathered from the learner undergoes a series of steps to ultimately come back to the learner in the form of feedback and actionable information for the user. Despite, such steps vary depending on the specifics of the learning domain, can group they can be grouped in the following "Big Five" challenges for MMLA: (1) data collection, (2) data storing, (3) data annotation, (4) data processing, and (5) exploitation (Di Mitri et al., 2019b). Engineering MMLA systems that tackle each of these challenges requires the use of complex and interconnected systems resulting often in a time-consuming development task.

3.1 Architecture Overview

To lower the entry point for MMLA and support current and future researchers in leveraging the potentials of multimodal data for learning, we introduced *the Multimodal Pipeline*, a generic technological framework that can serve as an architectural blueprint for MMLA/MLX projects (Di Mitri et al., 2019b). The Multimodal Pipeline introduces a set of generic solutions to the above-mentioned "Big Five" challenges introduced by multimodal data, namely the collection, storing, annotation, processing, and exploitation of multimodal data. The Multimodal Pipeline is a process model, which is directed primarily at research purposes, i.e. the data of the learners are collected to research and unfold particular aspects of learning. In the future, we can imagine that systems following the blueprint of the Multimodal Pipeline can reach and enter a production stage in which the data from learners are systematically collected in their learning contexts.

In the following sections, we describe the components of the Multimodal Pipeline by analysing four functional layers as represented in Fig. 1: (1) the interaction layer, (2) the data collection layer, (3) the feedback layer, and (4) the data layer.

3.2 Interaction Layer

The interaction layer refers to the user interface (UI) indicating the set of sensors and actuators, including the applications with which the target user (most often the learner and the tutor) interacts. In multimodal interaction, the UI usually comprises – but not exclusively – non-graphic interface elements, meaning that in addition to displays, various wearable sensors, cameras, or internet of things devices are involved. In some examples, the multimodal interfaces are accompanied by some

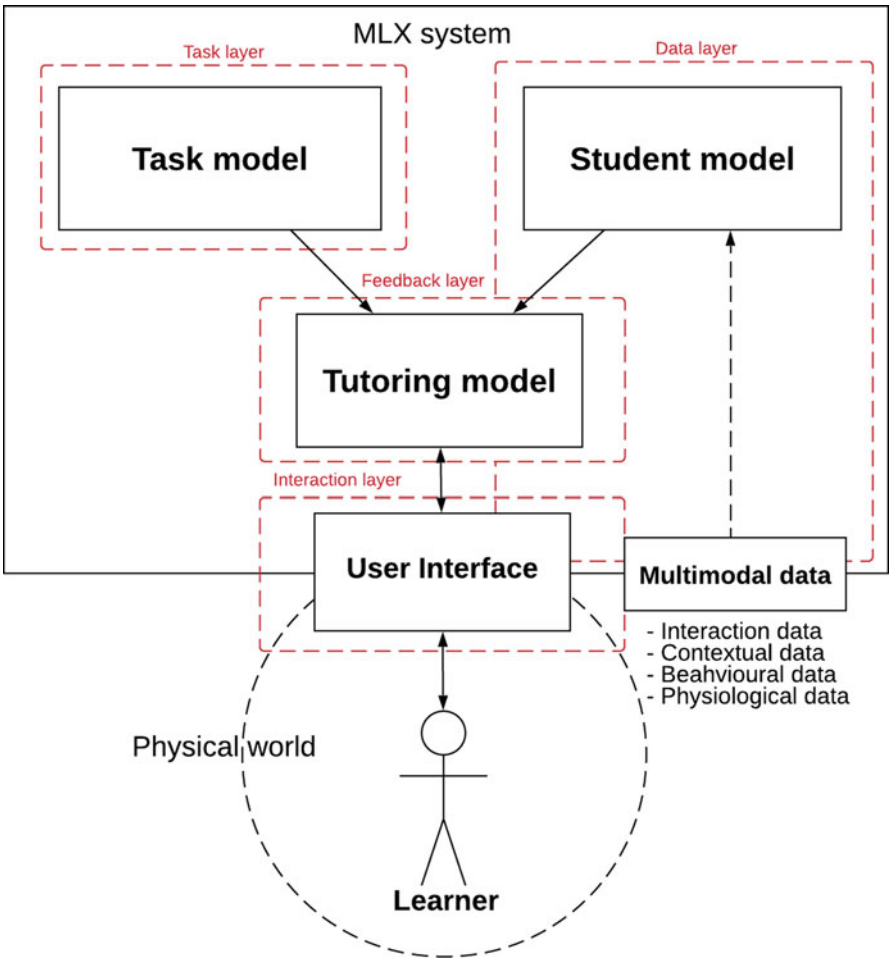

Fig. 1 Structure of the MLX system

types of displays for visual feedback to the learner. One example is the Calligraphy tutor (Limbu 2020), which uses the capacitive display of a tablet in conjunction with the digital pen and the myographic device to train character writing skills in foreign alphabets. Another example is the Presentation Trainer, which uses a large display connected to a depth camera that provides real-time feedback on the speaker's body position and presentation style (Schneider et al., 2015). Both examples are further described in Sect. 5.

3.3 Data Layer

The data layer deals with data collection, storing, annotation, and processing. The first two challenges are addressed by the *Multimodal Learning Hub* (LearningHub), a research prototype that allows for a customised collection of learning experiences (Schneider et al., 2018). The LearningHub enables the researcher to collect data from various sensor applications and to obtain synchronised session files into a custom file format. In its current version, the LearningHub supports the recording and storage of data from a limited number of commercial sensor devices.

Multimodal data are noisy and difficult to interpret. The *Visual Inspection Tool* (VIT) (Di Mitri et al., 2019b) allows visualising and annotating the data collected by the LearningHub. This tool helps experts to design interventions. It is the basis for further developments towards general learner data analysis support.

3.4 Feedback Layer

The feedback layer implements the Tutoring model, which characterises the approach for providing feedback. The feedback of MLX systems is distinguishable between (near) real-time and retrospective feedback. The real-time approach is typically used for steering the learner towards the ideal learning trajectory. This feedback can consist of correction, nudges, or other forms of automatic intervention which is popular among the Intelligent Tutoring System community. The retrospective feedback instead includes either the novice or the learner in the assessment process. Examples of retrospective feedback are learning analytics dashboards which can brief the execution of the task for stimulating self-reflection.

3.5 Task Layer

The task layer represents the learning and training tasks a learner needs to go through. Learning tasks can be highly diverse. Consequently, there is not a *one-size-fits-all* solution for all use cases where MLX systems can be employed. Every experimental scenario has specific requirements and constraints that are unique from other scenarios. Before deciding to utilise a particular system for a learning session, the engineering part should collect system requirements and constraints. We talk in this case about the 'task model', which means considering all aspects related to the learning task, including how long it lasts, what age are the learners, what meaningful actions are they supposed to take during the task, which modalities are best to track during this task. Yet, there is limited understanding of which sensors and modality are best matching the sensors. The literature review from Schneider et al. (2015a) provides the first analysis to best match the sensor devices to types of psychomotor

learning tasks. Yet, at this stage, the matching between learning activities, sensor devices, and modalities of interaction is best found as a result of compromise of what has been proven to work and what is technologically feasible.

3.6 The MLX System into Teaching

With the multimodal architecture presented in Fig. 1 and described in the previous sections, we have detailed a general MLX architecture. It is relevant to describe how this architecture is embedded into the teaching and learning process of psychomotor skills training. The use of the MLX system for psychomotor skills training promotes Deliberate Practice which is the sequence of learning sessions with graduating criteria that the learner takes autonomously that feature both real-time feedback and retrospective feedback at the end of the session. They also provide the mentor with high fidelity representation of the student's practice session, which allows them to provide more effective feedback asynchronously (e.g. hours or days after the learner has practiced with the MLX system). The AI-generated feedback of the MLX system does not replace human feedback but can complement it.

4 Research Methodologies

The mere act of developing an MLX system for learning is no guarantee that it will support learners in improving their skills. It might as well cause adverse effects by distracting, confusing, or forcing the learner to perform cumbersome interactions with the MLX system. Therefore, it is important to follow a methodological approach for developing and evaluating MLX systems for deliberate practice.

MLX for deliberate practice exploits the use of emerging sensing and display technologies to support the deliberate practice of specific skills. This exploitation of technologies requires the design, development, and evaluation of educational prototypes. Therefore, the overall methodology commonly conducted in this field is based on Design-Based Research (DBR), using the proposed engineering aspect of MLX systems as a blueprint to develop technological solutions. DBR is an iterative research methodology, which consists of designing, developing, and evaluating prototypical solutions (Anderson & Shattuck, 2012). This methodology is commonly used in the learning sciences to create and refine educational interventions. High-quality design-based research studies commonly have characteristics such as the involvement of multiple iterations, being situated in a real educational context, close collaboration between researchers and practitioners, focusing on the design and evaluation of the significant intervention, use of mixed methods, etc.

Currently, MLX for deliberate practice is an emerging field of study. At the moment no established best practices indicate the number of iterations needed nor well-defined procedures for each of the iterations. Based on our studies on MLX for

deliberate practice, the following subsections present basic descriptions and pointers about important considerations for the first few iterations when researching MLX for deliberate practice.

4.1 First Iteration

In the case of the research of multimodal technologies for deliberate practice, the first iteration following DBR consists of the design, development, and evaluation of a proof of concept. Before starting with the first design phase, it is important to do a requirement analysis where one selects a skill to be developed through deliberate practice supported by multimodal experiences. In other words, the development of which skill do we want to support?

Once the skill is selected, it is important to do a part of the task analysis because skills usually consist of sub-skills. We define sub-skills as unit aspects of the complex task that cannot be further broken down and can be practised by itself (van Merriënboer et al., 2002). For this step, first, one needs to identify the attributes and performance indicators that the prototype will support. For example in the scenario of practicing public speaking skills, voice volume, body posture, etc. can be the selected attributes or performance indicators. After the requirement acquisition, one needs to design the intervention that the tutor will provide to learners based on the selected performance indicators. In the example of practicing public speaking, the intervention could be the display of a message indicating the learner to speak softer when the learner is speaking above a predefined threshold. To guide this fourth step, models like Instructional design for Augmented Reality (ID4AR) from Limbu et al. (2018) can be used. ID4AR, which is based on the 4C/ID model (van Merriënboer et al., 2002), supports the holistic design of multimodal learning environments that support deliberate practice, with individual interventions for each attribute or performance indicator.

After the task analysis is conducted, it is important to design the setup configuration for the MLX system for deliberate practice. This step requires the identification and selection of hardware devices required for the tutor. There are three distinctive hardware devices needed for the development of such a tutor. Sensors to track the learners' performance, a processor to analyse the captured performance, and output devices that present the results of the analysis back to the learner/teacher. The selected hardware highly depends on the skill, for example, if the skill is something like running, cycling, etc., to train the skill, the learner needs to move in space. In such a case, wearable hardware is more suitable than static. While in cases where the learner remains stationary such as practicing Cardiopulmonary Resuscitation(CPR), ambient sensors and displays can be a better option.

The development of the proof of concept MLX system for deliberate practice comes after its design phase. This phase is usually more problematic than expected,

especially for this first iteration. Rather than reinventing the wheel, the MLX system (See Sect. 3) presents a model that can serve as a blueprint to develop MLX system prototypes.

The evaluation of the MLX system proof of concept is the last phase of the first research iteration. This last phase comprises two main steps: conducting user tests and evaluating the results of the tests. Conducting user tests is an activity that needs to be carefully planned. Before starting the tests, one needs to define the goals of the evaluation, target users of the tutor, and the evaluation procedure.

The goals of the evaluation must align with the insights that user tests can provide. For this first iteration, user tests can reveal important aspects regarding the interaction with the tutor and their usability issues. They also provide crucial information concerning the appropriate setup of the tutor for its effective use and evaluation.

In this first iteration, explorative instruments such as questionnaires and surveys are helpful to provide a general idea of how the interaction with the multimodal tutor is perceived by the users. Questionnaires regarding user experience, usability, and technology acceptance are useful for this goal. Nonetheless, at this point of the research, an analysis of the interactions between users and the tested multimodal tutor and qualitative data provided by the users that tested the system can bring better insights into the necessary improvements for future versions of the tutor.

Another possible goal of the user tests is to investigate whether the data collected by the sensor setup is good enough to gain insights into the learners' performance. Artificial intelligence, more specifically machine learning approaches, can be very helpful for this goal. To implement these machine learning approaches, it is important to first annotate the collected sensor data. The sensor data is in many cases not possible to be directly interpreted by humans, therefore, to perform the annotation task, video recordings of the user tests are needed. By looking at the videos, human experts can analyse the learners' performance and annotate the sensor data accordingly. For example, in the study of Di Mitri et al. (2019a), a proof of concept of a multimodal tutor for the deliberate practice of CPR was developed. User tests were conducted to investigate if the data collected by the proposed sensor setup allowed the identification of a good CPR technique. In this study, human experts examined the video recordings of the user tests, identified correct and incorrect techniques, and correspondingly annotated the sensor data with the help of the VIT (Di Mitri et al., 2019b).

4.2 Second Iteration

The first iteration can provide sufficient insights to move an MLX system prototype beyond the proof of concept. The design phase of this second iteration includes a prototype design and a user interaction design. The first step in the design of the prototype consists of designing solutions to address the challenges pointed out by the results of the first iteration. In this second iteration, the prototype design might

include requirement acquisition studies with experts in the field to identify important aspects that the multimodal tutor should support based on what is technologically feasible to implement. Whether experts should be included for the requirement acquisition in the first iteration or in this second one is a chicken and the egg problem. Some arguments for including experts after the first iteration, is that they can see and experience a working prototype and based on their expertise, once they see what is possible, they can provide valuable input for the MLX system. An example of this type of requirement acquisition is depicted in the study of Schneider et al. (2017), where authors interviewed experts on public speaking to identify non-verbal communication aspects that could be practiced with the use of their Presentation Trainer prototype.

Results from the previous iteration can also provide insights that support the design of the user interaction with the MLX system. For example, results from the study of Schneider et al. (2015b), show that, for the effective use of the Presentation Trainer, learners should be familiar with the feedback that they might receive, and should not improvise while practicing their presentations.

The development phase of this second iteration usually is more straightforward than the first one because researchers and developers are already familiar with the used technologies and the system architecture to follow. However, it still requires considerable time and effort on fine-tuning several aspects, such as UX, software reliability, of the prototype to reach a state that is good enough to test its effectiveness. The specific type of fine-tuning depends on the prototype.

For this second iteration, the MLX system prototype might have reached a stage where it is possible to evaluate its effectiveness. Experiments or quasi-experiments can be used to conduct this evaluation. A tricky part of this evaluation is identifying a valid procedure to follow for the control group because, in most of the cases, there are no comparable treatments. Schneider et al. (2016) used as a control group the system with feedback disabled.

To examine the effectiveness of the treatment provided by the MLX system, one can look at log files that register the captured performance of the participants. Log files can point out evidence that learners correct some mistakes after receiving feedback, as is shown in the study of Di Mitri et al. (2020), where participants corrected their CPR technique accordingly after receiving audio feedback instructions provided by the CPR tutor. In the study of Schneider et al. (2016), log files show how participants from the treatment group significantly reduced the proportion of time making nonverbal communication mistakes for public speaking after each practice session. In contrast to participants from the control group whose proportion of time making mistakes remained constant throughout the different practice sessions.

It is not possible to predict whether the evaluation of this second iteration will provide results showing the effectiveness of the evaluated prototype. Therefore, it is important to keep using user experience, usability, and technology acceptance questionnaires. Even when the study shows clear effects of the tutor, these types of questionnaires will help to improve it for future iterations.

4.3 Following Iterations

The following iterations are used to improve the MLX system by adding new features and addressing old challenges. These following iterations can also help to evaluate different aspects of the tutor like long-term effects, long-term use, effects on a different population, etc.

5 Application Use Cases

In previous sections, we have summarised the theory of deliberate practice, defined the MLX system, an engineering model that helps as a guide to developing specific MLX applications, and presented a methodology that allows us to evaluate and refine these applications. To show the potential of MLX for deliberate practice, in this section, we present an overview of concrete use cases that have used MLX to support deliberate practice.

5.1 Presentation Trainer

The Presentation Trainer (PT) is an MLX application designed to support the development of non-verbal communication skills for public speaking through deliberate practice. Learners can practice the delivery of their presentations in authentic settings, which is a vital aspect of deliberate practice (Neelen & Kirschner, 2016). The PT will analyse the non-verbal communication of the learner. Based on the results of the analysis, the PT then provides the learner with real-time feedback instructions to foster conscious practice and correct some basic non-verbal communication mistakes for public speaking such as voice volume, body posture, facial expressions, use of gestures, and use of pauses. The feedback instructions are displayed on a screen which should be placed in front of the learner (Schneider et al., 2016) or on an AR display showing also a virtual audience (Schneider et al., 2019; see also Fig. 2, left). A report of the mistakes performed during a practice session and the practice session is recorded and can be shown to the learner for self-reflection.

The research of the PT has gone through multiple iterations from the DBR methodology. Results from the different evaluations show that learners practicing with the PT significantly reduce the proportion of time of making non-verbal communication mistakes after each practice session (Schneider et al., 2016).

5.2 CPR Tutor

The CPR Tutor (Di Mitri et al., 2019a) is a real-time multimodal feedback system for cardiopulmonary resuscitation (CPR) training. The CPR Tutor detects mistakes using recurrent neural networks for real-time time-series classification. It automatically recognises and assesses the quality of the chest compressions according to five performance indicators such as, correct locking of the arms, correct use of the body position, right compression rate, release, and depth. The chest compressions are classified from a multimodal data stream consisting of kinematic and electromyographic data (see Fig. 2, right-end). Based on this assessment, the CPR Tutor provides audio feedback to correct the most critical mistakes and improve CPR performance. The CPR Tutor was designed by running two experiments, the first which aimed to design the neural network architecture for detecting the mistakes in the chest compression (Di Mitri et al., 2019b); the second experiment embedded the mistake detection system in a real-time feedback architecture. A dataset from 10 experts was used for model training. The impact of the feedback functionality, we ran a user study involving ten participants. Although long-lasting learning effects cannot be acknowledged given the insufficient number of participants, the results of the feedback study show a short-term positive improvement in the error rate on the five target performance indicators.

5.3 Calligraphy Tutor

The Calligraphy Tutor (Limbu et al., 2019b) is an MLX application designed to allow learners to practice handwriting deliberately. It was developed using the Microsoft Surface tablet and Myo armband. It was designed using the ID4AR

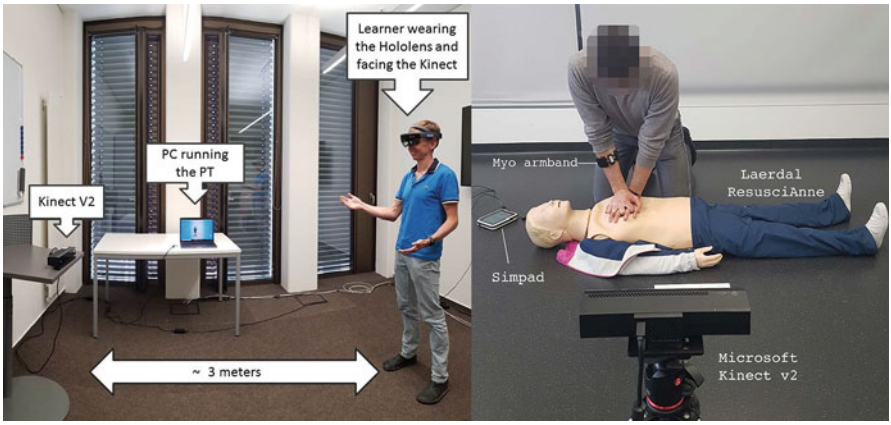

Fig. 2 Left: Learner practicing a presentation with the PT in front of a virtual audience. Right: Learner practicing CPR with the CPR tutor

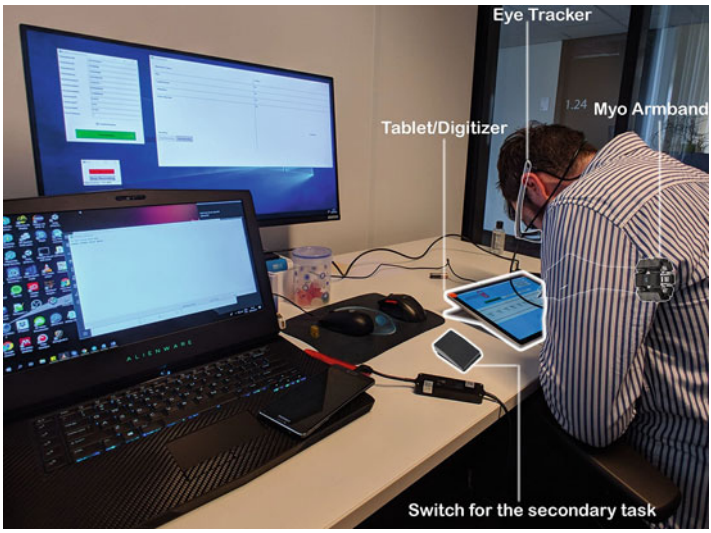

Fig. 3 Set up for Calligraphy tutor

framework (Limbu et al., 2018) for fostering deliberate practice. While writing notes by hand is often considered beneficial for a deep understanding of concepts, handwriting is a complex perceptual-motor skill that needs numerous hours of practice to master. The calligraphy tutor application allows the teacher to create practice materials quickly using the pen and the Myo data. It also allows the learner to practice from the material while receiving guidance and feedback. The application provides feedback on various perceptual-psychomotor attributes involved in performing calligraphy.

The research on the calligraphy tutor only went through one iteration. The focus of this iteration was to ensure that the design of feedback would not demand excessive mental effort from the learner because practicing deliberately is already cognitively demanding. The results of the study show that the application and the feedback do not impose additional mental effort (Limbu et al. 2019b; see Fig. 3 for study setup).

5.4 Table Tennis Tutor

The Table Tennis Tutor (T3) (Mat Sanusi et al., 2021) is an MLX application designed specifically for psychomotor skills development in the table tennis domain. The application was built to address the sport's complexity which involves numerous dynamic movements and difficult techniques that require the learners to repeatedly practice the specific skill to the extent that the muscle memory is trained

Fig. 4 Left: Learner practicing the forehand table tennis stroke with the T3. Right: Learner practicing how to provide maintenance to a model of a Mars rover

(Feder & Majnemer, 2007), which automates them. In this study, the T3 allows beginners to practice the most fundamental technique – the forehand stroke. The application uses the Multimodal Pipeline framework (Di Mitri et al., 2019b) for the creation of a data corpus by collecting the motion data and tracking the human body using smartphone sensors (accelerometer and gyroscope) and the Kinect's depth camera sensor, respectively. Moreover, the study intended to explore the potential of using only motion data from a smartphone for learners to practice the stroke since such a device is commonly owned (the study setup is shown in Fig. 4, left).

The results show that the combination of smartphone sensors and the Kinect achieved a higher classification rate than the smartphone sensors alone, which emphasises the importance of the multimodal approach in classifying complex activities. The research of the T3 only went through one iteration but provides a significant step into designing the feedback system. It is vital that the feedback system should be designed with the least demanding mental effort as the sport itself, and practicing it deliberately, can be physically and mentally exhausting (Mat Sanusi et al., 2021).

5.5 Astronaut Training

The WEKIT (Ravagnolo et al., 2019) application is an augmented reality application for training astronauts, developed in the context of a European project called Wearable Experience for Knowledge Intensive Training.[1] It was developed for

[1] http://wekit.eu/

Microsoft Hololens and was designed to be used with a wearable harness consisting of various sensors. The WEKIT application uses the ID4AR framework (Limbu et al., 2018) for fostering deliberate practice. The WEKIT application has two major interfaces, the "expert mode" for the experts to create a learning material and the "student mode" for the student to learn from the recorded material. The objective of the WEKIT application was to reduce the time and cost associated with training the astronauts on the ground and potentially train the astronauts during their journey. By using AR and recorded expert data, the astronaut can train by himself in space.

The WEKIT application was tested in two iterations. In the first iteration, the usability of the system was tested, which was found to be acceptable. In the second iteration, the learning effectiveness of the application was tested using a control group. The results found no difference in learning between the group that used the WEKIT application and the control group that used instructional manuals (Limbu et al., 2019a; see Fig. 4, right for the study setup).

5.6 Commonalities and Differences

Table 1 displays the presented use cases showing their commonalities and differences based on our MLX system architecture as presented in Sect. 3.1. As seen in Table 1, there are three main different Task Models: specific structured, semi-structured, and generic unstructured. Specific structured Task Models focus on well-defined tasks that have been repeated multiple times in order to master them. Supervised classification with Neural Networks can be used to identify the performance indicators for the Student Models. This specific structure allows the Tutoring Model of our use cases to objectively assess the learners' performance and provide feedback based on the objective assessment.

Semi-structured Task Models in our use cases follow a rule-based Student Model approach. The tasks practiced by the learners are, to a certain degree, quite flexible, e.g. there are multiple ways to do a presentation and multiple ways to write a text. The semi-structured approach allows the Tutoring Model of our use cases to provide learners' with feedback based on the predefined rules. However, the assessment of the performance cannot be completely objective.

Unstructured Task Models, on the other hand, focus on providing generic instructions based on the recorded expert model in which then students imitate the defined tasks. They offer support for multiple types of learning tasks. However, this flexibility of tasks restricts the Tutoring Model of the system to process-based guidance without any performance feedback.

Table 1 MLX systems applied to the different use cases

Case	Modality	Learner Model	Task Model	Tutoring Model	User Interface	Number of DBR cycles
PT	Motion, audio, posture	A rule-based model derived from experts to assess the performance	Semi-structured: Practicing the delivery of a presentation	One instructional feedback at a given time (Realtime) Retrospective feedback about the recorded session	Depth camera, Screen, VR display	4
CPR	Motion, posture, myographic data	Supervised performance classification with Neural Networks	Specific, structured: Chest compressions	Real-time audio feedback based on 5 performance indicators	Depth camera, myographic band, audio	2
Calli- graphy	Hand movement, pen pressure	Rule-based measurement of the derivation of student performance to the expert task model	Specific, semi-structured: Expert recorded calligraphy task	Real-time visual feedback highlighted in student's calligraphy attempt	Tablet with a sensor-enabled pen with calligraphy application	1
T3	Motion, posture	Supervised performance classification with Neural Networks	Specific, structured: Expert recorded correct and incorrect table tennis stroke	Automatic audio feedback when an error is detected (future work)	Smartphone motion sensors, smartphone application, depth camera	1
Astro- naut	Motion, posture, hands	The student re-enacts the recorded expert model. Defined task completion status.	Generic, unstructured: Mixture of authored and recorded task models created by a domain expert	Process-based task guidance	Wearable augmented reality application embedded in a physical training environment	2

6 Conclusions and Future Work

In this book chapter, we introduced the reader to *multimodal learning experiences for deliberate practice* to address a research gap in educational technology research: the tendency of designing systems fitted on technological devices rather than on the learning process. MLX systems aim to address this gap through multimodal data collection and analysis, and the use of ubiquitous and immersive systems. MLX systems rely on multi-sensor networks and data collection that leverage immersive and ubiquitous technologies to support embodied learning and deliberate practice in authentic settings. In Sect. 2, we have described the theoretical foundations of MLX systems, arguing why the current educational theories such as the dual-coding theory must be updated. In Sect. 3, we have detailed the engineering aspect of MLX systems and introduced a prototypical architecture design. In Sect. 4, we have described the experimental procedure for evaluating and validating these systems. Finally, in Sect. 5, we have listed some relevant use cases for the MLX system presenting their tested potential to support deliberate practice.

When contemplating the future of MLX systems, it is inevitable to consider the existing philosophical discussion on whether AI systems are set to augment human abilities or replace them completely. This discussion applies without exceptions to learning, about the degree of activation the AI systems assume in the learning process and to what extent they are set to imitate the cognition of the human expert and replace human feedback. Educational researchers have different stands concerning the autonomy the AI systems should take for education and practice. Beyond the work on autonomous and intelligent tutoring systems in the last decades, new approaches based on human-AI hybrid intelligence have been recently proliferating, especially within the CrossMMLA community. These systems do not necessarily rely on *automatic feedback* but rather on *intelligence augmentation* (Cukurova et al., 2019). For example, multimodal data indicators can be displayed in the learning analytics dashboard for both the teachers and learners (Jivet et al., 2018); similarly, using data storytelling approaches, it is possible to create narratives to better communicate the insights of data to the learners (Martinez-Maldonado et al., 2020). Such hybrid approaches keep the human in the loop of the sense-making process and are aimed to engage in reflection and debriefing moments where learners and teachers can discuss and gain insights into their performance.

It is also important to state that it is not affordable, nor feasible, to have human experts provide learners with personalised feedback and instruction whenever learners want to engage in deliberate practice. MLX systems can bridge this gap enabling learners to practice their skills and receive personalised feedback at any time. Moreover, while conducting our research, we have experienced subtle, nonetheless important differences between feedback provided by human tutors and MLX systems. Feedback from human tutors inevitably contains an emotional aspect, which can be incredibly motivating for learners. However, it can also be overwhelming and, in some cases, detrimental. The emotional aspect of the

feedback from MLX systems is reduced, allowing learners to make as many mistakes as needed to improve their skills.

Humans, as teachers and learners, have natural virtues and limitations. The same applies to MLX systems. We can observe that MLX systems have to compromise the flexibility of their Task Model with the capabilities of their Tutoring model. High levels of flexibility demand a loss in the objectivity of the assessment and thus capabilities of instruction. We, therefore, consider that human tutors and MLX systems are not competing for mutually exclusive variables. Hence, we argue that learners, human tutors, and MLX systems together can conform to an eco-system able to take digital learning to unprecedented levels.

References

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*(1), 16–25.

Chandler, P., & Tricot, A. (2015). Mind your body: The essential role of body movements in children's learning. *Educational Psychology Review, 27*(3), 365–370. https://doi.org/10.1007/s10648-015-9333-3

Clark, R. E., & Voogel, A. (1985). Transfer of training principles for instructional design. *ECTJ, 33*, 113. https://doi.org/10.1007/BF02769112

Clark, J., Crandall, P., Pellegrino, R., & Shabatura, J. (2019). Assessing smart glasses-based foodservice training: An embodied learning theory approach. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, 45. https://doi.org/10.21432/cjlt27838

James M. Clark, and Allan Paivio. "Dual coding theory and education." Educational Psychology Review 3, no. 3 (1991): 149–210. https://doi.org/10.1007/BF01320076.

Cukurova, M., Kent, C., & Luckin, R. (2019). Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring. *British Journal of Educational Technology*, bjet.12829. https://doi.org/10.1111/bjet.12829

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning, 34*(4), 338–349.

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2019a). Detecting mistakes in CPR training with multimodal data and neural networks. *Sensors, 19*(14), 3099.

Di Mitri, D., Schneider, J., Klemke, R., Specht, M., & Drachsler, H. (2019b). Read between the lines: An annotation tool for multimodal data for learning. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 51–60).

Di Mitri, D., Schneider, J., Trebing, K., Sopka, S., Specht, M., & Drachsler, H. (2020, July). Real-time multimodal feedback with the CPR tutor. In *International conference on artificial intelligence in education* (pp. 141–152). Springer.

Dillenbourg, P. (2016). The evolution of research on digital education. *International Journal of Artificial Intelligence in Education, 26*(2), 544–560. https://doi.org/10.1007/s40593-016-0106-z

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard Business Review, 85*(7–8), 114–121. https://doi.org/Article

Feder, K. P., & Majnemer, A. (2007). Handwriting development, competency, and intervention. *Developmental Medicine & Child Neurology, 49*(4), 312–317.

Goldman Schuyler, K. (2010). Increasing leadership integrity through mind training and embodied learning. *Consulting Psychology Journal: Practice and Research, 62*(1), 21–28.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology and Society, 15*(3), 42–57.

Gruber, H., Law, L.-C., Mandl, H., & Renkl, A. (1995). Situated learning and transfer. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines*. Elsevier Science.

Guest, W., Wild, F., Vovk, A., Fominykh, M., Limbu, B., Klemke, R., Sharma, P., Karjalainen, J., Smith, C., Rasool, J., Aswat, S., & Schneider, J. (2017, September). Affordances for capturing and re-enacting expert performance with wearables. In *European conference on technology enhanced learning* (pp. 403–409). Springer.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). *License to evaluate: Preparing learning analytics dashboards for educational practice* (p. 40). https://doi.org/10.1145/3170358.3170421

Juntunen, M. L. (2020). Embodied learning through and for collaborative multimodal composing: A case in a Finnish lower secondary music classroom. *International Journal of Education & the Arts, 21*(29).

Kiefer, M., Schuler, S., Mayer, C., Trumpp, N., Hille, K., & Sachse, S. (2015). Handwriting or typewriting? The influence of pen or keyboard-based writing training on reading and writing performance in preschool handwriting. *Advances in Cognitive Psychology, 11*(4), 136–146. https://doi.org/10.5709/acp-0178-7

Limbu, B. H. (2020). *Multimodal interaction for deliberate practice*. PhD thesis, Open University of the Netherlands, Heerlen, The Netherlands.

Limbu, B. H., Jarodzka, H., Klemke, R., & Specht, M. (2018). Using sensors and augmented reality to train apprentices using recorded expert performance: A systematic literature review. *Educational Research Review, 25*, 1–22.

Limbu, B., Vovk, A., Jarodzka, H., Klemke, R., Wild, F., & Specht, M. (2019a). WEKIT. One: A sensor-based augmented reality system for experience capture and re-enactment. In *Transforming learning with meaningful technologies*. EC-TEL 2019. LNCS11722. Springer.

Limbu, B. H., Jarodzka, H., Klemke, R., & Specht, M. (2019b). Can you ink while you blink? Assessing mental effort in a sensor-based calligraphy trainer. *Sensors, 19*(14), 3244. https://doi.org/10.3390/s19143244

Longcamp, M., Anton, J.-L., Roth, M., & Velay, J.-L. (2003). Visual presentation of single letters activates a premotor area involved in writing. *NeuroImage, 19*(4), 1492–1500. https://doi.org/10.1016/S1053-8119(03)00088-0

Mat Sanusi, K. A., Mitri, D. D., Limbu, B., & Klemke, R. (2021). Table tennis tutor: Forehand strokes classification based on multimodal data and neural networks. *Sensors, 21*(9), 3121.

Martinez-Maldonado, Roberto, Vanessa Echeverria, Gloria Fernandez Nieto, and Simon Buckingham Shum. "From data to insights: A layered storytelling approach for multimodal learning analytics." In Proceedings of the 2020 CHI conference on human factors in computing systems, 1–15. : ACM, 2020. https://doi.org/https://doi.org/10.1145/3313831.3376148.

Neelen, M., & Kirschner, P. A. (2016). *Deliberate practice: What it is and what it isn't – 3-Star learning experiences*. Retrieved March 18, 2018, from https://3starlearningexperiences.wordpress.com/2016/06/21/370/

Odena, O. (2012). Creativity in secondary music classroom. In G. E. McPherson & G. F. Welch (Eds.), *Oxford handbook of music education* (Vol. 1, pp. 512–527). Oxford University Press.

Ravagnolo, L., Helin, K., Musso, I., Sapone, R., Vizzi, C., Wild, F., Vovk, A., Limbu, B., Ransley, M., Smith, C., & Rasool, J. (2019). *Enhancing crew training for exploration missions: The WEKIT experience*. International Astronautical Federation.

Rikers, R. M. J. P., Van Gerven, P. W. M., & Schmidt, H. G. (2004). Cognitive load theory as a tool for expertise development. *Instructional Science, 32*(1), 173–182. https://doi.org/10.1023/B:TRUC.0000021807.49315.31

Robbins, P., & Aydede, M. (2009). A short primer on situated cognition. In *The Cambridge handbook of situated cognition*. Cambridge University Press.

Schneider, J., Börner, D., Van Rosmalen, P., & Specht, M. (2015a). Augmenting the senses: A review on sensor-based learning support. *Sensors, 15*(2), 4097–4133.

Schneider, J., Börner, D., Van Rosmalen, P., & Specht, M. (2015b). Stand tall and raise your voice! A study on the presentation trainer. In *Design for teaching and learning in a networked world* (pp. 311–324). Springer.

Schneider, J., Börner, D., Van Rosmalen, P., & Specht, M. (2016). Can you help me with my pitch? Studying a tool for real-time automated feedback. *IEEE Transactions on Learning Technologies, 9*(4), 318–327.

Schneider, J., Börner, D., Van Rosmalen, P., & Specht, M. (2017). Presentation Trainer: What experts and computers can tell about your nonverbal communication. *Journal of Computer Assisted Learning, 33*(2), 164–177.

Schneider, J., Di Mitri, D., Limbu, B., & Drachsler, H. (2018, September). Multimodal learning hub: A tool for capturing customisable multimodal learning experiences. In *European conference on technology enhanced learning* (pp. 45–58). Springer.

Schneider, J., Romano, G., & Drachsler, H. (2019). Beyond reality—Extending a presentation trainer with an immersive VR module. *Sensors, 19*(16), 3457.

Van Merriënboer, J. J., Clark, R. E., & De Croock, M. B. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development, 50*(2), 39–61.